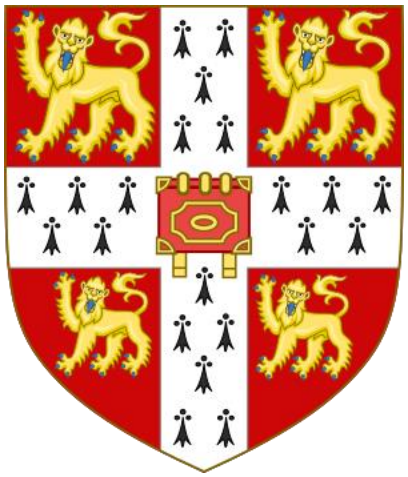
Tryptophanase Regulatory Mechanisms in *Escherichia coli*

Ellis Caitlin Kelly

Gonville and Caius College

Department of Genetics



This thesis is submitted for the degree of Doctor of Philosophy

January 2024

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted, or, is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Title

Tryptophanase Regulatory Mechanisms in *Escherichia coli.*

Abstract

Indole is a heterocyclic signaling molecule, synthesized by a wide array of bacterial species, but most studied in *Escherichia coli*. Indole is formed by the degradation of L-tryptophan through the action of the enzyme tryptophanase (TnaA). The *tnaA* gene is part of the tryptophanase (*tna*) operon, which is regulated by catabolite repression and tryptophan-induced transcription antitermination. Recent studies have unveiled two distinct kinetic modes of indole signalling in *E. coli*: a long-lasting but low-level (persistent) signal, and a transient, high-level (pulse) signal. The underlying mechanisms that regulate the indole pulse phenotype have not been fully characterised. Pulse signalling is thought to be linked to the increased expression of tryptophanase during the transition from exponential to stationary phase. Nonetheless, empirical data from the literature indicates the presence of tryptophanase during early to mid-exponential phase when indole synthesis is low or non-existent. This contradictory evidence indicates the existence of additional regulatory mechanisms controlling tryptophanase activity, and the indole pulse.

This work has combined conventional shake-flask culture approaches with single-cell analysis to investigate tryptophanase expression and activity across all *E. coli* growth phases. From these studies a hypothesis has emerged that the indole pulse is triggered, at least in part, by post-translational activation of tryptophanase that results in the surge of indole production. Initial experiments focused on characterising the indole pulse. Significant variability was observed in the timing of the indole pulse in shake-flask culture, its timing varying within a window of about 40 min and its duration ranging from 30 to 45 minutes. The inability experimentally to regulate the timing of the pulse by mutation of the *tna* operon or by plasmid-based tryptophanase expression indicated a post-translational regulatory mechanism might be acting on tryptophanase. Tryptophanase expression measured throughout growth showed tryptophanase protein levels *per* cell are similar in

exponential and early stationary phase. This suggests the pulse is triggered by both *de novo* synthesis and activation of pre-existing tryptophanase enzyme. Tryptophanase activity assays on enzyme harvested at different growth phases were consistent with this idea. An indication of the mechanism of post-translational regulation came from the use of single-cell microfluidic studies. Within individual cells, tryptophanase was either concentrated at a polar focus or dispersed throughout the cell. The dispersal of polar foci appears to be critical for the triggering of the indole pulse.

As a result of the COVID-pandemic, laboratory access was restricted for a substantial portion of this study, prompting the use of alternative research methods. Consequently, an exploration of tryptophanase distribution and potential evolutionary history was undertaken in *E. coli* and other gut microbial species. The results indicate TnaA is maintained in *E. coli*, and there is evidence that it is a component of the core genome. TnaA extends beyond *E. coli*, being identified in 10% of the bacterial species sequenced in the microbiome. It is present across various phyla of the gut microbiome, including Proteobacteria, Firmicutes, and Bacteroidota.

Acknowledgments

To my esteemed supervisor, dedicated coworkers and supportive family,

I am deeply grateful for the unwavering support, guidance, and encouragement that each you gave me during this academic journey.

To my supervisor, David Summers, your mentorship, and expertise have shaped my research, challenged my critical thinking, and propelled me to become a better scientist. Your unwavering commitment to my growth as a scholar has been incredible. I will cherish every conversation, story, and debate with fondness; it has been an honour.

As well, I would like to thank the Cambridge Trust and Gonville and Caius for the generous support to finance my journey as I developed as a scientist.

To my collaborators, Somenath and Georgeos. Working with you added so much to my thesis, and I always enjoyed learning about your insights. Your willingness to share ideas and tackle challenges, enriched the quality of my thesis and I am very fortunate.

To the members of the Summers' lab (Ash, Cameron, Marta and Junyan), as well as honorary members (Claudio and Jehangir), each person imparted valuable wisdom and had insightful questions, contributing significantly to my learning. Thank you to my students (Part II and supervision students) for giving me the opportunity to develop as a teacher.

Lastly, to my loving family, your unwavering support, understanding and unshakeable belief in my abilities have been the bedrock of my academic pursuit. To my parents for believing in me, even during times when I lacked that confidence in myself, and for their steadfast support as I moved across the world to pursue my passion for studying bacteria. I owe my love of microbiology to my sister, who instilled this passion in me, and for constantly reminding me to appreciate the beauty of science in every aspect of life. Thank you to my siblings, Ciaran, Naomi, and Jefferson for their patience and for always lending support. Thank you to my partner and love of my life, Ben, who has stood by my side through every high and low, and who has always served as my foremost proof-reader.

TABLE OF CONTENTS

List of Figures

4.2. Experimental procedure for estimating the per-cell fluorescence of TnaA-GFP in *E. coli* cells grown in LB medium.

4.3. Fluorescence of TnaA-GFP over time in BW25113 *ΔtnaCAB pCA24N* cultures with corresponding supernatant indole concentrations.

4.4. TnaA-GFP fluorescence measurements of *E. coli* cultures sampled during exponential or stationary phase.

4.5. Fluorescence of TnaA-GFP per cell in BW25113 *tnaA-GFP* kan^R cultures over a period of 5 hours.

4.6. Indole production from tryptophan in *E. coli* BW25113 *tnaA-GFP* kan^R cells harvested at two different growth phases: exponential (OD_{600} = approx. 0.2) and stationary (OD_{600} = approx. 2.0).

4.7. Indole production from tryptophan in *E. coli BW25113 tnaA-GFP* kan^R samples taken at three different growth phases, equalized for TnaA-GFP fluorescence.

4.8. Indole production from tryptophan (0.3 mM) in *E. coli* BW25113 WT cultures in the presence (teal) and absence (orange) of glucose.

5.1. The mother machine experimental platform.

5.2. MI13 *tnaA-GFP* kan^R cells are shown in channels at significant optical densities (panels a-f) throughout the growth curve.

5.3. The panels show MI13 *tnaA-GFP* kan^R cells in deep stationary phase that have been loaded into the mother machine microfluidic chip.

5.4. A mother cell with a focus containing TnaA-GFP that appears to dilute over time from lag phase to mid-exponential.

5.5. A mother cell observed throughout a growth cycle that appears to not divide or lose any of its TnaA-GFP.

5.6. Cells from an overnight culture of *E. coli* MI13 *tnaA-GFP* kan^R were added to the mother machine microfluidic chip and the associated Erlenmeyer flask containing LB medium with an additional 0.5 mM tryptophan and pluronic fluid (0.085%).

5.7. Mean fluorescence intensity of GFP in *E. coli* MI13 *tnaA-GFP kan*^{R} cells plotted over time in the mother machine microfluidic device.

5.8. Mean fluorescence intensity of GFP with confidence intervals in *E. coli* MI13 *tnaA-GFP kan*^{R} cells plotted over time in the mother machine microfluidic device.

5.9. Percentage of *E. coli* MI13 *tnaA-GFP kan*^{R} cells that contains a focus at each time point.

5.10. MI13 *tnaA-GFP kan*^{R} cells are shown in channels during exponential phase, small foci appear in numerous cells, as identified by the red arrows.

5.11. The distribution of TnaA-GFP fluorescence at time = 0. *E. coli* MI13 *tnaA-GFP kan*^{R} cells were inoculated at OD_{600} = 0.05 in LB medium with additional 0.5 mM tryptophan and seeded into channels to the Mother Machine device.

5.12. The distribution of TnaA-GFP (presence or absence of foci) within 15,736 individual cells in the Mother Machine microfluidic device at time = 0, following the introduction of cells into the microfluidic chip.

5.13. TnaA-GFP fluorescence at time = 0 of cells that contain a focus (blue bars) or do not contain a focus (orange bars).

5.14. The mean time to division (min) for cells that have a focus versus cells that do not.

6.1. Metadata represented in bar plots obtained from NCBI BioSample from the 1251 *E. coli* genomes in the dataset.

6.2. A phylogenetic tree of the *E. coli* strains showing genetic relatedness (inner ring) and sequence types (STs) annotated on the outer ring.

6.3. Histograms that represent the BLASTn matches from tnaA and tnaB and their distributions based on the following parameters: percentage identity, alignment length and E-values before alignment thresholds were applied.

6.4. Phylogenetic trees depicting tnaA diversity in *E. coli.*

6.5. The Unified Human Gastrointestinal Protein (UHGP) database represents over 4,616 bacterial species from across the different phyla that were detected from the human gut.

6.6. Comparison of the enzymatic reactions and crystal structure between the tetrameric enzymes tryptophanase (TnaA) and tyrosine phenol lyase (TPL).

6.7. Presence and frequency of TnaA in different phyla of the gut microbiome.

6.8. Evolutionary history of TnaA in the human gut microbiome.

7.1. A new model for triggering the indole pulse by the dispersal of tryptophanase foci and activation of the enzyme.

List of Tables

6.2. Species information from the 1251 genomes in the BLAST database. The table shows the number and percentage of the genomes that was represented by each species which was determined by PathogenWatch.

6.3. A measure of nucleotide diversity determined by the maximum and mean number of single nucleotide polymorphisms (SNPs) per site among *tnaA, tnaCAB* and MLST genes.

Abbreviations

AHL: Acyl-homoserine lactone

AHR: Aryl hydrocarbon receptor

AI-2: Auto-inducer 2

AIP: Auto-inducer peptide

AMR: Anti-microbial resistance

AR2: Adaptive response 2

ATP: Adenosine triphosphate

A/E: Attaching and effacing

cAMP: Cyclic adenosine monophosphate

CAP: cAMP receptor protein

DNA: Deoxyribonucleic acid

EDTA: Ethylenediaminetetraacetic acid

EHEC: Enterohaemorrhagic E. coli

FMT: Faecal microbiota transplantation

GAD: Glutamate decarboxylase

GF: Germ-free

GFP: Green fluorescence protein

GLP-1: Glucagon-Like Peptide-1

IAA: Indole-3-acetic acid

IBD: Inflammatory bowel disease

IBS: Irritable bowel syndrome

IECs: Intestinal epithelial cells

ILCs: Innate lymphoid cells

IL: Interleukin

IPA: Indole-3-propionic acid

IPTG: Isopropylthio-β-galactoside

IS: Indoxyl sulphate

LB: Luria-Bertani

LEE: Locus of enterocyte effacement

MLST: Multi-locus sequence typing

mRNA: Messenger ribonucleic acid

NADH/NAD+: Nicotinamide adenine dinucleotide

NF: Nuclear factor

OD600: Optical density measured at 600 nm

PBS: Phosphate-buffered saline

PDMS: Polydimethylsiloxan

PCR: Polymerase chain reaction

PLP: Pyridoxal 5'-phosphate

QS: Quorum sensing

Rcd: Regulator of cell division

RNA: Ribonucleic acid

Rut: Rho utilization

SNPs: single nucleotide polymorphism sites

SCFA: short-chain fatty acid

SOC: Super optimal broth with catabolite repression

ST: Sequence type

UHGP: United Human Gastrointestinal Protein

WT: Wild-type

1. Introduction

1.1 Overview of the human microbiome

“It is microbes who will have the last word” (Louis Pasteur). Most biodiversity on Earth is and always has been microbial. The human body is composed of trillions of microorganisms, slightly outnumbering the host cells (5). Microbes populate surfaces found within and surrounding the body, from skin and mucosal layers to the epithelial cells that line the human gastrointestinal (GI) tract. An enigmatic environment, the human microbiome is a fluctuating ecosystem, that encompasses microorganisms (bacteria, archaea, lower and higher eukaryotes, and viruses) and their genomes, as well as the surrounding environment (6, 7). The inherent dichotomy between the microbiome’s dynamic nature and the robustness of the microbiome is difficult to understand until the sensitivity to environmental factors is considered. These ecological factors are capable of disrupting or influencing its equilibrium. In the 21st century, there has been a substantial loss in diversity within the human intestinal microbiome compared to the average microbiome found within fecal samples from preindustrial societies (8, 9). The industrial revolution has significantly altered dietary habits, leading to a more widespread adoption of the Western Diet, which is typically low in polyunsaturated fatty acids and fibers, while rich in glucose and saturated fatty acids. This type of diet has been linked to an overall reduction in microbiota count and a shift in the abundance of certain microbial species (10). Research in both animal and human models has shown that the high-fat diets primarily cause an increase in the Firmicutes/Bacteroidetes ratio, often accompanied by a rise in *Proteobacteria* abundance, a decrease in Bacteroidete*s* (particularly families Prevotellaceae and Rikenellaceae) and a reduction in Actinobacteria (particularly Bifidobacterium species) (11). The shift in microbiota ratios is associated with a rise in chronic conditions, although the exact relationship between specific bacterial species and disease is complex. Metagenomic analyses, which study the nucleotide sequences of all organisms in a sample, have revealed reduced diversity of the bacterial phylum Firmicutes in the faecal microbiota of patients with Crohn’s disease (12). While causation between certain bacterial species and disease is unclear, the alarming trend of microbiota diversity loss over the last century has been linked to a plethora of emerging health

issues, including gastrointestinal cancers, autoimmune diseases, infectious, liver diseases, metabolic disorders, respiratory conditions, and mental health disorders (13, 14). The gut microbiota has a mutualistic relationship with its host and this relationship is quintessential in maintaining health and metabolic homeostasis and when disrupted, can result in disease. This dense, resident microbial community within the gut has co-evolved with the host and remains essential throughout life. It is linked to physiological processes that include promoting the health of the intestinal epithelial barrier, development of the immune system, and nutrient acquisition (15).

The complex and dynamic population of microorganisms that inhabit the GI tract constitutes the gut microbiota. In recent years there has been precedent to define the microbiota as a human organ, due to its metabolic activity and mass. The combined microbial genome contains approximately 500 times more genes than the human genome, leading to the intriguing notion of human genes being akin to mere background noise amidst the magnitude of microbial signals (16). Bacterial populations colonize the entirety of the GI tract, spanning from the oral cavity to the rectum. The concentration of microbial cells inhabiting the colon has been estimated at 10^{11} to 10^{12} per milliliter, making the colon one of the most densely populated bacterial habitats known on earth (17).

The density and composition of microbial communities exhibit significant variation across the anatomical sites, contingent upon various factors such as gut motility, host secretions, environmental conditions, substrate availability, and the integrity of the gastrointestinal mucosal barrier (18).

1.1.1 Microbiome composition

The exact composition of a person's microbiota is shaped by both inter-individual differences and fluctuations occurring within an individual over the course of their lifetime. The core microbiota is the quantity and identity of bacteria that are commensal among different individuals. These native bacteria are shaped early in life; established after birth, the microbiota continues to develop throughout life and is composed of many species of microorganisms, including bacteria, yeast, and viruses (19).

Taxonomically, bacteria are classified according to their phyla, classes, orders, families, genera and species. While only a few phyla are represented within the gut microbiome, they contain more than 160 species (20). The dominant gut microbial phyla are Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, Fusobacteria, and Verrucomicrobia. Obtaining a representative microbiome proves challenging due to the dynamic nature of species presence and abundance, which exhibit fluctuations both short-term and long-term in an individual, as seen in Figure 1.1. The composition of the gut microbiota is influenced by various factors over both the long term and short term. Throughout one's lifetime, chronic health conditions, infections, dietary habits, genetics, environmental exposure and medications establish the foundational composition of gut microbiota, that is often more stable after birth. However, short-term factors such as circadian rhythm can lead to daily fluctuations and variations in the microbial population (3). The microbiome exhibits both stability and plasticity, showing daily fluctuations while maintaining a personalized composition (see Figure 1.1). Even over a few days, these fluctuations occur, yet each person retains a unique microbiome. Remarkably, it is possible for two individuals to have no overlapping microbial species in their microbiomes (3).

The Firmicutes and Bacteroidetes represent 90% of the human gut microbiota, with the ratio between these two dominant bacterial divisions known to shift dramatically with age (21). Studies have shown that aging is associated with an increase in the proportion of Bacteroides, Proteobacteria, Eubacteria and Clostridiaceae. In contrast,

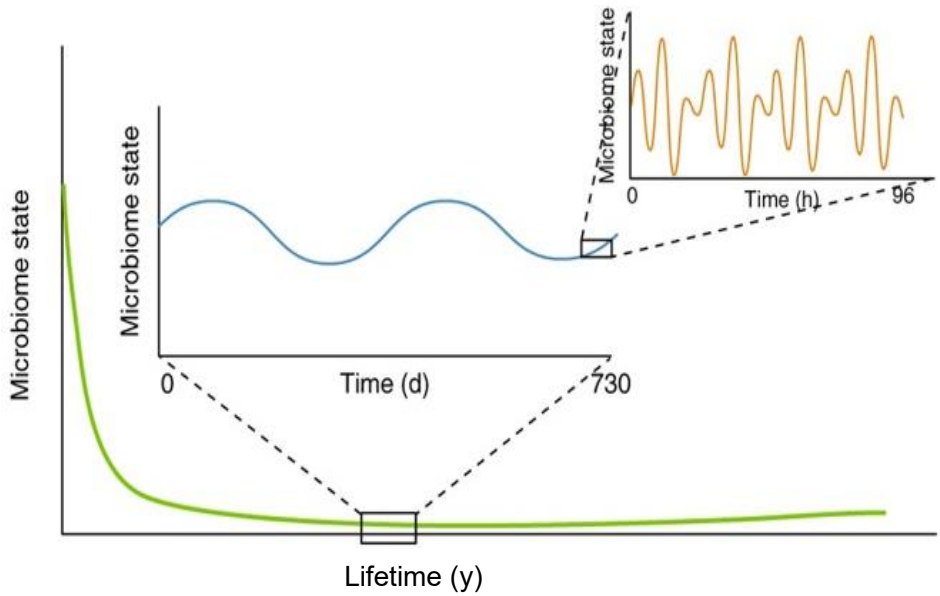


Figure 1.1 The dynamic composition of the human microbiome is shown over the course of a lifetime. A representative sample at any given time is difficult to obtain as the microbiome state settles after birth (green line), but the composition can vary over a short and long timescale (orange and blue line). (3)

Enterobacteriaceae are abundant in infants and the elderly, while Bifidobacterium is reduced in the elderly and enriched in infants (22). In the African turquoise killifish (*Nothobranchius furzeri*) experiments have shown an extended lifespan in older fish treated with antibiotics and subsequently recolonized with the microbiota of younger fish (23). This research suggests that the microbiota from younger fish may play a role in promoting longevity, highlighting the potential influence of gut microbiota on aging and health.

The Firmicutes to Bacteroidetes ratio decreases from approximately 10.9 in middle-aged adults to 0.6 in the elderly (24). In studies examining microbiome diversity in centenarians, researchers observed a lower abundance of symbiotic bacteria typically associated with health in younger age groups (such as *Faecibacterium* and *Bifidobacterium* spp.), accompanied by an increase in the prevalence of commensals (such as *Akkermansia* and *Odoribacter* spp.) and disease-related pathobionts (ex. *Streptococcus* spp.) (25, 26).

Another factor contributing to composition and diversity is dietary intake of the host. Diet is widely acknowledged as a major factor of gut microbiota variability. Comprehensive genetic sequencing of a specific region of the 16S ribosomal RNA gene (V3-V5 region) in stool samples showed associations between the consumption of different food groups and the overall diversity of microorganisms present in the sample (27). The Western diet is characterized by a higher intake of animal protein, saturated fats, sugar, and salt, and a reduced consumption of fiber. It is also linked to a reduction in microbial diversity, which could potentially contribute to the onset of a diseased state (28, 29).

While critical for the treatment of infections, antibiotics are a double-edged sword. Research findings have revealed antibiotics indiscriminately eliminated both pathogenic and beneficial microorganisms. Antibiotic use also disrupts community structure and function which can prolong susceptibility of the microbial gut environment to pathogens (30, 31). A study by Pérez-Cobas and colleagues explored the effects of antibiotic therapy on the microbial ecology of the GI microbiota and demonstrated an oscillation of Gram-negative population dynamics, ultimately leading to the overall collapse of microbial diversity (32).

1.1.2 Microbiome and the disease state

The composition and balance of the microbiome can have a significant impact on an individual's susceptibility to various diseases and their progression. Gut dysbiosis refers to an imbalance in the microbial communities residing in the gastrointestinal tract and has been linked to a variety of gastrointestinal conditions including inflammatory bowel disease (IBD) and irritable bowel syndrome (IBS), as well as a wider systemic manifestation of disease such as obesity, type 2 diabetes, cancer, mental illness and autoimmune diseases (33). In the absence of a gut microbiota, such deleterious effects do not manifest, as observed in germ-free mice. However, when a dysbiotic microbiome is transplanted, it can transfer the associated phenotype (34, 35). For instance, germ-free mice colonized with microbiota harvested from the caecum of either obese or lean donor mice showed significant differences. 16 rRNA gene sequencing confirmed a higher amount of Firmicutes in the obese microbiota. Mice colonized with these obese microbiota

exhibited a significantly greater percentage of body fat after two weeks, independent of food consumption or initial body fat, demonstrating the transfer of the obese phenotype through the dysbiotic microbiota (36).

These studies highlight the strong link between gut microbiota dysbiosis and the disease states. In another study, germ-free Apc^Min/+ mice, which are predisposed to developing intestinal neoplasms, received microbiota from Apc^Min/+ mice with colorectal tumors or from healthy wild-type mice. The transfer of microbiota from mice with colorectal cancer to germ-free mice resulted in significantly more intestinal tumours compared to those who received microbiota from healthy mice. Along with increased tumorigenesis, these mice also had elevated inflammatory markers: interleukin-6, tumor necrosis factor-alpha, and interleukin-17A. The increased bacterial taxa associated with the colorectal cancer microbiota included *Escherichia coli (E. coli)* and *Shigella* species, members of the Bacteroides genus, *Prevotella* species and *Akkermansia muciniphila*. While these bacterial species are part of the normal gut microbiota, in dysbiotic states, they may contribute to inflammation and cancer (37). These findings underscore a potential critical role of microbiota in health and disease, particularly in the context of dysbiosis.

In contrast to the dysbiotic state, the healthy gut microbiota is dominated by Firmicutes and Bacteroidetes phyla, which produce metabolites that promote immune homeostasis and colonization resistance. In the absence of critical commensal microbes, such as in germ-free (GF) and antibiotic-treated mice, the hosts are more vulnerable to enteric pathogen infection. A study by Darko and colleagues postulated that *C. difficile* infections persist by manipulating the indigenous microbiota to produce a metabolite called indole that acts to halt growth of competitors and subsequently allows *C. difficile* to flourish (38).

To combat the dysbiotic state of the microbiome, and the adverse effects on human health, research on probiotics and the ability to regulate commensal bacterial populations have been researched. Furthermore, in recent years, the proliferation of antimicrobial resistance (AMR), stemming from a multitude of factors, has resulted in a diminishing efficacy of antibiotic treatment (39). Given the emerging challenges posed by AMR, the utilization of microbiome modulation and other therapeutic intervention strategies will be

critical in the battle against infections. Fecal microbiota transplantation (FMT) from healthy donors into the GI tract of a recipient has been shown to be highly effective when used to treat patients with recurrent *C. difficile* infections and induce remission in ulcerative colitis patients (40), highlighting the role of FMTs as a therapeutic intervention. As scientists aim to provide therapeutic options for microbiome-associated diseases, it is important to consider and address the difficulties and limitations of microbiome research.

Studying the microbiome and its related diseases presents inherent difficulties, particularly in distinguishing causation from correlation. Only a fraction of intestinal bacteria can be assessed through standard cultivation techniques, with an estimated 80% not yet cultured (41). Current methods for studying the microbiome aim to overcome this limitation through advanced methods like metagenomics and sequencing, which may provide snapshots of microbial communities, but there is still a gap in understanding their role in disease states. Changes in the microbiome composition, which is primarily based on sequencing data, can be correlated with various conditions, but these associations might reflect the consequences of disease rather than its causes.

Proving causation is challenging due to the microbiome's complexity and its interactions with the host's immune system, metabolism and other physiological processes. Disentangling the direct effects of the microbiome from the host's response is difficult. Microbial interactions, which can be synergistic or antagonistic, further complicate attributing specific effects from individual species. Additionally, most analyses focus on fecal samples rather than mucosal samples, and often use 16S ribosomal RNA sequencing instead of more detailed shotgun metagenomics.

Human microbiome studies face additional challenges due to genetic and lifestyle differences among individuals, leading to significant variability. Factors like diet, medications, illness, and lifestyle changes cause temporal variability, making most human studies observational and not designed to prove causal relationships.

Animal model experiments have demonstrated the proof of principle that interfering in host-microbe interactions can delay or prevent diseases. These models allow for tight control over variables; using germ-free animals or genetically manipulated animals, strict

environmental controls, and the ability to sacrifice animals at the desired time-point. However, the translation these findings have to humans remains complex and challenging. The microbiome field is working to decipher host-microbe interactions and bacterial signalling to better understand their impact on host physiology and health.

1.1.3 Bacterial signalling in the gut microbiome

The gut bacteria form an intricate and interconnected relationship between various components of the body, including distant organs, the immune system, hormones, the brain, and the host metabolism often referred to as the gut-brain-axis (42). Apart from the influence that the gut microbiota has on host physiology, the microorganisms can communicate with one another through a variety of mechanisms. These communication processes that are crucial for the functioning and stability of the microbial community.

Central to the structure and evolution of the resident gut microbiota community are the various and complex interactions among microbes and their chemical environment. One well-characterised example of intercellular communication, called quorum sensing, is a classic example of bacterial signalling. The Quorum sensing system (QS) was discovered initially and characterised in the bioluminescent marine bacteria, *Vibrio fischeri* and *Vibrio harveyi,* where it was reported that it modified the expression of target genes in response to changing cell density (43, 44). Quorum sensing is used by gut bacteria to coordinate behaviour and response to environmental stimuli. This process involves the production and detection of specific signalling molecules known as autoinducers. When the bacteria population density in an environment increase*,* the concentration of these signalling molecules also rises*,* eventually reaching a threshold which triggers a coordinated response within the community. These collective responses encompass the control of gene expression, the potential formation of biofilms, and/or the production of virulence factors (45-47).

To date, three types of quorum signalling have been described: LuxR/LuxI-type quorum sensing system that uses acyl-homoserine lactones (AHL) as its signal; the auto-inducer peptide (AIP) system that uses small peptides, and the *luxS*/AI-2 quorum sensing system (48-53). The *luxS*/AI-2 quorum sensing system functions in both Gram-negative and

Gram-positive bacteria and exists in approximately half of all sequenced bacterial species (53). Autoinducer-2 (AI-2) is a signal molecule generated in a core metabolic pathway and produced by the enzyme, LuxS (54, 55). The action of AI-2 is species dependent. In Enterohemorrhagic *E. coli* (EHEC), the signal molecule is responsible for motility and modulation of type III secretion systems (56, 57). In *E. coli* K12, AI-2 influences biofilm formation (58, 59). AI-2 has also been found to control the expression of genes involved in bacterial metabolism, DNA repair, cell growth and division, metabolism, and nucleotide and protein biosynthesis (60, 61).

Signalling molecules produced by specific commensal bacteria have the potential to uphold the integrity of the intestinal barrier, regulate inflammatory processes, and enhance resistance against pathogen colonization (62). A metabolite produced by some members of the gut microbiota, indole, can increase the expression of genes that are involved in strengthening the mucosal barrier and mucin production. Further, treatment with this gut-microbiota derived molecule results in the reduction of the expression of pro-inflammatory genes, an increase in the expression of ant-inflammatory genes and a reduction in the colonization of pathogens in vitro (63).

Interactions among microbial species within the gut microbiota exhibit a range of effects, spanning from positive to deleterious. The positive interactions can be categorized into distinct groups, including cooperation, commensalism, and cross-feeding. Commensalism is a type of symbiosis in which one organism benefits from another without affecting the other organism positively or negatively. In contrast, true cooperation is defined as one species having evolved specifically to increase the fitness of another (64). Cooperation between microbial species may or may not be reciprocal; however, if one species incurs a cost, it must receive indirect fitness benefits for the interaction to be evolutionarily stable. An example of true and evolved interspecies cooperation in the gut microbiota is seen in the case of *Bacteroides ovatus.* These microbes engage in non-essential, costly polysaccharide digestion to assist other species, which is a rare case of genuine interspecies cooperation (65). In the gut microbiota, positive interactions are most often cross-feeding or also known as syntrophy. Syntrophy is classically defined as a form of obligately mutualistic metabolism, where together the partners exploit substrates

that neither could metabolize alone (66). Firmicute species like *Faecalibacterium prausnitzii* produce butyrate, a short chain fatty acid (SCFA), as an end-product of fermentation. This fermentation process also produces intermediate metabolites, like acetate and lactate, which can be utilized by other microbes. Bacteroidetes species, in turn, use the acetate produced by Firmicutes as a carbon source to support their own growth and metabolic activities (67, 68). This cross-feeding mechanism ensures efficient utilization of available nutrients. The metabolic by-products of one species become substrates for another, creating a synergistic relationship that enhances the overall metabolic output of the microbial community. Cross-feeding plays a pivotal role in molding the composition of the microbiota, its response to changes, the combined metabolic profile of the community, and its collective interactions with the host (64).

Bacteroides species play a disproportionately large role in the microbial ecosystem through cross-feeding mechanisms by degrading complex polysaccharides, strongly suggesting they are a keystone species (69). While there is no currently uniformly accepted definition, a bacterial keystone species may be described as microbial species that plays a disproportionately large role in maintaining the structure, function, and stability of a microbial community or ecosystem, despite its relatively low abundance (70). To determine the importance of key species in the microbiome's metabolic role and the potential impact of their absence on dysbiosis, researchers have created synthetic consortia. In one study, a synthetic consortium of 14 gut microbes was examined in batch bioreactors. Fermentations were repeated, each time leaving out one species out to assess the impact of its removal. The removal of *Bacteroides dorei* resulted in lower biomass and negatively affected 10 other species (71). This analysis showed *B. dorei* was essential for the consortium's function and may be a keystone species in the wider gut environment.

Detrimental interactions within the gut microbiota encompass both competitive and amensalistic relationships. An amensalistic relationship involves one bacterial species inhibiting or killing another bacterial species, while the other remains unaffected. In the context of bacteria, this usually involves the production of antibiotics of bacteriocin production. *Lactobacillus* species produce antimicrobial peptides like lactacin B which can

inhibit the growth of other bacteria, contributing to its dominance in this environment (72). These amensalistic relationships can shape the microbiota, by leading to the suppression or elimination of competing bacterial species, thereby influencing the structure and function of the microbiota.

Competition manifests in two forms: exploitative and interference. Exploitative competition is a scenario where one microbe species consumes resources more efficiently or at a higher rate than other species, thereby limiting the resources of the others. An example of exploitative competition is *C. difficile,* a Gram-positive bacterium that can cause serious disease by *exploiting* the gut environment when the normal microbiota is disrupted, such as after antibiotic treatment. *C. difficile* is capable for forming resilient spores that can withstand antibiotic treatment. When the commensal bacteria decrease, there is a reduction in colonization resistance, allowing spores to germinate into active cells. In this altered environment with less competition for nutrients, *C. difficile* proliferates and produces toxins, TcdA and TcdB. These toxins damage the intestinal lining, leading to inflammation and further disrupting the gut microbiota, exacerbating dysbiosis. This cycle of damage and dysbiosis creates a pathogenic environment, allowing *C. difficile to* thrive and cause severe gastrointestinal disease (73). In contrast, interference competition entails one species directly impacting the fitness of another (64). *Bdellovibrio bacteriovorus* is an aerobic Gram-negative bacterium, which preys upon other Gram-negative species, and acts as an ecological stabilizer within a microbial ecosystem by maintaining the stability of the ecosystem (74, 75). *Bdellovibrio* and similar microorganisms have been associated with a healthy human gut microbiome and have recently been proposed as a method to modulate the microbiome for medical applications (76, 77). *Ruminococcus obeum*, a bacterium residing within the gut microbiota, plays an important role in the fermentation of dietary fibres and complex carbohydrates in the intestinal environment. Notably, the expression of *R. obeum*'s autoinducer-2 (AI-2) synthase triggers quorum-sensing mediated repression of multiple colonization factors in *Vibrio cholera* (78).

Bacteria in the microbiome often engage in metabolic cooperation through the production and consumption of various metabolites. Colonic microorganisms are frequently

responsible for the degradation of proteins and peptides, yielding a diverse array of end products. These end products encompass short-chain fatty acids, ammonia, amines, phenols, indoles, mercaptans, carbon dioxide, hydrogen and hydrogen sulfide; a substantial portion of which possess toxic properties (79).

Microorganisms communicate with the host's immune system which is crucial for maintaining a balance between the host's immune responses and the presence of beneficial microorganisms while defending against potential pathogens. Commensal gut bacteria such as *Lactobacillus* have been suggested to catabolize the amino acid tryptophan into the metabolite indole-3-aldehyde. This specific metabolite serves as a ligand for the aryl hydrocarbon receptor (AHR) (80). AHR is expressed by innate lymphoid cells (ILCs), and upon activation, interleukin-22 expression is stimulated, a crucial cytokine responsible for maintaining intestinal mucosal homeostasis and enhancing resistance against the fungus *Candida albicans.* Impairment to the production of the AHR ligand by the gut microbiota is associated with IBD (81).

Microorganisms residing within the GI tract engage in multifaceted modes of intercellular communication, and among these mechanisms, the utilization of metabolites is an important aspect. Metabolites, as biochemical intermediates and end products of microbial reactions, assume an important role within the intricate milieu of the gut microbiota, shaping its environment and intricately modulating the dynamics of microbial interactions.

1.1.3.1 An introduction to metabolites in the gut microbiota

In the gut microbiome, it is estimated there are nearly 22 million bacterial genes. Of these, more than half occur only once, resulting in approximately 12.6 million unique genes (7, 82). Each of these genes have different functions. The genes commonly shared among bacteria in the gut microbiome are typically essential for the microbes' daily survival. In contrast, unique genes are often responsible for specialized function such as acquiring antibiotic resistance or producing specific metabolites. Researchers have predicted that 10% of metabolites found in mammalian blood are derived from gut microbes.

Leveraging these genes, the gut microbiota synthesizes a myriad of multifunctional enzymes to ferment undigested or indigestible compounds, yielding a diverse array of bioactive metabolites. Microbial-derived metabolites are mediators in microbe-host communication, which plays a pivotal role in the preservation of host physiology (83).

The most common metabolites produced by the action of gut microbiota are short-chain fatty acids, bile acids, amino acid derivatives, gases, polyamines, neurotransmitter precursors, secondary metabolites, and urolithins (83, 84). These metabolites serve as vital sources of energy (85), contribute to immune regulation (86, 87), modulate inflammation (88), aid in detoxification processes (89), support nutrient production (90), impact metabolism and absorption (91), facilitate communication along the gut-brain axis (92, 93), maintain gut health (94) and contribute to cellular growth and proliferation (95). Among these crucial metabolites is indole-3-propionic acid (IPA). Its significance is illustrated in studies involving germ-free mice. In these studies, germ-free mice were monocolonized with *Clostridium sporogenes,* which is capable of producing IPA (96). Notably, when the gene responsible for IPA production was deleted, it led to adverse host consequences, including intestinal permeability, inflammation and the production of antibodies against the targeting mutant strain (96, 97).

1.1.3.2 Indole as a metabolite in the microbiome

Indole is synthesised by the breakdown of tryptophan, through the action of the enzyme tryptophanase. L-tryptophan is among the nine essential amino acids that must be supplied through the diet to fulfill the body's nutritional needs. Most tryptophan is absorbed in the small intestine (98). A small portion of ingested tryptophan is used in protein synthesis and the remainder is metabolized by the host cells through two primary pathways: the kynurenine pathway (99) and the serotonin pathway (100). A small proportion (4-6%) of tryptophan is catabolized into indole and its derivatives by the intestinal flora in the colon. In mammals, indole originates exclusively from metabolic activity in the gut microbiota since host cells do not have the metabolic capacity for its production.

Indole is capable of crossing and binding to lipid membranes, thereby enabling its entry into any cells, independent of their capacity to produce the metabolite (101). In an environment such as the human gut, it is likely that bacteria lacking the ability to produce indole will still encounter indole. Measurement of the indole content present in fecal matter, reveals a wide variation in indole concentrations, ranging from 0.30 to 6.64 mM (102). As indole is absorbed through the colonic epithelium (103), it is anticipated that concentrations within the colon are considerably higher than those observed in fecal matter. Furthermore, concentrations in the distal regions of the intestine are expected to be notably higher than the proximal section (104).

Indole is frequently regarded as an enigmatic molecule, owing to its effects which range from detrimental to beneficial, and can vary significantly depending on the species in question (105). Most research pertaining to indole has been centered on the model bacterium, *E. coli*. *E. coli* is found in the intestinal tracts of over 90% of individuals, harmlessly residing as a commensal organism (106). In adults, 90% of the gut microbiota is composed of the phyla Firmicutes and Bacteriodetes (107) and Proteobacteria are far less common, with less than 1% of the microbiome being made up by *E. coli* (41). *E. coli*'s prevalence in the human gut microbiome, has naively and historically led to its recognition as the primary source of indole production, however this thesis will question this assumption. This is attributed to the fact that while other bacteria possess a similar *tnaA* gene, some of these bacteria are incapable of independently synthesizing indole (108).

Indole is the parent substance of a large number of important compounds that occur in nature. Indoles, collectively refer to derivatives of indole where one or more hydrogen atoms have been replaced by substituent groups. Indoles are directly linked to the homeostasis of intestinal flora and GI tract health. They exert their influence on the host through several crucial processes. Indoles modulate the intestinal barrier to uphold intestinal homeostasis by activating immune cells, thereby stimulating the release of anti-inflammatory factors such as interleukin-22 (IL-22) (80), and inhibiting the colonization of pathogenic bacteria including EHEC (63). Indole has been shown to have anti-inflammatory properties through the suppression of the pro-inflammatory cytokine interleukin-8 (IL-8). Indole can also attenuate the activation of nuclear factor (NF-κB)

signalling pathway, which plays a central role in the regulation of inflammation. Furthermore, it promotes secretion of the anti-inflammatory cytokine IL-10 (63) which counterbalances the pro-inflammatory effects of other cytokines. Oral administration of indole-capsules resulted in increased gene expression of molecules associated with tight junctions and adheren junctions in the colons of GF mice, leading to a reduction of intestinal permeability (109). Furthermore, indoles increase mucin expression to bolster the mucus barrier's protective function (110). In addition, indole derivatives enhance the integrity and function of intestinal epithelial cells (IECs) while concurrently suppressing inflammation and regulating gut insulin secretion. Indoles facilitate the restoration of depleted goblet cells via AHR and interleukin-10 (IL-10), thereby mitigating the effects typically associated with aging in mammals (111, 112).

Indole has been identified as a metabolite capable of modulating hormones such as Glucagon-Like Peptide-1 (GLP-1) secretion from enteroendocrine cells (111). At a concentration of 1 mM, indole exhibits a dual effect on GLP-1 secretion: it initially stimulates GLP-1 release, but over prolonged periods of time it exerts a suppressive influence on the rate of hormone secretion. Following the incubation period, it was observed that the cumulative release of GLP-1 in indole-treated cells was 35% lower compared to control cells (111). These findings raise the possibility of indole's potential involvement in metabolic diseases, such as type 2 diabetes.

While exerting a multitude of effects on the host's health, indole can also give rise to a wide array of consequences for bacteria residing within the microbiota. Indole can aid in detoxifying the gut by the reduction of absorption of harmful compounds and toxins. Indole mitigates cytotoxicity by *Klebsiella spp*. by suppressing toxin production and mitigates pathogenicity (113). Further, indole and its derivatives can exhibit antimicrobial properties by inhibiting the growth of potentially harmful bacteria and pathogens, which can help alleviate diseases such as hemorrhagic colitis (114). For example, indole may enhance the resilience of human intestinal cells, HCT-8, to colonization of pathogenic EHEC bacteria when exposed to norepinephrine (63).

The effects of indole signalling on interspecies connections among commensal bacteria and enteric pathogens have been increasingly investigated in recent years. A study by Rattanphan and colleagues found that chemically produced indole or indole-rich medium diminished biofilm formation and the virulence of *Listeria monocytogenes* by reducing cell motility and aggregation (115). In the non-indole-producing species, *Salmonella enterica* (116) and *Pseudomonas aeruginosa* (117), indole reduces virulence and increases host clearance of these pathogens. The presence of indole also attenuates virulence of the fungal species *Candida albicans* by regulating transcriptional factors that influence filamentation and biofilm formation (118).

In one of its extensively investigated model systems, indole is a factor that influences virulence of EHEC O157:H7, a pathogenic strain notorious for its propensity to induce severe infections in humans. In a study by Hirakawa and colleagues, indole elevated the secretion and production of EspA and EspB, which are proteins encoded by genes residing within the locus of enterocyte effacement (LEE4) operon(119). These proteins are essential for the formation of the type III secretion system. Additionally, indole was observed to increase the production of attaching and effacing (A/E) lesions, a pathogenic indicator linked to the disruption of the gut's normal structure and function (120). This study posits that indole has the potential to enhance virulence of EHEC by facilitating the secretion of virulence factors and enhancing the formation of the hallmark A/E lesions. These actions may contribute to the development of more severe infections in human.

However, indole exhibits a dual nature, and both indole and its derivatives can exert adverse effects upon bacteria and the host. Research associated with indole and its influence on pathophysiological processes demonstrated that chronic overproduction of indole in rats monocolonized with indole-positive *E. coli* (as opposed to an indole non-producing mutant) has been linked to anxiety-like behaviour and depression in these animals. Furthermore, a sustained production of high concentrations of indole by intestinal microbiota was observed to exacerbate the effects of chronic stress (121, 122). Indoxyl sulfate (IS), produced through the hepatic metabolism of indole exhibits renal toxicity and cardiovascular toxicity at high concentrations (123). These elevated

concentrations can induce multi-system dysfunction by promoting oxidative stress, inflammation, and other pathological alterations.

Despite extensive research on the role of indole within the microbiome, numerous aspects remain unexplored. All the specific bacteria responsible for indole production in the gut microbiota have not been definitively identified. Furthermore, its interactions within other microbiomes, such as the oral and vaginal microbiomes, are largely unknown. Additionally, the consequences of modulating indole-producing bacteria in the gut microbiota are not yet fully understood. For an inclusive understanding of indole's role in health and disease, it is imperative to comprehend the fundamental mechanisms governing the production and regulation of indole and tryptophanase.

1.2 Indole: an overview

Indole is a heterocyclic aromatic compound that is widely distributed in nature and produced by over 85 bacterial species (108, 124). It is formed of a five membered ring, which contains a nitrogen atom, attached to a six membered benzene ring, as seen in Figure **1.2** (124). While primarily synthesized by bacteria, indole is recognized as a universal signalling molecule with roles in communication across various domains, encompassing mammals, plants, insects and bacteria (108). Thus, indole serves as a signalling molecule that is effective not only within and between species, but also across different biological kingdoms.

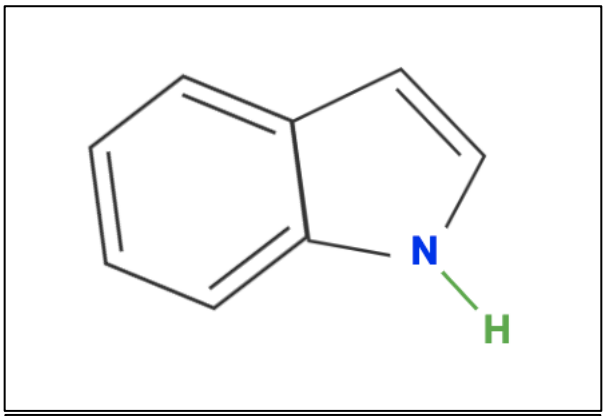


Figure 1.2. The structure of the molecule indole. *Created in Biorender.*

1.2.1 Indole signalling outside of bacteria

Indole is known primarily for its role in bacterial communication; however, indole and its derivatives can also be found in many diverse ecological niches.

As a response to an herbivore insect attack, plants can produce volatile compounds as a defense mechanism. Indole is one of these volatile compounds in maize, as it functions as a rapid and powerful aerial priming signal, readying systemic tissues and nearby plants for imminent threats (125). Indole is the most prolific nitrogen-containing volatile found in the petals of flowering plants and it is synthesized and emitted by over 30 distinct flowering plant species (126). The indole derivative, known as indole-3-acetic acid (IAA), is a type of plant hormone essential for regulating and orchestrating in growth and developmental processes. Indole compounds promote root branching and root architecture, stimulate fruit formation and enhance plant resistance to stress (127) .

In addition to plants, various insect classes exhibit the ability to recognize indole, such as the grey-backed mining bee (*Andrena vaga*), hummingbird moth (*Hyles lineata)* and the housefly *(Musca domestica)* (128, 129). In these insects, indole has been observed to produce significant reactions in the antennae during foraging and possibly during pollination.

An additional example of indole's enigmatic behaviour is its role in *Caenorhabditis elegans.* Indole-producing bacteria influence the egg-laying behaviour of *C. elegans*, with indole derivatives modulating chemotaxis, egg-laying behaviour and stress responses (130). At elevated concentrations, indole has the potential to be lethal to *C. elegans,* prompting *C. elegans* to employ detoxification mechanisms involving oxidation and glucosylation in response (130).

Clearly, indole and its derivatives are pervasive in both prokaryotic and eukaryotic communities, yet there is still much to uncover regarding this signalling molecule.

1.2.2 Current understanding of indole in E. coli

A diverse array of both Gram-negative and Gram-positive bacteria are known to synthesize indole as an intracellular signal within microbial communities. Historical reports from 1897 revealed that *Bacillus coli (E. coli)* and Asiatic cholera (*Vibrio cholerae*) cells produced indole during the stationary growth phase (131). Since this report, the indole test has been a frequently employed diagnostic marker for the identification of *E. coli* (131, 132). IMViC reactions are four biochemical tests used to differentiate and identify members of the *Enterobacteriaceae* family, particularly *E.* coli. One of these tests is the indole test, which detects the production of indole by bacteria. This test utilizes Kovac's reagent, a biochemical reagent employed to detect indole production. When indole reacts with the p-dimethylaminobenzaldehyde in Kovacs reagent, it results in the formation of a red complex (133). This reaction is used to confirm the presence of indole in bacterial cultures and is often part of the plethora of tests used to differentiate and identify bacterial species.

Indole is most widely and well-studied in *E. coli* where it plays a significant role in the bacterium. In *E. coli,* indole is synthesized from tryptophan via the enzymatic activity of tryptophanase (TnaA), resulting in the simultaneous generation of pyruvate and ammonia (134). Tryptophanase is encoded by the *tnaCAB* operon, and its expression and regulation are subject to control mechanisms including catabolite repression and transcription anti-termination (135, 136). Transcription of the tryptophanase operon commences from a CAP-dependent promoter. Activation of CAP, and up-regulation of transcription, occurs in response to increased cAMP concentration (137). The advancement of transcription into the structural gene region requires the presence of external tryptophan, which is detected through a *tnaC*-dependent mechanism. This mechanism prevents Rho-dependent termination before RNA polymerase reaches the *tnaA* and *tnaB* genes (138, 139).

Indole is capable of diffusing through the lipid membrane (140), however other proteins may be involved in transport, such as the tryptophan transporter Mtr. The Mtr-mediated

transport of tryptophan is modulated by the composition of the growth medium, and there are indications that Mtr may also facilitate the transport of indole (141-143).

Indole increases the survival of bacteria subjected to antibiotic stress, and the production of indole is stimulated upon exposure to various antibiotics, including ampicillin and kanamycin (144). Indole protects bacteria from the effects of antibiotics by inducing multidrug exporter genes (145). Indole may additionally protect bacterial cells by acting as an ionophore. The ionophoric properties result in a reduction of the proton motive force across the *E. coli* cytoplasmic membrane, thereby impeding cell division (146). At indole concentrations of 3-5 mM, there is severe inhibition of growth and cell division. In a study by Chimerel and colleagues (146), evidence is presented that the cell membrane serves as a direct target of indole action, conferring advantages to the indole-producing cells.

In *E. coli*, recent studies have unveiled two distinct modes of indole signalling and both have been implicated in stress responses (147). Indole signalling operates through one of two kinetic modes: a long lasting but low-level (persistent) signal, or a transient, high-level (pulse) signal.

1.2.3 Indole persistent signalling

E. coli cultured in LB medium and allowed to grow until stationary phase accumulates approximately 0.5 mM indole in the growth medium. In persistent signalling, the indole endures in the culture supernatant for an extended period at this relatively low concentration. Under these conditions, indole has been demonstrated to exert regulatory control over the formation of biofilms (148, 149), virulence (119) and responses to various stresses including antibiotics (150).

A substantial body of literature exists regarding persistent signalling, and at times it can be intricate, perplexing, and marked by contradictions when it comes to learning about the effects of low levels of indole signalling in *E. coli* physiology.

1.2.3.1 Indole persistent signalling phenotypes

1.2.3.1.1 The role of indole in acid and heat resistance

Optimal conditions for the growth and survival of *E. coli* are typically around neutral pH. In contrast, while the stomach's highly acidic environment, with a pH of 2 or less, is lethal to most bacteria (151), *E. coli* is notable for its ability to tolerate acidity (152). Among *E. coli*'s acid resistance systems, the glutamate decarboxylase (GAD) system stands out as the most effective. Indole has been shown to significantly enhance the survival rates of *E. coli* when exposed to acidic stress (153). In their study, pre-incubation of *E. coli* with 1-2 mM indole led to the upregulation of acid resistance genes linked to the GAD system. Consequently, this heightened gene expression resulted in a significant improvement in the percentage of cell survival when confronted with acidic conditions. However, it's important to note that the precise relationship between acid resistance and indole remains a subject of ongoing research, and our understanding of their complex relationships is not yet complete. In apparent contradiction, a recent study demonstrated that indole and tryptophan supplementation inhibits the development of the adaptive response 2 (AR2) system in wild-type *E. coli*, compared to a *tnaA* mutant (154). The AR2 (Adaptive Response 2) system in *E. coli* refers to a set of genes and regulatory mechanisms that enable the bacterium to survive under acidic conditions. The AR2 system is one of several acid resistance systems in this bacterium, which are crucial for the bacterium's survival in highly acidic environments, such as the stomach. Researchers demonstrated that indole functions to hinder the pH-triggered activation of several genes in AR2. It is probable from this study, that indole binds to the EvgS protein, which plays an important role in initiating the AR2 system and conferring resistance against acidic stress at a pH 2.5 (154).

These contrasting results demonstrate that indole can affect acid resistance in *E. coli* through various means and via diverse channels. However, in these specific studies, the authors used different types of media (LB versus M9) and strains with substitutions in the EvgS system. These means they were examining different aspects of acid resistance. These variations highlight the challenges of studying indole signalling, as the type of media or strain used can significantly affect indole production.

LB culture medium is a complex medium primarily used for the growth of bacteria, known for permitting fast growth and high yields. It contains tryptone, yeast extract, and NaCl.

Tryptone is a pancreatic digest of casein, providing oligopeptides as the main carbon source, while yeast extract is derived from autolyzed yeast. Variability between batches and manufacturers arises due to the complex biological origins of these components (155).

In contrast, minimal media is a defined medium with specific compositions based on the microorganism, typically using glucose as the carbon source along with inorganic salts. Minimal media allows for reproducible bacterial experiments but is less suitable for indole signaling studies because glucose inhibits tryptophanase expression, essential for indole production. This limitation can lead to differences in experimental outcomes when comparing studies using different media types, particularly in indole phenotype observations.

The ability of *E. coli* to withstand high temperatures poses an increasing challenge to the food industry (156). Several scientific reports have indicated that indole may be involved in a mechanistic pathway contributing to enhanced survival under elevated temperature conditions (157). In apparent support of this hypothesis, when an indole-negative *E. coli* strain (*ΔtnaA*) was subjected to heat stress during the exponential growth phase and subsequently supplemented with relatively low concentrations of indole (20 μM), there was a significant increase in the proportion of surviving bacteria (158). Exponential cells exhibit significantly higher sensitivity to heat stress when compared to their stationary counterparts. After subjecting *E. coli* cells to a 50°C incubation for 24 hours, it was observed that wild-type cells displayed a remarkable tenfold increase in their survival compared to the *ΔtnaA* mutant strain. During this incubation period, the cells accumulated an approximate concentration of 20 μM indole, which was synthesized by the TnaA already present in the cells at the time of temperature shift. To further substantiate this claim, the addition of 20 μM indole had a substantial impact on the survival of the mutant strain, rendering it indistinguishable from the wild-type cells in terms of survival rates. This observation suggests that a low level of persistent signalling found in the culture supernatant is sufficient to account for the greater survival of exponentially growing wild-type cells under the stressful conditions of a 50°C environment (158).

However, once again illustrating the perplexing behaviour of indole, a more recent study reached a contrasting conclusion (159). In this study, it was observed that the *ΔtnaA* mutants exhibited a tenfold increase in resistance to heat stress compared to the WT cells when exposed to conditions of 60°C for 10 min. Additionally, the study revealed that both wild-type and mutant strains experienced a significant reduction in heat tolerance following a brief pre-treatment of just 10 minutes with indole and its derivatives. It's noteworthy, however, that these effects were observed at much higher concentrations that previous research (2.0 mM) and under different media conditions, underscoring the concentration-dependent nature of indole's impact on heat tolerance and acid resistance.

1.2.3.1.2 The role of indole in response to antibiotic treatment

Indole has been shown to augment bacterial tolerance to certain antibiotics by influencing the expression of efflux pumps. Efflux pumps are specialized proteins responsible for actively pumping antibiotics out of bacterial cells, thereby reducing the intracellular concentration of these drugs and diminishing their effectiveness. Indole has the capacity to either upregulate or downregulate the activity of these efflux pumps, thereby impacting the susceptibility of bacteria to antibiotics. Indole has been shown to induce the expression of multidrug exporters, such as *acrD, acrE, cusB, mdtE, yceL* and *emrK* (145), to increase antibiotic tolerance of *E. coli*. The study conducted by Han and colleagues, revealed that *E. coli* produces significantly more extracellular indole/cell density when treated with ampicillin (2 μg mL^{-1}) and kanamycin (2 μg mL^{-1}) (144). Additionally, the study demonstrated that the supplementation of indole (1 mM) in cultures of both *E. coli* BW25113 wild-type and *ΔtnaA* strains led to a 2- to 5- fold increase in cell growth within the first 8 hours when ampicillin was present. This study concluded that the presence of indole has a significant positive impact on the proliferation of *E. coli* cells, particularly when challenged with ampicillin (144).

There is an ongoing debate within the scientific community regarding the role of indole and tryptophan metabolism in antibiotic persistence. Bacterial persistence is the phenomenon where genetically sensitive cells survive in the presence of high concentrations of antibiotics. Persisters constitute a minor sub-population of cells within

a culture (>1%). Their persistence is physiological adaptation rather than a genetic one. Descendant cells emerging from persisters upon the removal of antibiotic display the same antibiotic sensitivity as the original population (160). Some studies have reported that indole can lead to an increase in the fraction of persister cells (142, 161) while others suggest the opposite phenotype (162, 163). These conflicting findings in studies might be due to the complex and multifaceted effects that tryptophan metabolism has on bacterial metabolism and pH regulation. Further investigation is needed to fully understand the underlying mechanisms and implications of inhibiting tryptophan metabolism.

A recent study by Goode and colleagues combined microfluidics, microscopy and an intracellular reporter plasmid to investigate bacterial pH in individual cells before and after ampicillin treatment (164). That this microfluidic approach does not introduce cellular stress, was shown by the consistent doubling time and fraction of persisters compared to a flask culture (165, 166). The researchers demonstrated that, before ampicillin treatment, *E. coli* cells that would subsequently be identified as antibiotic persisters had a notably lower intracellular pH than cells that would in due course be susceptible to the antibiotic. This appeared to be linked to the production of tryptophanase (164) because, in a Δ*tnaA* strain, the intracellular pH of both persister and susceptible cells was markedly elevated compared to WT *E. coli.* However, indole supplementation, whether mimicking the pulse or persistent signalling, failed to restore the regulation of intracellular pH in the Δ*tnaA* strain.

While quantifying persisters during exponential growth, Goode and colleagues noted a greater proportion in the WT strain compared to the Δ*tnaA* strain (164). However, during the stationary growth phase, this phenomenon was reversed, and a higher proportion of persisters were discovered in the Δ*tnaA* than in the WT strain. Interestingly, mimicking persistent indole signaling by adding indole to the culture medium restored the persister levels in the mutant to those observed in the WT strain.

1.2.3.1.3 The paradoxes of indole persistent signalling and biofilm formation

Biofilms are bacterial populations in which cells embedded in a secreted extracellular matrix, adhere both to one another and to surfaces or interfaces. These communities

exhibit remarkable resilience to host immune responses, leading to the establishment of persistent infections that demonstrate notable resistance to therapeutic interventions involving antibiotic agents (167).

The impact of indole on biofilm formation has been documented across a broad spectrum of both indole-producing and non-producing bacterial strains. However, a consensus regarding its effect has yet to be reached (168). In *E. coli*, initial reports suggested that indole inhibited biofilm formation, likely through the activation of stress response and stationary phase regulatory pathways (168-170). In the case of *E. coli*, both an indole-negative mutant and a wild-type strain treated with a chemical inhibitor of tryptophanase exhibited reduced biofilm formation compared to the wild-type strain. The addition of indole to these cultures was found to restore the capacity for biofilm formation (149, 171). An overview of indole's contradictory roles in biofilms is presented in a recent review (105). These studies indicate that the involvement of indole in biofilm formation is likely strain and concentration specific.

1.2.4 Indole pulse signalling

Indole pulse signalling was first described in 2014, it is defined by intracellular indole transiently at a much higher concentration that in the culture supernatant (147). The indole pulse occurred during entry into the stationary phase, when the intracellular indole concentration reached 60 mM for a duration of 10-20 min. Following the pulse, the cell-associated indole concentration decreased and eventually equilibrated at approximately 15 mM. This transient, high concentration can be explained by the rate of indole production temporarily outpacing the rate at which indole dissipates from the cells by diffusion. The cell-associated concentration (15 mM) that remains after the pulse is significantly higher than the concentration present in the growth medium. This difference arises from the fact that indole exhibits an approximately 90-fold higher affinity for lipids compared to water, resulting in its concentration within the cell membrane (172). The pulse promotes the transition from exponential to stationary phase before the nutrients are completely exhausted, with benefits for long-term survival (147, 172).

1.2.4.1 Indole pulse signalling phenotypes

1.2.4.1.1 The indole pulse generates persisters in fluoroquinolone treatment

A prime illustration of indole pulse signalling as a response to external stress is its effect on the production of persister cells. As previously mentioned, the addition of 4-5 mM indole to an actively dividing *E. coli* population in a broth culture inhibits growth and cell division (111, 173). The cells maintain their viability, and upon removing indole, growth resumes. This parallels the natural mechanism employed by bacterial cells to evade death during antibiotic exposure, resulting in the formation of persister cells (174). Recent research has suggested that indole treatment drives the formation of persister cells in *E. coli* through an interaction with the antibiotic target DNA gyrase, a bacterial type II topoisomerase (175). DNA gyrase plays a pivotal role in overseeing the supercoiling and topological organization of bacterial DNA. The mechanism of action for quinolone antibiotics involves inhibiting or interfering with the function of DNA gyrase (176, 177). The indole concentration that is reached during pulse signalling is sufficient to inhibit supercoiling by DNA gyrase *in vitro* (178). Indole's inhibition of DNA gyrase occurs independently of the concurrent accumulation of double stranded breaks, resembling the mechanism of action seen with aminocoumarin antibiotics. This observation implies an interaction between indole and one of the subunits of gyrase (179).

A recent study, researchers compared the proportion of persister cells in cultures of *E. coli* BW25113 (WT) with indole-negative (*tnaA* knock-out) cultures of *E. coli* BW25113 following exposure to a range of antibiotics designed to target DNA gyrase (175). The absence of indole production led to a reduction in the number of persister cells following quinolone treatment, with ciprofloxacin demonstrating the most pronounced effect. The restoration of persister levels to wild type was accomplished through an experimentally applied indole pulse mimicking the concentration and timing of an indole pulse. Further evidence was provided by *in silico* docking simulations of indole with DNA gyrase that indicated indole occupies the ATP binding site of GyrB. This suggests that indole treatment interferes with the stabilization of double-stranded DNA breaks and induces a state of cellular quiescence by inhibiting gyrase activity (105).

1.2.4.1.2 The indole pulse regulates cellular pH

Indole serves as a proton ionophore, and the pulse is proposed to diminish the electrical potential across the cytoplasmic membrane. This, in turn, hinders bacterial growth and cell division (146). Another related consequence of indole pulse signalling is its impact on the intracellular pH of bacterial cells (180). *E. coli* cells in which the indole pulse occurs sustain their cytoplasmic pH at approx. 7.2. However, in the absence of indole production, these bacterial cells show an elevated cytoplasmic pH of 7.8. Upon the application of a synthetic pulse, experimentally administered to *ΔtnaA* cells to mimic the conditions of a typical pulse, the normal pH was restored. A similar result was achieved with the use of non-indole proton ionophores, suggesting that proton transport across the cytoplasmic membrane was influencing pH regulation. The authors proposed the presence of cellular "memory" that maintains the cytoplasmic pH, long after a stationary phase indole pulse has taken place. This is somewhat analogous to chemosensing, a process where a memory of changes in external stimuli is retained by adjusting the methylation of chemoreceptors, creating a working memory of the stimulus (181). *E coli* has previously been documented as possessing the ability to retain a memory of the timing of stationary phase entry. Subsequently, this memory influences the timing of resumption of growth upon favorable changes in culture conditions (182). The regulation of cytoplasmic pH through indole pulse signalling could yield significant implications for bacterial stress responses, including its ability to survive antibiotic treatment.

1.2.4.1.3 The indole pulse contributes to long-term cell survival

For wild-type *E. coli* cells in LB medium, the indole pulse causes cells to enter stationary phase before nutrients are exhausted (172). This correlates with heightened survival during prolonged periods of stationary phase, presumably because the cells have a greater reservoir of resources to use for long-term survival. In *ΔtnaA cell*s, where the tryptophanase gene is deleted and indole production is lost, cultures reach a higher density before entering stationary phase. These cultures also display diminished long-term survival, presumably because they have fewer resources available for repair and maintenance processes in stationary phase.

1.2.5 Considerations for interpreting indole assays

Persistent signalling reflects the indole present in the culture environment and is experienced by both producer and non-producer cells. In contrast, pulse signalling is experienced exclusively by the producer cells and only for a limited time. Being an intra-cellular phenomenon of limited duration, the indole pulse is considerably more difficult to assay than persistent indole. Until now, pulse signalling has only been detected at the level of the cell culture. It occurs at a predictable time (during entry into stationary phase) and is of approx. 20 min duration. However, it is not known whether all cells in a culture pulse simultaneously, or at varying times throughout the 20 min window.

When interpreting the results of indole assays, it is important to remember that when extra-cellular and cell-associated indole exist in equilibrium, they will be at very different concentrations. This is because the lipid-to-water partition coefficient for indole is approximately 100 so indole partitions selectively into a non-polar environment (147). Consequently, in a bacterial culture, indole shows greater affinity for cells than for a culture medium. The concentration within the cells is non-uniform since indole shows a preference for lipid membranes over the cytoplasm. It has been determined experimentally that when indole is in equilibrium between cells and the culture medium, the cell-associated indole concentration (i.e. the concentration of indole in the cell pellet) is approx. 16-fold greater than the indole concentration in the growth medium (147). However, in a non-equilibrium situation, such as during the pulse of indole synthesis during the transition from exponential to stationary phase, the cell-associated concentration can exceed the concentration in the culture medium by up to 80-fold.

1.3 An overview of the enzyme: tryptophanase

As previously mentioned, tryptophanase is an enzyme responsible for the degradation of L-tryptophan to indole, ammonia and pyruvate. It also has the capacity to act in reverse to synthesize L-tryptophan when there are elevated concentrations of both pyruvate and ammonia (183), but it is heavily thermodynamically favoured to produce indole (184). Tryptophanase relies on pyridoxal 5′-phosphate (PLP) as a cofactor, and the presence of monovalent cations (K^+ or NH_4^+) is essential for facilitating the binding of PLP to a lysine

residue at position 270 within the enzyme's active site, thereby making it functionally active (185, 186). Tryptophanase in *E. coli* functions as a tetramer of four identical 52.7 kDa subunits and each monomer binds one molecule of PLP. The disassociation of tryptophanase into its monomeric and dimeric forms makes it inactive, and the tetrameric state is necessary for enzymatic activity. Finally it is important to note that, with the exception of the reverse tryptophanase reaction, *E. coli* does not encode an enzyme for the degradation of indole (173).

1.3.1 Structure of the tryptophanase operon

Tryptophanase is encoded within the tryptophanase (*tna*) operon whose expression is under a series of rigorous regulatory mechanisms which include catabolite repression and transcriptional control. These mechanisms have been well studied and characterized in the literature for over thirty years (187). The tryptophanase operon (seen in Figure 1.3) comprises two major structural genes: *tnaA,* encoding tryptophanase and *tnaB*, encoding a low affinity, tryptophan-specific permease (188, 189). The operon is under the control of specific regulatory sequences within the *E. coli* genome (190). Researchers, Deeley and Yanofsky, used *in vitro* transcription experiments to detect a transcription initiation site that precedes *tnaA* by 320 bp within the operon (137). Also upstream of *tnaA* is a 24 bp region that encodes a leader peptide called *tnaC* (191). Downstream of *tnaC,* there is a non-coding region comprising multiple Rho-dependent terminator sequences.

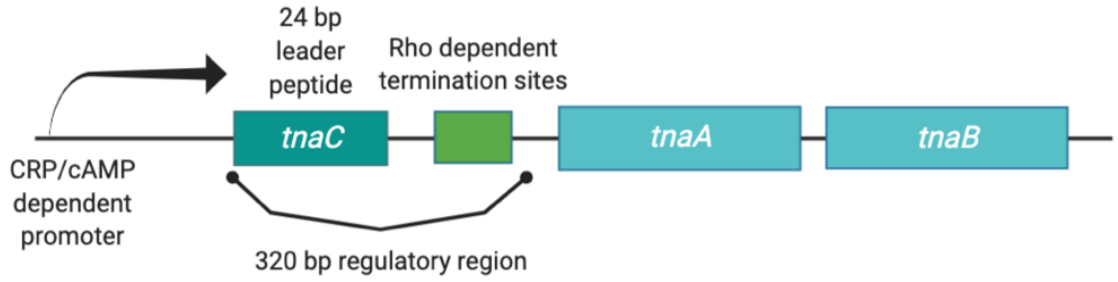


Figure 1.3. Organization of the *tna* operon of E. coli. The operon contains a CAP/cAMP dependent promoter, a 320 bp regulatory region which includes tnaC and Rho dependent termination sites. Downstream in the operon are genes encoding tryptophanase (tnaA) and a tryptophan specific permease (tnaB). Not drawn to scale. *Generated in Biorender.*

1.3.2 Tryptophanase operon regulation

1.3.2.1 Catabolite repression

Transcriptional regulation of the *tna* operon is subject to catabolite repression (192). Transcription initiation requires cAMP-dependent, CAP-facilitated binding of the RNA polymerase to the *tna* promoter (137, 193). Under conditions of limited glucose, cellular cAMP increases, initiating binding of cAMP and CAP, and leading to the formation of the active cAMP-CAP complex. Subsequently, the complex binds to the *tna* promoter, facilitating the recruitment of RNA polymerase and initiating transcription. Alternatively, in the presence of abundant glucose, the concentration of cellular cAMP decreases and cAMP-CAP is no longer formed or maintained in its active state. The binding of this complex to the *tna* promoter is impeded and transcription is prevented.

Based on prior research, it has been suggested that catabolite repression could explain the detection of substantial indole concentrations in the culture supernatant as the cells approach stationary phase, coinciding with nutrient depletion (172, 173). Moreover, the introduction of glucose to *E. coli* cultures serves as an effective way of delaying indole production (194).

1.3.2.2 Transcription anti-termination

There is substantial evidence for a positive correlation between the induction of the *tna* operon and the concentration of tryptophan in the culture medium. The level of induction increases in a dose-dependent manner, reaching its peak at 100 μg ml^{-1} (141). Transcription anti-termination controls the elongation of *tna* transcripts. In the absence of the inducer, tryptophan, transcription is halted prematurely at the Rho-dependent termination sequences (191). Conversely, when tryptophan is readily available, termination is suppressed, and allowing the *tna* transcript to proceed without attenuation at the Rho sites. This results in the production of mRNA that includes *tnaA* and *tnaB* (191).

In the case of tryptophan depletion, release factor-mediated cleavage occurs. The TnaC peptidyl-$tRNA^{Pro}$ complex is hydrolysed, and the TnaC peptide is released from the ribosome-mRNA complex (195). The spacer region of 220 nucleotides that separates *tnaC* and *tnaA* in the genome contains several transcription pause sites, which serve as regions for Rho-dependent transcription termination (136, 196). The Rho protein recognizes and binds to the Rho utilization (*rut*) site on the nascent RNA molecule. Once this binding takes place, Rho interacts with RNA polymerase, causing it to dissociate from the DNA template and resulting in transcriptional termination (195).

In the presence of abundant tryptophan levels, tryptophan binds near the ribosomal A site, obstructing the activity of TnaC-$tRNA^{Pro}$. The TnaC leader peptide adopts an alternative structure, exposing a short open reading frame (188). As the ribosome translates the *tna* transcript, it encounters the *tnaC* stop codon, leading to a stalling event. This stalling impedes the Rho-factor's access to its rut RNA-binding site, located just after the *tnaC* stop codon, thereby preventing Rho-catalyzed transcription termination. Due to the stalling of the ribosome, the translation of the TnaC peptide is terminated, and the truncated TnaC peptide interacts with the ribosome and triggers the expression of a downstream transcriptional antiterminator, allowing transcription and eventually expression of *tnaA* and *tnaB* (197, 198).

In the absence of the *tnaC* gene, represented by *ΔtnaC*, the regulation of the operon would be significantly impacted, and it is hypothesized that indole production would stop altogether. This assumption forms the basis of a hypothesis that will be explored in subsequent chapters. Specifically, the role of specific genes in the operon by examining genetic strains lacking *tnaA, tnaB* or *tnaC* in their ability to produce indole. Through the experimental approach, it is acknowledged the possibility of leaky expression stemming from basal transcription at the promoter, stochastic fluctuations in gene expression, and indirect regulation by other environmental or cellular signals (199). These factors may influence indole production independently of regulatory genes. By integrating the concept of leaky expression into our understanding of *tnaCAB* transcriptional regulation, we recognize the complexity of gene expression dynamics and the potential for subtle regulatory mechanism to influence indole synthesis.

1.3.2.3 The RpoS and tryptophanase relationship

There is a growing body of evidence suggesting the regulation of tryptophanase is intricate and not yet fully understood. The stationary phase sigma factor, σ^S (*rpoS*), is essential for normal expression of TnaA (200) and another transcription factor, MarA, can also increase TnaA expression (201). The regulation of tryptophanase by RpoS, the stationary phase sigma factor, has been characterized in *E. coli*. In response to specific stress conditions, RpoS can control the regulation of indole production through the tryptophanase enzyme (200). In an RpoS null mutant, tryptophanase mRNA levels are reduced approximately 14-fold compared to the wild-type strain. Indole production is severely impacted by inactivation of the *rpoS* gene and results in a 4-fold decrease in indole concentration. This correlated with a greater than 10-fold reduction in tryptophanase mRNA levels in the RpoS null mutant in comparison to the wild-type strain. In a thesis published by this laboratory, RpoS mutants exhibited reduced production of indole during the transition into stationary phase in comparison to the wild-type strain.

1.3.2.4 Evidence for post-translational control of tryptophanase

While not as extensively studied as other forms of tryptophanase regulation, there is evidence for the existence of post-translational regulation of tryptophanase. Dimers of

plasmid ColE1 produce an untranslated RNA called Rcd (202); part of a checkpoint that is activated by the presence of plasmid dimers, to prevent a dimer catastrophe (173, 203, 204). Using RNA affinity chromatography, researchers established that tryptophanase is the target of Rcd, resulting in a stimulation of indole production by TnaA (173). In the presence of Rcd, the affinity of tryptophanase for tryptophan increases approximately five-fold, leading to increased indole production. It is proposed that this produces an indole pulse in the dimer-containing cell, delaying its division (205). Although Rcd is absent in non-plasmid containing bacterial cells, the enhancement of tryptophanase activity in the presence of Rcd provides the best evidence of post-translational tryptophanase regulation.

In plasmid-free cells, post-translational tryptophanase regulation has been less well characterized. However, compelling evidence emerged in 2014, shedding light on this matter (206, 207). The researchers were studying the presence of exogenous cAMP and its relationship to *tnaA* production and activity. They initially hypothesized that the addition of carbon sources to cells with a localized focus would change their behaviour. The researchers introduced sugars into mid-exponential cultures and measured *tnaA* expression, as well as localization and activity of TnaA. As expected, they observed a cessation of *tnaA* expression in the presence of glucose, consistent with catabolite repression, but not in the presence of glycerol. However, the most intriguing finding was the cessation of indole production in the presence of both glucose and arabinose, in contrast to the presence of glycerol.

Furthermore, when cells were incubated with either arabinose or glucose, the TnaA-GFP protein diffused throughout the cytoplasm, contrasting with the tightly clustered foci observed under normal conditions. In stark contrast, normal foci formation was observed in the presence of glycerol (207). These results pointed towards the possibility of a previously unrecognized post-translational mechanism influenced by the presence or absence of sugars. Further, this observation highlights the differentiation of the metabolic pathways in arabinose and glucose to glycerol and suggests the potential involvement of glycolysis in post-translational regulation of tryptophanase.

Another set of researchers explored the possibility of post-translational regulation. They added late stationary cells containing abundant tryptophanase into tryptophan-containing PBS buffer or LB (158). The cells in the tryptophan-containing PBS buffer produced ample amounts of indole, while the cells added to LB produced little indole. This is consistent with the observations seen by Li and Young that arabinose and glucose can either indirectly, or by an ill-defined mechanism, inhibit the activity of tryptophanase (207).

1.3.2.5 Heterogeneity in tryptophanase expression and how to measure it

Liu and Summers investigated the TnaA-GFP content of single cells using flow cytometry (158). The authors observed considerable TnaA-GFP heterogeneity among exponential and stationary phase cells. This contrasted with greater homogeneity during the transition from exponential growth to stationary phase. In stationary phase, TnaA-GFP content gave a bimodal distribution. The population of cells fell into two discernible subgroups, one with lower fluorescence and the other exhibiting a 10-fold higher fluorescence intensity (158). The mechanism responsible for the emergence of this distribution, and precise reasons for the existence of two distinct groups within a single culture, remains unexplained and warrants further investigation.

The heterogeneity of tryptophanase content had been missed in previous investigations due to the reliance on bulk culture experiments in which variability within the population was effectively averaged out. In contrast, the use of flow cytometry allows the assessment of individual cells within a population. This allows for the precise identification and analysis of nuanced differences that exist between and among cells within the same population. Moreover, flow cell cytometry potentially allows for the detection of rare cell types such as persister cells (208), which otherwise remain concealed or overshadowed in bulk culture experiments. Although flow cell cytometry enables high-throughput cell measurements, it lacks the capability to continuously monitor cells throughout their growth.

To address this limitation and further delve into the intricacies of cellular heterogeneity and dynamics in microbial populations, researchers have increasingly turned to microfluidic single-cell technologies. The utilization of microfluidic systems offers a

promising avenue for studying individual cellular behaviour over time, making it a coveted technology in recent decades. This approach holds the potential to unlock a deeper understanding of cellular individuality at a single-cell level, shedding light on previously elusive aspects of bacterial signalling (209).

1.4 Aims of this work

The primary objective of this study was to conduct a more comprehensive examination of the indole pulse and to explore how tryptophanase regulation plays a role in it. In particular, the research sought to understand the mechanism behind the indole pulse, exploring its connection to the *de novo* transcription and potential post-translational activation of tryptophanase during the transition from exponential to stationary phase.

Initial evidence from the indole pulse assays revealed heterogeneity in indole production, suggesting potential heterogeneity in TnaA production. This heterogeneity was examined in detail, utilizing a combination of conventional and innovative microbiological techniques. In addition, the involvement of post-translational TnaA regulation was explored.

Lastly, because of the COVID-pandemic, laboratory access was restricted for a substantial portion of this study, prompting the adoption of computational research methods. This involved an exploration of tryptophanase distribution and evolutionary history in *E. coli* and a wide range of microbial species that make up the human gut microbiome.

2. Methods and Materials

2.1 Strains

Table 2.1: List of bacterial strains used in this work.

Species	Name of Strain	Genotype/Description	Source/Reference
E. coli	BW25113	*lacI^{q}rrnB$_{T14}$ ΔlacZ$_{WJ16}$hsdR514 ΔaraBAD$_{AH33}$ΔrhaBAD$_{LD78}$*	Laboratory Stock (210)
E. coli	BW25113 Δ*tnaA:: kan^{R}*	*BW25113 ΔtnaA:: kan^{R}* Tryptophanase knockout, Kanamycin resistant.	Laboratory Stock (210)
E. coli	BW25113 Δ*tnaA*	*BW25113 ΔtnaA* Tryptophanase knockout	(147)
E. coli	BW25113 tnaA–*GFP: kan^{R}*	*lacI^{q}rrnB$_{T14}$ ΔlacZ$_{WJ16}$hsdR514 ΔaraBAD$_{AH33}$ΔrhaBAD$_{LD78}$* *tnaA-GFP: kan^{R}* introduced from *MG1655 tnaA-GFP*	Hannah Gaimster Thesis, 2014 (147)
E. coli	BW25113 Δ*tnaB:: kan^{R}*	*E. coli K12 JW1253ΔtnaB:: kan^{R}* Tryptophan permease gene knockout, Kanamycin resistant.	(211)
E. coli	BW25113 Δ*tnaC:: kan^{R}*	*E. coli K12 JW1254 ΔtnaC:: kan^{R}* Leader peptide gene knockout, Kanamycin resistant.	(211)
E. coli	BW25113 Δ*tnaCAB:: kan^{R}*	*BW25113 ΔtnaCAB:: kan^{R}* Tryptophanase operon knockout, Kanamycin resistant. *kan^{R}* introduced from BW25113 Δ*tnaA:: kan^{R}*	This work
E. coli	NEB	Chemically Competent Cells	New England Biolabs (NEB)
E. coli	ATCC® 25922™	O6, Biotype 1	FDA strain Seattle 1946 [DSM 1103, NCIB 12210]
E. coli	NEB 5-alpha Competent E. coli (High Efficiency)	*DH5α^{TM}* derivative T1 phage resistant and *endA* deficient for high quality plasmid preparations	C2987, NEB

E. coli	*MI13*	*MG1655 7740 ΔmotA::PrpsL-mCherry* Using *pGRG37 (attR Tn7* integration vector); next to *glmS* into *attTn7*	Laboratory of Dr. Somenath Bakshi
E. coli	*MI13 tnaA-GFP :: kan^R*	*MI13 tnaA-GFP :: kan^{R}tnaA—GFP: Km^R* introduced from *MG1655 tnaA–GFP*	This work.

2.2 Plasmids

Table 2.2: List of plasmids used in this work.

Name of Plasmid	**Genotype/Description**	**Source/Reference**
pSLTS	*Ori SC101 (Ts) Amp^R; P$_{araB}$ for λ-Red; P$_{tetR}$ for I-SceI* Expresses the lambda Red recombinase under the control of an arabinose inducible promoter P(*araB*) Temperature sensitive replication Ampicillin resistant.	(212)
pBAD TOPO^{TM}	Arabinose inducible promoter P(*araB*) in *E. coli*, drives the control of the gene of interest in the presence of L-arabinose. TOPO® cloning system Ampicillin resistant.	Invitrogen
pEK_1	pBAD TOPO: *tnaA* Contains the tryptophanase gene P(*araB*)	This work.
pEK_2	pBAD TOPO: *tnaAB* Contains the tryptophanase gene and tryptophan permease gene P(*araB*)	This work.
pEK_3	pBAD TOPO: *tnaCAB* Contains the tryptophanase operon P(*araB*)	This work.
pXY027	IPTG inducible promoter pCA24N-FtsZ-GFP backbone	A gift from Jie Xiao (Addgene plasmid # 98915 ; http://n2t.net/addgene:98915 ; RRID:Addgene_98915)

2.3 Media

2.3.1 Luria broth base (LB)

Supplied by ForMedium™

Solution was autoclaved for 60 min at 121°C before use.

Table 2.3: Components of LB medium.

Component	**Concentration/ gL^{-1}**
Tryptone	10
Yeast Extract	5
NaCl	0.5

2.3.2 Luria broth base with agar

1.5% agar was added to LB medium

Solution was autoclaved at 121°C for 20 minutes before use.

2.3.3 Supplemented LB medium

Additional tryptophan was added to the media.

2.3.4 Super Optimal Broth with Catabolite repression (SOC)

Table 2.4: Components of S.O.C medium.

Component	**Final Concentration**
Tryptone	2%
Yeast Extract	0.5%
NaCl	10 mM
KCl	2.5 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM
Glucose	20 mM

Medium was obtained as 10 x 10 ml bottles of liquid medium from ThermoFisher Scientific and stored at 4°C until use.

2.4 Antibiotics

Table 2.5: List of antibiotics used in this work.

Antibiotic	Stock concentration and solvent	Working concentration	Supplier
Carbenicillin	100 mg ml^{-1} in distilled water	100 µg ml^{-1}	ThermoFisher
Kanamycin	50 mg ml^{-1} in distilled water	50 µg ml^{-1}	ThermoFisher

All antibiotics were filter sterilized using a 0.2 µm filter before use, aliquoted and stored at -20°C.

2.5 Chemicals

Table 2.6: List of chemicals used in this work.

Chemical	Stock concentration and solvent	Supplier
Indole	50 mM; H_2O	
Tryptophan	50 mM; H_2O	
Quantichrom™ Indole Assay Kit	N/A	Universal Biologics
Pluronic Fluid	10%; H_2O	

Indole was added to a Falcon™ conical tube covered in aluminum foil to protect from light and stored at 4°C.

Tryptophan was filter sterilized using a 0.2 µm filter before use, wrapped in an aluminum foil-covered Falcon™ conical tube to protect from light and stored at 4°C.

Quantichrom™ reagent was obtained from the supplier and stored at 4°C in a light-protected container.

2.6 Microbiological techniques

2.6.1 Stock maintenance

For short term storage, strains were streaked to single colonies on LB agar plates containing the appropriate antibiotics. The bacteria were left to grow overnight (~16-18 h) in a standing incubator set at 37°C, or 30°C for temperature sensitive plasmids. The plates were then sealed with parafilm and stored at 4°C.

For long term storage, 0.9 ml of an overnight culture was mixed with 0.9 ml of 50% sterile glycerol in a 1.8 ml cryotube. The cultures were then stored at -80°C.

2.6.2 Growth curves

Growth medium (5 ml), with appropriate antibiotic, was inoculated with a single bacterial colony from an agar plate. The culture was left to grow at 37°C at 150 rpm overnight. Overnight cultures were diluted to an OD_{600} of 0.05 in the appropriate medium and flask and left to incubate shaking at 150 rpm at 37°C. Growth was measured by recording optical density at 600 nm (OD_{600}) of the cultures over time, using a Gene Quant 1300, GE Spectrophotometer.

However, it is important to note that OD_{600} is not the most accurate measurement of cell concentration, as it does not differentiate between viable cells or dead cells and particles. As well, OD_{600} does not consider the morphological changes that occur to cells during growth such as size. However, the Mother Machine data explored in later chapters allows visualization of cells throughout growth which gives more confidence in our assumptions.

Optical density (600 nm) relies on absorbance and values over 1 are outside the dynamic range of spectrophotometers. This was overcome in these experiments by take 100 µL

samples and diluting them 10X in PBS. This allowed for more accurate measurements of OD_{600}.

2.6.3 Making electrocompetent cells

LB, lysogeny broth, (5 ml) and the appropriate antibiotic was inoculated with a single bacterial colony and incubated overnight with shaking at 150 rpm at 37°C, or 30°C for temperature sensitive strains/plasmids (Table 2.2). A sample of overnight culture (1 ml) was added to 100 ml of LB medium in a 500 ml flask. The culture was left to grow, shaking at 150 rpm at 37°C, or 30°C if temperature sensitive, until it reached an OD_{600} of 0.7-0.9 (~2-3 h). The cells were harvested by centrifugation (20 minutes, 4000 x g at 4°C, Eppendorf Centrifuge 5810 R). The supernatant was discarded, and the pellet was resuspended in 100 ml of 10% ice cold sterile glycerol. This wash step was repeated four times, until the final pellet was resuspended in 100 µl of ice-cold sterile glycerol, to make a stock of one hundred times the initial concentration. Cells (50 µl) were aliquoted in cryotubes and stored at -80°C until use.

2.6.4 Electroporation

Aliquots of the electrocompetent cells were placed on ice to thaw, 50 µl of cells were transferred to an electroporation cuvette. Plasmid DNA or DNA cassettes (50-100 ng) were added to the cells in the electroporation cuvettes and mixed well. The mixture was subjected to a single pulse of 1.68 kV (25 µFd, 200 Ω) with a Gene Pulser electroporator (Biorad). 500 µl of S.O.C medium was added, and the mixture transferred to a 1 ml Eppendorf tube. The mixture was left for 1 hour with shaking at 150 rpm at 37°C, or 30°C for temperature sensitive strains/plasmids. Appropriate dilutions were spread on LB plates containing appropriate antibiotics.

2.6.5 Chemical transformation

NEB 5-alpha Competent *E. coli* (High Efficiency) cells were placed on ice to thaw for 10 minutes. Gibson assembly mix (2 µl) was added to 50 µl of the cells, and the tube was mixed. The mixture was placed on ice for 30 minutes. The cells were heat shocked at 42°C for 30 seconds, and subsequently placed on ice for 2 minutes. 950 µl of S.O.C

media was added to the cells, and this was incubated, shaking at 150 rpm, at 37°C for 1 hour. Agar plates containing appropriate antibiotics was warmed to 37°C for one hour. Appropriate dilutions were spread on LB plates containing appropriate antibiotics.

2.7 DNA assembly and manipulation

2.7.1 Plasmid purification

LB medium (5 ml) containing the appropriate antibiotic was inoculated with a single bacterial colony and incubated overnight with shaking at 150 rpm at 37ºC, or 30ºC for temperature sensitive plasmids. Plasmids were extracted according to the manufacturer's instructions using the Qiaprep Spin Miniprep Kit (Qiagen, Hilden, Germany).

2.7.2 Polymerase chain reaction (PCR)

To produce a BW25113 tryptophanase operon knockout (see 2.7.4 and 2.7.5), starting from BW25113 *ΔtnaA*, a kanamycin cassette with regions flanking the tryptophanase operon was produced. Primer designs were obtained from Reham Sharkawy (pers. comm. 2019) and ordered from ThermoFisher Scientific.

Table 2.7 Primers for production of whole *tnaCAB* operon knockout in BW25113.

Name of primer	**Sequence 5' – 3'**	**Source**
Forward primer	TCCTGTTATTCCTCAACCCTTTTTTTAAACATTAA AATTC TATGGACAGCAAGCGAACCG	Sharkawy, pers. comm. 2019
Reverse primer	ACGCAAAGCATACGTGGTGAAGGAGGATCAGG CGAGTTGC TCAGAAGAACTCGTCAAGAAG	Sharkawy, pers. comm. 2019

Table 2.8 Components of PCR for production of kanamycin cassette for BW25113 *tnaCAB* knockout.

Component	**Concentration**	**Volume (50 µl/rxn)**
Water, nuclease free	-	Make up to 50 µl
5X Platinum™ II PCR Buffer	5X	10

dNTP mix	10 mM	1
Kanamycin cassette forward primer	10 mM	1
Kanamycin cassette reverse primer	10 mM	1
Platinum™ II *Taq* Hot-Start DNA Polymerase	-	0.4

Components, excluding Taq polymerase, were added to PCR tubes on ice. Candidate colonies from BW25113 Δ*tnaA:: kan*^{R} were resuspended in the PCR mix. Taq polymerase was added to PCR tubes and mixed. Tubes were subjected to the following conditions:

Table 2.9 Conditions for PCR using Platinum II™ System.

Step		**3-step protocol**	
		Temperature	**Time**
Initial Denaturation		94ºC	2 minutes
35 PCR cycles	Denature	94ºC	15 seconds
	Anneal	60ºC	15 seconds
	Extend	68ºC	15 seconds/kb
Hold		4ºC	hold

Final PCR product was analyzed by gel electrophoresis.

Once a *⊿tnaCAB* strain had been generated in BW25113, a PCR was performed to confirm the knockout. Primer designs were obtained from Reham Sharkawy (pers. comm., 2019) and ordered from ThermoFisher Scientific.

Table 2.10 PCR primers to confirm tryptophanase operon knockout and replacement with kanamycin cassette in BW25113.

Name of primer	**Sequence 5' – 3'**	**Source**
Forward primer	TCAGATGCGGCTTCGCTTCATTGT	Sharkawy, 2019

Reverse primer	TCAAAAAGCGGGACATGGGCTAAA	Sharkawy, 2019

To produce a *tnaA-GFP::kan*^{R} cassette in *E. coli* MI13, the cassette was amplified from BW25113 *tnaA-GFP::kan*^{R}.

Table 2.11 PCR primers to create TnaA-GFP and kanamycin cassette in MI13.

Name of primer	**Sequence 5' – 3'**	**Source**
Forward primer	CCCCGAACGATTGTGATTCGATTCACATTTAAACAATTT CAGAATAGACAAAAACTCTGAGTGTAATAATGTAGC	This work
Reverse primer	TACCTGCTATAACCATAACACCCCAAAATGCAGAGTGCT TTTTTTCAGCTCAGAAGAACTCGTCAAGAAGGCGA	This work

2.7.3 Gel electrophoresis

Electrophoresis through 1% (w/v) agarose gels in 1x TAE buffer (40 mM Tris-acetate, 1 mM EDTA) was used to analyze PCR products and plasmid preps. Gels were pre-stained with Invitrogen™ SYBR™ Safe Stain (10 000X concentrate in DMSO). Samples were loaded after addition of 6X loading dye (Sigma) and gels were run at 10 volts cm^{-1}. Gels were visualized by UV transillumination.

2.7.4 Gibson Assembly®

Gibson assembly was used to generate the pBAD TOPO® plasmids containing different parts of the tryptophanase operon. pBAD TOPO®, pJMK001, was linearized using PmeI and NcoI and purified.

2.12 Primers used to generate insert DNA for Gibson Assembly® of pBAD TOPO::*tnaA.*

Name of primer	**Sequence 5' – 3'**	**Source**

Forward primer	aagaaggagatatacataccATGGAAAACTTTAAACATCTCCCTG	This work
Reverse primer	agccaagctggagaccgtttTTAAACTTCTTTAAGTTTTGCGG	This work

2.13 Primers used to generate insert DNA for Gibson Assembly® of pBAD TOPO::*tnaB.*

Name of primer	**Sequence 5' – 3'**	**Source**
Forward primer	aagaaggagatatacataccATGACTGATCAAGCTGAAAAAAA	This work
Reverse primer	agccaagctggagaccgtttTTAGCCAAATTTAGGTAACACG	This work

2.14 Primers used to generate insert DNA for Gibson Assembly® of pBAD TOPO::*tnaC.*

Name of primer	**Sequence 5' – 3'**	**Source**
Forward primer	aagaaggagatatacataccATGAATATCTTACATATATGTG	This work
Reverse primer	agccaagctggagaccgtttTCAAGGGCGGTGATCGACAA	This work

2.15 Primers used to generate insert DNA for Gibson Assembly® of pBAD TOPO::*tnaAB.*

Name of primer	**Sequence 5' – 3'**	**Source**
Forward primer	aagaaggagatatacataccATGGAAAACTTTAAACATCTCCCTG	This work
Reverse primer	agccaagctggagaccgtttTTAGCCAAATTTAGGTAACACG	This work

2.16 Primers used to generate insert DNA for Gibson Assembly® of pBAD TOPO::*tnaCAB*.

Name of primer	Sequence 5' – 3'	Source
Inser_fwd	aagaaggagatatacataccCGTGTCTTGCGAGGATAAG	This work
Inser_fwd 2	gtagtattaaTTAAACTTCTTTAAGTTTTGCGG	This work
Insert_rev 2	agaagtttaaTTAATACTACAGAGTGGCTATAAG	This work
Insert_rev	agccaagctggagaccgtttTTAGCCAAATTTAGGTAACACG	This work

PCR followed Tables 2.8 and 2.9 conditions for generation of insert DNA for Gibson Assembly® of pBAD TOPO::*tnaA*, pBAD TOPO::*tnaB*, pBAD TOPO::*tnaC*, pBAD TOPO::*tnaCAB* pBAD TOPO::*tnaAB*.

Gibson Assembly® was performed following the manufacturer's guidelines using the Gibson Assembly® Protocol (E5510), New England Biolabs.

2.17 Components of Gibson Assembly®.

	2-3 Fragment Assembly	Positive Control
Total Amount of Fragments	0.02-0.05 pmol X µl	10 µl
Gibson Assembly Master Mix (2X)	10 µl	10 µl
Deionized Water	10 – X µl	0
Total Volume	20 µl	20 µl

Following the Gibson Assembly® protocol, samples were transformed into NEB 5-alpha

Competent *E. coli* cells following the Gibson Assembly® Protocol (E5510), New England Biolabs.

2.7.5 Production of DNA knockouts/fusions

LB medium (5 ml) was inoculated with a single bacterial colony and incubated overnight with shaking at 150 rpm at 37°C. Cells were made electrocompetent (refer to 2.6.3). Aliquots were thawed, and subsequently underwent electroporation to introduce the pSLTS plasmid. The pSLTS plasmid is necessary for the production of lambda Red recombinase under the control of the arabinose inducible P(*araB*) promoter. Cells were grown on agar plates at 30°C overnight containing carbenicillin to select for positive colonies. LB media (5 ml) containing carbenicillin was inoculated with a single bacterial colony and incubated overnight with shaking at 150 rpm at 30°C. Overnight culture (1 ml) was diluted in 100 ml of LB medium and incubated for 1 hour shaking at 150 rpm in 30°C. Sterile arabinose (1 mM) was added and the cells were incubated until an OD600 of 0.7-0.9 was reached (~2-3 h). The cells were harvested by centrifugation (20 minutes, 4000 x g at 4°C, Eppendorf Centrifuge 5810 R). The supernatant was discarded, and the pellet was resuspended in 100 ml of 10% ice cold sterile glycerol. This wash step was repeated four times, until the final pellet was resuspended in 100 µl of ice-cold sterile glycerol, to make a stock of one hundred times the initial concentration. Cells (50 µl) were transferred to an electroporation cuvette. DNA cassettes (50-100 ng) were added to the cells in the electroporation cuvettes and mixed well. The mixture was subjected to a single pulse of 1.68 kV (25 µFd, 200 Ω) with a Gene Pulser electroporator (Biorad). 450 µl of LB media was added, and the mixture transferred to a 15 ml Falcon tube. The cultures were incubated for 3 h with shaking at 150 rpm at 30°C. Appropriate dilutions were spread on LB plates containing kanamycin and carbenicillin and grown overnight at 30°C. Colonies were selected and streaked onto carbenicillin plates and kanamycin plates to select for colonies without the pSLTS plasmid. Selected colonies were confirmed for correct DNA editing using PCR amplification of the edited tryptophanase operon.

2.7.6 Genomic DNA extraction

LB medium (5 ml) containing the appropriate antibiotic was inoculated with a single bacterial colony and incubated overnight with shaking at 150 rpm at 37ºC. Genomic DNA

was extracted according to the manufacturer's instructions using the DNAeasy UltraClean Microbial Kit (Qiagen, Hilden, Germany).

2.7.8 PCR Purification

PCR products were purified according to the manufacturer's instructions using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).

2.8 Biochemical Assays

2.8.1 Kovacs assay for internal indole measurements

Kovacs Solution:

10 g p-dimethylamino-benzaldehyde
150 ml iso-amyl alcohol
50 ml hydrochloric acid (analytical grade reagent)

HCl-amyl alcohol solution:

150 ml iso-amyl alcohol
50 ml hydrochloric acid (analytical grade reagent)

Both solutions were wrapped in aluminum foil to protect from light and stored at room temperature

2.8.1.1 Measurement of extracellular indole

To assay the extracellular indole, or the indole in a culture supernatant, a sample (1 ml) of culture was removed, the optical density was measured, and the sample was centrifuged to pellet cells at 1137 x g for 1 minute (Eppendorf, Minispin Plus, Hamburg, Germany). The resulting supernatant was removed and assayed. The supernatant was diluted in LB medium 1:5 for a final volume of 100 µl in a clear flat bottom 96 well-plate and mixed well. The indole measurements were estimated using the Quantichrom™ Indole Assay kit (DIND- 100) from BioAssay Systems in accordance with the manufacturer's instructions. In summary, a 96-well plate was added to a SpectraMax ®

iD3 Plate Reader Molecular Devices, Sunnyvale, USA), and a program was set up to add 100 µl Quantichrom™ reagent (BioAssay Systems) to each well. The optical density was measured at 565 nm, and background absorbance was subtracted from the obtained measurement. The concentration of indole in the supernatant was calculated using a calibration curve.

2.8.1.2 Measurement of cell associated indole

To assay the indole contained within the cell pellet, i.e. cell associated indole, a modified method was used. A sample (1 ml) from a growing culture was removed, the optical density was measured using a microplate reader and the sample centrifuged at 1137 x g for 1 minute. The supernatant was discarded, 300 µl of Kovacs Reagent (5 g *p*-dimetylaminobenzaldehyde, Sigma, Belgium; 75 ml amyl alcohol, Sigma, Germany; 25 ml 37% w/w HCl, Fisher Scientific, Loughborough, UK) was added to the cell pellet and mixed well. The mixture was left for 2 minutes at room temperature. This achieved lysis of the cell pellet, as well as complete reaction of indole with the Kovacs reagent. This mixture was pipetted onto 1 ml LB before 10 µl was removed and added to 190 µl of HCl-amyl alcohol solution in a 96 well plate. The plate was inserted into the microplate reader, and the optical density was measured at 540 nm. The absorbance of control cells at the same OD_{600} was subtracted from the obtained measurement to account for absorbance of the cells. The concentration of indole was calculated using a calibration curve.

The concentration of indole obtained from the calibration curve was used to calculate the number of moles of indole present in the cell pellet. Using optical density measurements, the volume of the bacterial cell within the pellet could be calculated (213), and hence the cell-associated indole concentration.

2.8.2 Measurement of TnaA-GFP expression: fluorescence intensity

LB medium (10 ml), was inoculated with a single bacterial colony and incubated overnight at 37°C with shaking at 150 rpm. Overnight cultures were diluted to OD_{600} = 0.05 in LB medium with an additional 0.5 mM tryptophan. Samples were placed in a 96-well black bottom plate. Plates were left to incubate at 37°C with shaking. Measurements of

absorbance at 600 nm and GFP fluorescence measurements using 480 nm excitation and 510 emissions were taken every 15 minutes for 18 h.

For single-read experiments, LB medium (10 ml) with additional 0.5 mM tryptophan, was inoculated with a single bacterial colony and incubated overnight at 37°C with shaking at 150 rpm. Overnight cultures were diluted to an OD_{600} = 0.05 in LB medium with additional 0.5 mM tryptophan. Samples were grown to an OD_{600} = 0.5 (exponential phase) and OD_{600}= 2.0 (stationary phase). Samples were taken from the different time points, washed twice in PBS and concentrated to a final OD_{600} = 2.0. The washed cells were added to cuvettes and measured in a fluorometer. Measurements of GFP fluorescence were taken using a fixed excitation (488 nm) and a range of emission values (490-530 nm) (Cary Eclipse Fluorescence spectrophotometer). Cells not containing GFP were used as a control.

2.8.3.1 Induction of pEK plasmids

L-(+)-Arabinose (Sigma-Aldrich Corp, Milwaukee, USA) was added to an inoculated culture to a final concentration of 0.2% and incubated for the duration of the growth curve (8 h).

2.8.3.2 Induction of pCA24N

IPTG (isopropylthio-β-galactoside, Thermo Fisher Scientific, Newport, UK) was added to an inoculated culture to a final concentration of 0.1 mM and incubated for the duration of the growth curve (8 h).

2.8.4 Reaction rate experiments

LB medium (10 ml), was inoculated with a single bacterial colony of *E. coli* BW25113 WT and relevant cultures were incubated overnight at 37°C with shaking at 150 rpm. Overnight cultures were diluted to OD_{600} = 0.05 in LB medium with an additional 0.5 mM tryptophan. Cells were grown to exponential phase OD_{600} = 0.50 and stationary phase OD_{600} = 2.0. Samples were taken at these stages of growth and concentrated to the same final optical density OD_{600} = 2.0. These cultures were washed twice with PBS to remove

any traces of indole present. The cells were resuspended in PBS. Cultures (100ul) were added to a 96-well black/clear bottom plate. A protocol was set up on the SpectraMax ® iD3 Plate Reader. In summary, tryptophan (2 mM) was added to each of the wells. Quantichrom™ reagent (100 ul) was added to wells after set periods of time (5 min, 8 min, 15 min, 22 min) to calculate the rate at which the tryptophanase in the cells could convert tryptophan to indole as a proxy for enzymatic activity. Final concentrations of indole were calculated using a calibration curve.

2.8.5 Carbon sources and indole production

LB medium (10 ml) was inoculated with a single bacterial colony of *E. coli* BW25113 WT and incubated overnight at 37°C with shaking at 150 rpm. Overnight cultures (1 ml) were washed twice in PBS to remove any trace indole present. Cultures (100 ml) were added to a 96 clear-bottom plate and 1 mM of tryptophan was added to each well, with a control that did not have tryptophan. Rows in the plate had either glucose (2%), pyruvate (2%) or both added to the wells. Cells were left to incubate with the tryptophan and carbon sources for 20 min. Quantichrom™ reagent was added to the wells at different time points (0 min, 5 min, 10 min, 15 min, 20 min) Absorbance (565 nm) was measured in the SpectraMax ® iD3 Plate Reader. Final indole concentrations were calculated using a calibration curve.

2.9 Mother Machine Experiments

2.9.1 Strain construction

E. coli MI13 is a derivative of MG1655 7740 *ΔmotA*. It contains the fusion, *PrpsL-mCherry*, under the control of the promoter, *rpsL*, responsible for encoding the ribosomal protein S12. The fusion was inserted next to the gene, *glmS*, into the attTn7 site with the vector pGRG37. The TnaA-GFP fusion was transformed into MI13 to allow the measurement of tryptophanase production throughout growth. The strain was constructed under the same procedure described previously (Methods and Materials 2.7.5).

2.9.2 Mother machine chip experimental set-up

The mother machine is a microfluidic device that is capable of high-throughput study of individual bacterial cells throughout a growth cycle. The device is made from polydimethylsiloxane (PDMS) and fabricated from a patterned silicon wafer. Electron beam lithography is used to create channels that are just wide enough to fit a single cell, and long enough to fit approx. 10 cells (channel length = 25 µm, channel width = 1.5 µm, channel height = 1.28 µm).

The entire set-up was housed in an incubator that kept the flask culture and mother machine chip at 37°C. For optimum growth, large batch cultures must be shaken well to enable thorough mixing and aeration. The flask culture was grown in a 100 ml Erlenmeyer flask with 20 mL of LB broth that contained an additional 0.5 mM tryptophan. Pluronic fluid was added to the culture (0.085%) to prevent the cells from sticking and clogging the channels. The flask culture was aerated with a magnetic stir bar and a magnetic heat platform. A set of pumps takes the culture through the mother machine device loaded on the microscope stage. Another set of peristaltic pumps circulates the culture to and from a syringe tip that allows extraction of the culture to assay optical density and indole concentration (214).

2.9.3 The mother machine experiment

A single colony of *E. coli* MI13 *tnaA-GFP* kan^R was inoculated in a growth culture tube overnight in 5 mL of LB with additional 0.5 mM tryptophan and 0.085% pluronic fluid. This culture was grown for 16 hours, shaking (200 rpm) at 37°C. A 1 mL aliquot of an overnight culture was centrifuged at 4 000 rpm (Eppendorf™ MiniSpin™) for 5 min and the majority of supernatant was discarded. The cells were resuspended in the remaining 100 mL of liquid into a viscous mixture. Using a micropipette tip, the cells were added to the microfluidic chip into the parallel channels using osmotic pressure. An Erlenmeyer flask containing 100 mL of LB with additional tryptophan (0.5 mM) and 0.085% pluronic fluid was inoculated with the overnight culture to a final OD_{600} = approx. 0.05 and incubated at 37°C with a stir bar in the Erlenmeyer flask. Medium was pumped between the flask culture and the mother machine chip at a constant flow rate. The cells in the microfluidic chip were photographed approx. every 3 minutes using 1.49NA Plan Apo phage contrast

and 100x oil objective. GFP fluorescence was measured at an excitation of 490 and an emission wavelength of 510 (as set-up by Georgeos Hardo). The cells were monitored for over 6 hours until they reach stationary phase, at which point the culture is too dense to get readable data.

Samples of media (0.2 mL) were taken at 30 min intervals to measure optical density (600 nm) and indole in the supernatant. Half of the sample (0.1 mL) was diluted in PBS and measured for optical density (600 nm). The optical density was recorded, and a growth curve was made. The rest of the sample was centrifuged for 1 min at 13 000 rpm (Eppendorf™ MiniSpin™). The cells were discarded, and the supernatant was kept on ice until samples had been collected at all designated time points. The supernatant was added to a 96-well clear plate. Quantichrom™ reagent was added to the wells and allowed to incubate for 1 min. Absorbance (565 nm) was measured in the SpectraMax ® iD3 Plate Reader. Final indole concentrations were calculated using a calibration curve.

2.9.4 Analysis of the mother machine data

Python (v 3.11) scripts (Georgeos Hardo) were used to analyse the images and count foci. A learning program was developed by Georgeos Hardo, and the necessary supervised learning to create synthetic data was performed by the author of this thesis to identify foci and segmentation of the cells (215, 216). SyMBac was used to generate 4000 synthetic training images using phase contrast to match the experiment (216). SyMBac rapidly and automatically creates a detailed synthetic training data, combining models of cell growth, physical interactions and microscope optics to resemble real micrographs, aiding in training accurate image segmentation models. A U-net deep learning network, which is a type of deep learning neural network used for image segmentation, was trained onto this to segment the experimental data into masks. The mother cell's mask was isolated, and fluorescence channels across all time points and cells were analyzed. A threshold of intensity isolated foci, with noise filtering to reject foci that were too small, too large, or beyond 1.5x the interquartile range (IQR). Python (v 3.11), MATLAB R2024a and Microsoft Excel were used to create the necessary statistical tests and graphs.

2.10 Bioinformatics

2.10.1 *E. coli tnaA* analysis

2.10.1.1 Acquisition of genome sequences and relevant metadata

A collection of 1974 complete *E. coli* genomes was acquired and downloaded from NCBI Assembly (https://www.ncbi.nlm.nih.gov/assembly/, August 2020 and February 2021) along with the Assembly and Annotation Report table from NCBI Genome. The Biological Source information for the *E. coli* genomes were retrieved from NCBI BioSample in August 2020/February 2021 and a Python (v3.9.2) script was created to extract the relevant metadata. In the program RStudio, the assembly tables were combined with the metadata. The genomic sequences were concatenated into a single sequence through Terminal and any plasmids were subsequently removed. Any strain duplicates in the collection were removed to create a compilation of 1251 genomes.

2.10.1.2 Alignments

A custom BLAST database was generated from the remaining 1251 sequences. The sequences were aligned through BLASTn (https://ftp.ncbi.nlm.nih.gov/blast/executables/LATEST/), downloaded August 2020 and February 2021) and the following outputs were used: subject sequence ID, start and end of alignment in query, start and end of alignment in subject, percentage identity, alignment length, E-value and nucleotide sequence of the alignment.

A nucleotide BLAST was performed for each individual gene in the *tnaCAB* operon *(tnaC, tnaA* and *tnaB*), as well as the operon itself. *E. coli* BW25113 was used for a reference sequence of TnaA and was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/nuccore/CP009273).

The reference sequences for seven MLST genes (*adk, fumC, gyrB, icd, mdh, purA* and *recA*) (217) from *E. coli BW25113* were extracted from NCBI. A BLASTn search was conducted against the created database to generate alignments for the seven genes and the results were concatenated by Dr. Marta Matuszewska.

The following was performed by Hannah Lin (HL) under the supervision of Ellis Kelly (this author) and Dr Matuszewska. Cut-offs for percent identity match (%) and length (bp) to eliminate short and non-specific matches were chosen based off the distribution of results. The alignment thresholds were made for length at 80% for the *tnaCAB* genes and 75% for the operon. Subsequently, a percent identity threshold of 90% was set for hits that fit the length threshold.

Unknown genes were identified through a BLASTn search of the nucleotide sequences against the NCBI database (218). The author of this thesis used Geneious (https://www.geneious.com) to translate the nucleotide sequences. Biostrings (v2.58.0) was used to create protein sequence alignments in R.

2.10.1.3 Multi-locus sequence typing (MLST)

The sequence type (ST) of the genomes were provided by Dr. Matuszewska using the MLST tool on PathogenWatch (https://pathogen.watch) that determines ST type based on the unique allele profile for the seven standard MLST genes. The speciator tool on PathogenWatch was used on the MLST alignments to assign unknown species.

The ST and species information was added to the assembly and metadata table in RStudio (HL).

2.10.1.4 Statistics analysis and visualization

Statistical analysis was conducted in Microsoft Excel, RStudio (v.4.0+) and Python (3.11.0) by both HL and author of this thesis. The RStudio applications, Ggplot2 (v3.3.3) and dplyr (v1.0.3), were used data manipulation and the creation of any plots. Another RStudio program, ape (v5.4-1) was used to create pairwise distance matrices that provided information and calculations on single nucleotide polymorphisms (SNPs) per site.

2.10.1.5 Phylogenetic Analysis of *tnaA* in *E. coli*

Using the program ape (v5.4-1), a MLST tree was generated (HL). Neighbour-joining trees for *tnaA* and *tnaCAB* based on the alignments above were created using RAxML

(author and MM). The trees were annotated in FigTree (v1.4.4) (http://tree.bio.ed.ac.uk/software/figtree/) and iTol (v6.0) (219, 220) (HL).

2.10.2 Bioinformatics of the human gut microbiome

2.10.2.1 Microbiome database and query sequences

The most comprehensive sequence resource of the human gut microbiome to date, which is the entire collection of the United Human Gastrointestinal Protein (UHGP) catalogue was downloaded (4). Protein sequences (aa) were used in this analysis due to the large amount of diversity represented in the dataset. A list of 48 TnaA different query sequences were assembled that ranged from across the phylum to capture the diversity of the gene found in the microbiome were extracted from NCBI Protein Database (https://www.ncbi.nlm.nih.gov/protein/) this is seen in Table 2.18. The lineages of the 48 TnaA query sequences were extracted from the metadata provided by the UHGP catalog. The amino acid sequences were concatenated into a single sequence through Terminal and made into a BLAST protein database using the following parameters: subject sequence ID, start and end of alignment in query, start and end of alignment in subject, percentage identity, alignment length, E-value and amino acid sequence of the alignment.

Table 2.18. The 48 TnaA query sequences from a range of distinct phyla and species that were used for phylogenetic analysis in the human gut microbiome.

Query Number	Phylum	Order	Species
1	*Proteobacteria*	*Enterobacterales*	*Escherichia flexneri*
2	*Proteobacteria*	*Enterobacterales*	*Aeromonas hydrophila*
3	*Firmicutes*	*Paenibacillales*	*Paenibacillus_B alvei*
4	*Firmicutes_A*	*Clostridiales*	*Clostridium_H novyi*
5	*Proteobacteria*	*Enterobacterales*	*Escherichia flexneri*
6	*Firmicutes_A*	*Peptostreptococcales*	*Alkaliphilus_A metalliredigens*
7	*Firmicutes_A*	*Clostridiales*	*Clostridium_G tetani*
8	*Bacteroidota*	*Bacteroidales*	*Bacteroides thetaiotaomicron*
9	*Actinobacteriota*	*Propionibacteriales*	*Cutibacterium acnes*
10	*Proteobacteria*	*Enterobacterales*	*Yersinia enterocolitica*
11	*Proteobacteria*	*Enterobacterales*	*Haemophilus influenzae*
12	*Proteobacteria*	*Enterobacterales*	*Edwardsiella tarda*
13	*Bacteroidota*	*Flavobacteriales*	*Flavobacterium sp000813005*
14	*Proteobacteria*	*Enterobacterales*	*Klebsiella_A huaxiensis*
15	*Firmicutes_B*	*Desulfitobacteriales*	*Desulfitobacterium hafniense*
16	*Proteobacteria*	*Enterobacterales*	*Morganella morganii_A*
17	*Actinobacteriota*	*Propionibacteriales*	*Nocardioides sp001627335*
18	*Proteobacteria*	*Enterobacterales*	*Proteus hauseri*
19	*Firmicutes_E*	*Symbiobacteriales*	*Symbiobacterium thermophilum_1**
20	*Firmicutes_E*	*Symbiobacteriales*	*Symbiobacterium thermophilum**
21	*Proteobacteria*	*Enterobacterales*	*Vibrio cholerae*
22	*Proteobacteria*	*Enterobacterales*	*Klebsiella aerogenes*
23	*Proteobacteria*	*Enterobacterales*	*Proteus inconstans*
24	*Firmicutes_A*	*Thermoanaerobacterales*	*Caldanaerobacter subterraneus*
25	*Proteobacteria*	*Enterobacterales*	*Vibrio vulnificus*
26	*Desulfobacterota*	*Desulfovibrionales*	*Nitratidesulfovibrio vulgaris*
27	*Proteobacteria*	*Enterobacterales*	*Pasteurella multocida*
28	*Proteobacteria*	*Enterobacterales*	*Histophilus somni*
29	*Proteobacteria*	*Enterobacterales*	*Vibrio campbellii*
30	*Proteobacteria*	*Enterobacterales*	*Citrobacter_B koseri*
31	*Bacteroidota*	*Rhodothermales*	*Salinibacter ruber*
32	*Proteobacteria*	*Enterobacterales*	*Aeromonas salmonicida*
33	*Proteobacteria*	*Caulobacterales*	*Hyphomonas neptunium*
34	*Proteobacteria*	*Enterobacterales*	*Shewanella sediminis*
35	*Proteobacteria*	*Rhodobacterales*	*Rhodobacter capsulatus*
36	*Firmicutes_C*	*Propionisporales*	*Pelosinus fermentans_B*
37	*Firmicutes*	*Staphylococcales*	*Staphylococcus aureus*
38	*Firmicutes_C*	*Sporomusales*	*Acetonema longum*
39	*Verrucomicrobiota*	*Verrucomicrobiales*	*Akkermansia muciniphila*
40	*Verrucomicrobiota*	*Opitutales*	*Opitutus terrae*
41	*Fusobacteriota*	*Fusobacteriales*	*Fusobacterium nucleatum*
42	*Spirochaetota*	*Brachyspirales*	*Brachyspira suanatina*
43	*Cyanobacteria*	*Cyanobacteriales*	*Nostoc commune*
44	*Synergistota*	*Synergistales*	*Thermovirga lienii*
45	*Elusimicrobiota*	*Elusimicrobiales*	*Elusimicrobium minutum*
46	*Myxococcota*	*Myxococcales*	*Myxococcus stipitatus*
47	*Bdellovibrionota*	*Bdellovibrionales*	*Bdellovibrio bacteriovorus*
48	*Campylobacterota*	*Campylobacterales*	*Aliarcobacter trophiarum*

* This species had two TnaA sequences that were distinct.

2.10.2.2 BLAST database for gut microbiome

BLASTp (v 2.12.0+) (221) was used to compare the TnaA query sequences against the created database. The sequences were aligned in Clustal Omega (v 1.2.3) in Geneious Prime (Java Version 11.09+11) and any sequences that contained a high proportion of gaps (25%) were excluded to ensure quality of the alignment (Python script by MM). Any alignment sites nearly identical or present in >90% of the columns were removed to provide a more informative phylogenetic analysis. Any low coverage score, due to insertions, and any hits with an e-value of 0 were removed. The graphs to represent the data was formed in Python, R (v 4.1.0), Ggplot (v 3.3.5), Ggrepel (v 0.9.1) and Tidyverse (v 1.3.1).

2.10.2.3 Phylogenetic analysis of TnaA in the gut

Using the RAxML (v 8.2.12) software tool, a phylogenetic tree (Parameters: m–PROTGMMAWAG, -N 100) was inferred from the above amino acid alignment(222). Any node with a bootstrap of <70 was removed to ensure only branches that were strongly supported by the data were included (223). iTOL (v 6.3) (219) and GrapeTree (v 3) (224) were used to visualise and annotate the phylogenetic trees.

Chapter 3. Dissecting indole pulse signalling

3.1 Introduction

The indole pulse is a transient high concentration of cell-associated indole that occurs during the cell's entry into stationary phase (147). Following the pulse, cell-associated indole declines and eventually an equilibrium is reached between the cells and growth medium. Previous studies have suggested the pulse is a result of indole production outpacing the ability of indole to dissipate from the cells by diffusion (147). High indole production correlates with a high rate of tryptophanase production in late exponential phase (172). Early researchers investigating the indole pulse focused on measuring the production of tryptophanase over time in the minutes proceeding and following the indole pulse. They observed a fivefold increase in tryptophanase production within 20 minutes preceding the indole pulse. This led to the hypothesis that the up regulation of tryptophanase expression is primarily responsible for the increased indole production during the entry into the stationary phase. However, the mechanism behind this up-regulation remains poorly understood.

The indole pulse has previously been theorized to elicit a variety of phenotypic responses in wild-type cells (seen section 1.2.4). The initial objective of this thesis was to exert control over this phenomenon in order to gain deeper insights into these diverse phenotypes. Previous studies of the pulse and its effects used labour intensive, low-throughput protocols that were subject to human error. The first goal of this chapter is therefore to develop an experimental procedure that could standardize measurements of the pulse using a high-throughput approach and commercial reagents that increase the sensitivity of the assay.

In previous experiments, the generation of the pulse was under cellular control, rather than under the control of the experimenter. Further investigation would be improved by the ability of an experimenter to induce a pulse upon demand. By controlling the production of tryptophanase from a plasmid or in appropriate mutant strains, it might be possible to achieve an indole pulse on demand. The object of this chapter is therefore firstly to reproduce prior observations of the indole pulse and then to explore its potential

to be manipulated experimentally. The hypothesis is that inducing tryptophanase production earlier in cell growth will trigger an indole pulse*, leading to a* quiescent like state in cells.

To test this, tryptophanase will be expressed under the control of a different promoter, specifically one located on a plasmid, to precisely control the timing of its production. Once the ability to express tryptophanase at a chosen time is confirmed, the levels of extracellular and cell-associated indole would be measured at various times throughout growth and correlated with the timing of tryptophanase production. If early production of tryptophanase and an early indole pulse is achieved, the long-term survival of the cells will be evaluated, following protocols from previous experiments (147, 172).

3.2 Developing and testing an improved indole pulse assay

A robust and reproducible cell-associated indole assay was designed, with standardized media and conditions (Figure 3.1). Adaptations made to the approach of Gaimster & Summers (2014) included the use of a 96-well plate reader and optimization of extra-cellular indole measurement using Quantichrom™ reagent (147). Also, the method of obtaining culture samples was modified and optimized.

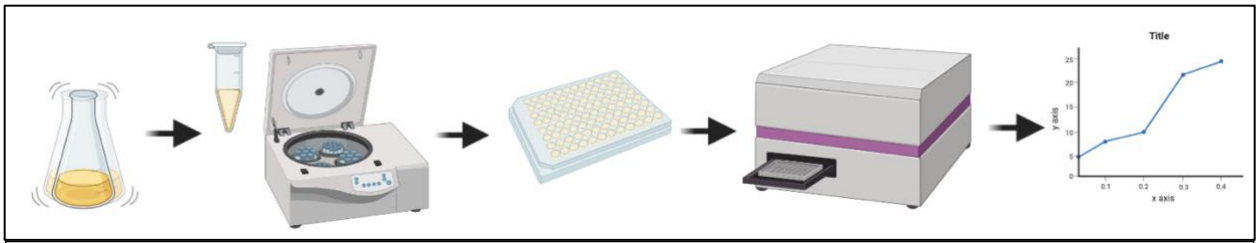


Figure 3.1. A standardized version of the indole pulse assay. *E. coli* BW25113 cells are grown in LB medium in Erlenmeyer flasks for approx. 340 min and shaking at 37 °C. Samples (1 mL) are taken throughout growth and centrifuged at 13 000g for 1 min. The supernatant and cell pellet are separated and measured independently. The supernatant samples are diluted and added to a 96-well plate where Quantichrom™ reagent is added in a 1:1 (v/v) ratio and the absorbance is measured at 565 nm. The indole within the cell pellet is measured by lysing the cells with Kovacs reagent, diluting in HCl-amyl alcohol and measuring the absorbance at 540 nm. The obtained measurements were then converted into indole (mM) and cell-associated indole (mM) using calibration curves. *Generated in BioRender.com.*

The standard pulse detection procedure was as follows. An overnight culture of *E. coli* BW25113 wild-type (WT) was diluted, in triplicate, into LB medium supplemented with additional tryptophan (0.5 mM) to an optical density at 600 nm (OD_{600}) of 0.05. Additional tryptophan was added to replicate the indole pulse, following previous experimental protocols where LB medium naturally contained much higher concentrations of tryptophan. The final concentration of indole in these experiments will be like ones previously seen in indole pulse assays (147, 172). These cultures were grown at 37 °C with shaking for 340 min, until cultures were in stationary phase. Initially the cultures were sampled and assayed for optical density (OD_{600}) and indole (supernatant and cell-associated) every 60 min. Samples (1 mL) were taken every hour until the OD_{600} indicated the culture was nearing the transition (OD_{600} = 1.5) into stationary phase (approx. 220 min after the initial sub-culture). During the transition, samples were collected every 10 min for a total of 6 samples during the transition, and 30 min thereafter, for a total of approx. 340 min. To assay extra-cellular indole (the indole in a culture supernatant), a sample (1 ml) of culture was removed, the OD_{600} measured, and the sample centrifuged at 13 000g (1 min) to pellet the cells. The supernatant was removed and assayed for indole content. Supernatant samples were diluted in LB medium in a 1:5 ratio in a 96 well-plate. Quantichrom™ reagent was added to each well in a 1:1 (v/v) ratio and the

absorbance was measured at 565 nm. Background absorbance (LB medium) was subtracted from the measurement. The concentration of indole in the sample was calculated using a calibration curve.

To assay indole contained within the cell pellet (cell-associated indole) the following method was used. A sample (1 mL) was taken from a growing culture, the OD_{600} measured, and the sample centrifuged for 1 min at 13 000g. The supernatant was removed, and Kovacs reagent was added to the cell pellet and mixed well. The mixture was left for 2 min at room temperature. This achieved lysis of the cell pellet, as well as the complete reaction of indole with the Kovacs reagent. This mixture was pipetted into LB to standardize all samples and conditions, before a portion was removed (10 μl) and added to HCl-amyl alcohol solution in a 96-well plate, to achieve a 20x dilution. The absorbance of each well was measured at 540 nm. The absorbance of BW25113 *ΔtnaA kan*^{R} at the same OD_{600} following the process noted above was subtracted from the measurement to account for absorbance of the cells. The concentration of indole was calculated using a calibration curve.

The obtained indole concentration was used to calculate the number of moles of indole present in the cell pellet. Using OD_{600} measurements, the total volume of the bacterial cells within the pellet could be calculated (213), and hence the cell-associated indole concentration.

This technique was used to assay indole production in cultures of *E. coli* BW25113 (Figure 3.2). The creation of this assay to use a plate reader and Quantichrom™ allowed for the comparison of nine different culture samples, compared to just three previously. This provided a more comprehensive view of indole pulse signalling, and its corresponding heterogeneity.

Supernatant indole (Figure 3.2A) showed the expected pattern; it was initially low but increased rapidly after exponential phase but prior to the stationary phase transition. There was very little variability among repeat experiments. Cell-associated indole measurements showed evidence of the pulse, with cell-associated indole rising to a maximum of approx. 35 mM during the exponential to stationary phase transition (Figure

3.2B; 200-300 min). However, the error bars indicate substantial variation among replicates. The cell-associated indole concentration at a state of expected equilibrium when all of the available tryptophan has been converted into indole and growth is minimal (approx. 22 mM) was calculated by measuring the indole in BW25113 *E. coli* WT cells after 16 hours of growth (OD_{600} = approx. 6). The final cell-associated indole concentration in Figure 3.2, is approx. 28 mM which is slightly higher than the equilibrium concentration.

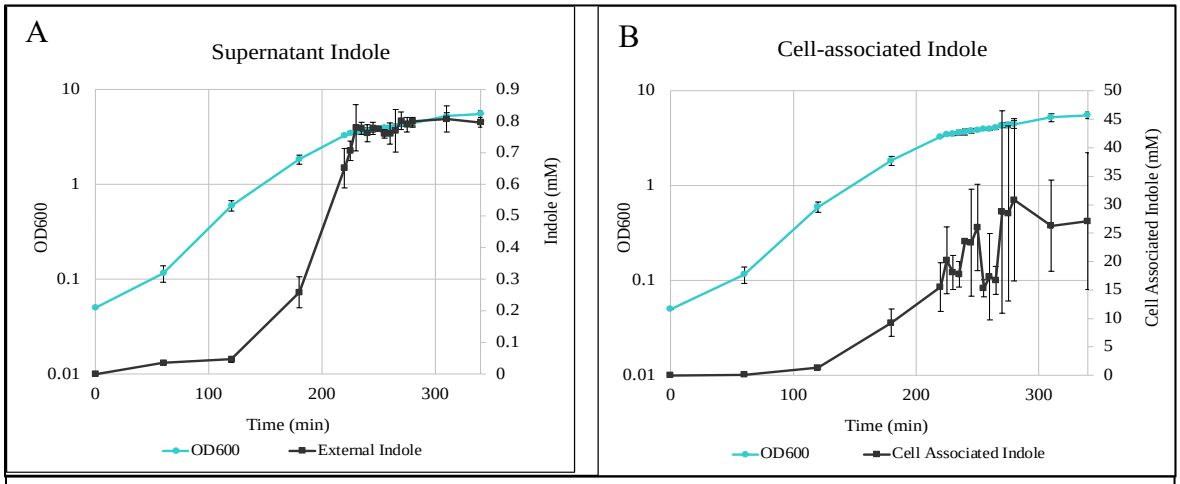


Figure 3.2. Supernatant and cell-associated indole measurements of *E. coli* BW25113 WT cultures throughout growth over 5-6 hours. Cultures were grown in LB medium plus additional tryptophan (0.5 mM) and sampled at intervals throughout the growth phase. Optical density (600 nm) was measured, and the supernatant was assayed for external indole concentrations using Quantichrom™ reagent (A) and internal cell-associated indole using Kovacs reagent (B). Data shown are the mean values ± standard error for three independent repeats.

In order to investigate further the nature of the variation in the indole pulse, the data from the three replicas were plotted separately (Figure 3.3). By separating experiments, the pulse is revealed more clearly and shows similar kinetics to what was initially reported in the literature (147). In all three experiments, supernatant indole rises to a maximum of ~0.8 mM which is consistent with observations in previous experiments. During the pulsing period (200-300 min), cell-associated indole rose from 10 to approx. 40 mM. Both the timing of the pulse and maximum cell-associated indole concentration show variation in the different experiments. Heterogeneity is seen in the exact time of the peak (approx.

240-280 min), as well as the pulse duration (approx. 30-45 min). The pulse lies approximately within the transition from exponential to early stationary phase but varies between cultures. The period of elevated cell-associated indole in these experiments last approx. 40 min which is much longer than the 10-15 min previously described (147). Following the pulse, cell-associated indole decreases in two of the replicas to a level that is associated with cells reaching a state of equilibrium (approx. 22 mM) as seen by Gaimster and colleagues (147). There appears to be considerable heterogeneity among the replica experiments that, when the data are combined, obscures the indole pulse in the individual cultures. Within an individual sample taken from the same overnight culture on a given day, phenotypic variation still exists. For example, in one indole pulse assay, three samples reached their maximum cell-associated indole concentrations at different times and levels: one sample peaked at 64 mM at 310 min, another at 63 mM at 280 min and the third at 60 mM at 275 min. However, when the data is averaged, these pulses are obscured because the cell-associated indole concentrations may have already dropped form the peak or had yet to peak. At 310 minutes, the indole concentrations were 64 mM, 32 mM and 36 mM for the three samples, respectively. This variation is not due to genetic differences or LB variation, as all samples originated from the same culture and used the same media. The observed heterogeneity might be due to the difficulty of measuring indole within cells at any given moment and/or the transient nature of indole pulses, which may vary between cultures and samples.

Additionally, previous studies have highlighted the variation in tryptophanase levels in late stationary phase cells (158). If a small number of these cells are taken for subculture, it could result in very different subpopulations. To study the phenotypic heterogeneity in the initial overnight culture and subsequent subcultures, one could examine the variation in cell-associated indole immediately after subculturing. This would reveal any initial differences in indole concentrations in subpopulations. Comparing this to the TnaA-GFP

levels in each subculture could further highlight the presence of heterogeneity in these samples.

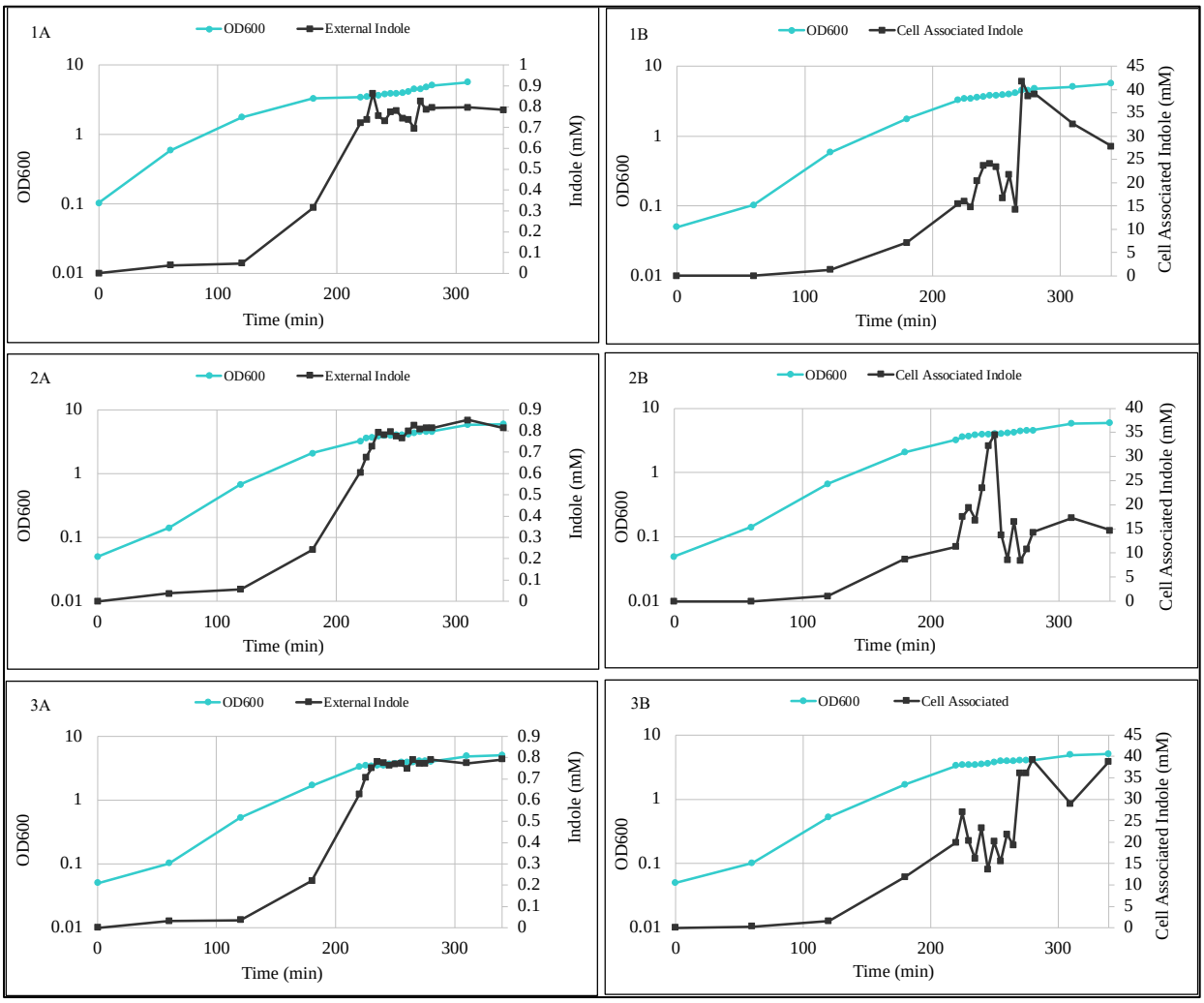


Figure 3.3 Supernatant and cell-associated indole measurements of independent *E. coli* BW25113 WT cultures throughout growth over 5-6 hours. Representative experiments are shown (1-3), as LB cultures supplemented with tryptophan (0.5 mM) were sampled, OD (600 nm) was measured, and the supernatant was assayed for external indole (A) using Quantichrom™ reagent and internal cell-associated indole (B) using Kovacs reagent.

3.3 Additional tryptophan in LB medium alters indole production dynamics

To further explore the indole pulse results obtained above, supernatant and cell-associated indole concentrations were with additional 1 mM supplement of tryptophan. Sterilized tryptophan (1 mM) was added to the LB medium at the beginning of the experiment and indole production was monitored as described previously (Figure 3.4). The final supernatant indole concentration of 1.2 mM, seen in two of the three independent experiments, is consistent with the addition of 1 mM of tryptophan to supplement the tryptophan already present in the LB medium. The assay of cell-associated indole show that the higher concentration of tryptophan creates a longer and wider pulse. The peak indole concentration was 60-80 mM and, in two of the three replicas, there was no clear indication of the concentration returning to its equilibrium value (approx. 22 mM) before the end of the experiment. This result is consistent with previous observations that increased tryptophan in the growth medium resulted in a longer “pulse period” and higher peak cell associated indole (147, 172).

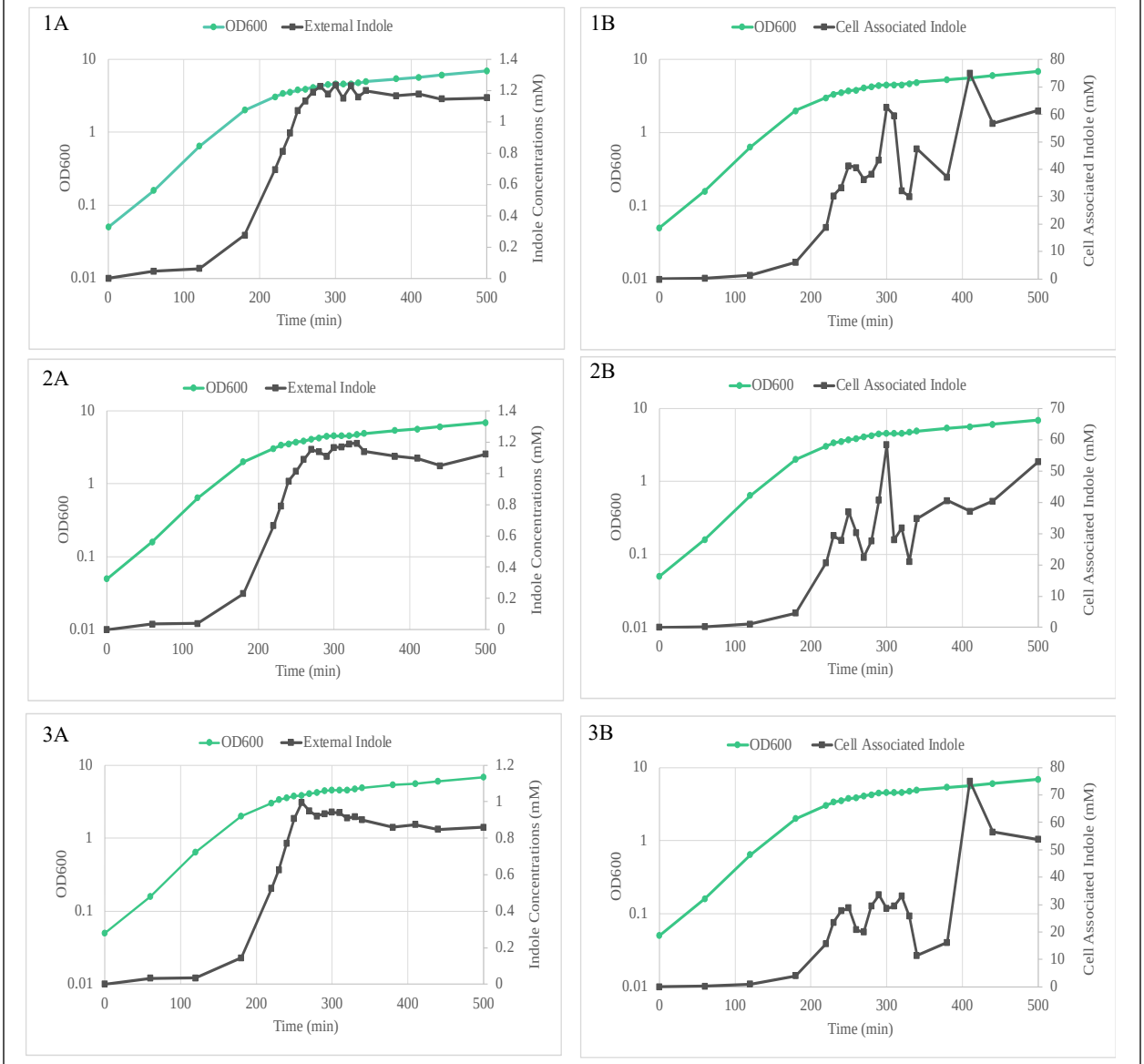


Figure 3.4. Supernatant and cell-associated indole measured throughout growth in LB media that contains additional tryptophan (1 mM). Representative independent experiments are shown (Experiments 1-3), cultures of BW25113 *E. coli* were sampled, OD (600 nm) was measured, and the supernatant was assayed for external indole (A) using Quantichrom™ reagent and cell-associated indole (B) using Kovacs reagent.

3.4 Measuring the indole pulse in a biofilm forming *E. coli* strain

The indole pulse has previously been measured exclusively in the non-pathogenic laboratory *E. coli* strain BW25113. Most *E. coli* strains are known to produce

tryptophanase and thus able to convert tryptophan to indole (168), but it has not previously been investigated whether non-laboratory strains are capable of generating an indole pulse during stationary phase entry. ATCC® 25922™ (American Type Culture Collection) was originally isolated as a human clinical sample and is commonly used as a control strain for measuring biofilm formation and testing pathogenic UTI strains. It is closer to what would be present in the human microbiota than the laboratory strains tested so far. The strain was cultured in LB medium with additional 0.5 mM of tryptophan and indole assays (both supernatant and cell-associated) were conducted as described previously (Figure 3.5). *E. coli* BW25113*ΔtnaA kan*^{R} was used as the control in this data set, as ATCC® 25922™*ΔtnaA* was unavailable at the time of measurement. Data are shown for three independent experiments.

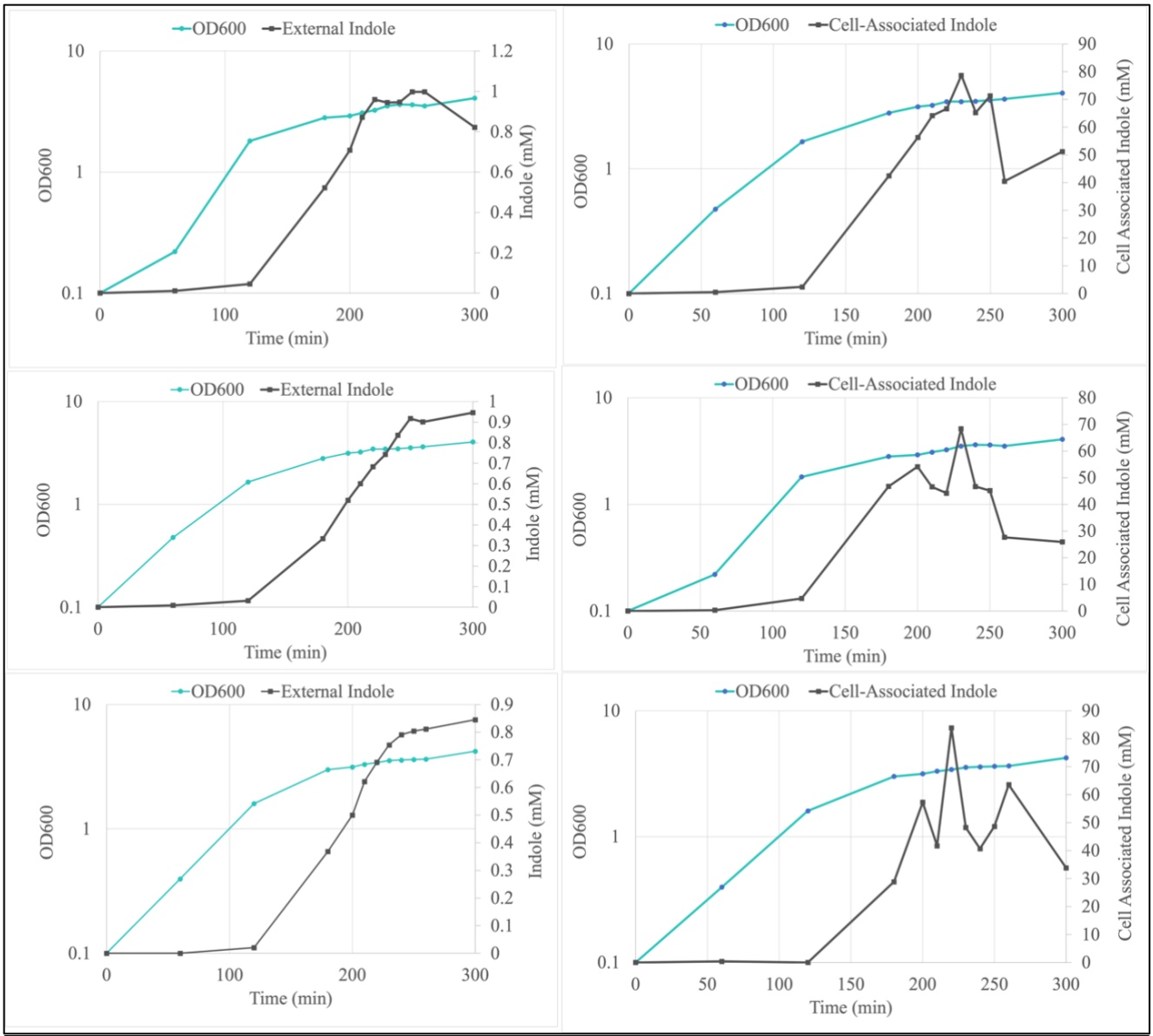


Figure 3.5. Supernatant and cell-associated indole measurements of a pathogenic *E. coli* strain (ATCC® 25922™). Representative experiments are shown (1-3), cultures of ATCC® 25922™ cells were sampled, OD (600 nm) was measured, and the supernatant was assayed for external indole (A) using Quantichrom™ reagent and internal cell-associated indole (B) using Kovacs reagent.

Supernatant indole is similar to that seen for the laboratory strain *E. coli* BW25113 WT, starting to rise after 120 min and reaching a final concentration of approx. 0.8 mM. Cell-associated measurements show a clear indole pulse at around 200-250 min. Just as was seen with the laboratory strain, there is heterogeneity in the timing of the pulse and the peak cell-associated indole concentration in *E. coli* ATCC® 25922™. Peak cell-

associated indole concentrations (70-90 mM) are higher than those seen in *E. coli* BW25113 WT and decrease to ~30 mM, which is slightly higher than the approx. 22 mM seen at equilibrium. However, it can be concluded that *E. coli* ATCC® 25922™ has an indole pulse qualitatively similar to that seen in BW25113. It seems plausible that any *E. coli* strain containing the *tnaCAB* operon will follow similar indole production kinetics.

3.5 Exploring the genetic basis of the indole pulse

Previous studies have focused on the indole pulse in the *E. coli* BW25113 wild-type strain containing the entire tryptophanase operon (*tnaCAB*). However, it is not known whether all elements of the operon are required to produce an indole pulse. To further understand the pulse, what initiates it, and how it might be manipulated, those elements of the tryptophanase operon necessary for the pulse should be identified. The kinetics of indole production were therefore explored in mutant derivatives of *E. coli* BW25113 lacking different components of the operon.

First, indole production was examined in *E. coli* BW25113 *ΔtnaB* kan^R which lacks the TnaB tryptophan importer (Figure 3.6). When the optical density (600 nm) of mutant cultures was compared to WT cells (Figure 3.3), there appeared to be no difference in growth rate. The supernatant indole reached similar final concentrations for both strains (~0.8 mM) but in the *ΔtnaB* cultures it accumulated at a slower rate. In addition, the peak cell-associated indole concentration was reached almost an hour later by the *tnaB* mutant than by the wild-type. These differences are presumably a consequence of the slower up-take of tryptophan from the growth medium by the mutant strain, due to lacking a tryptophan importer gene.

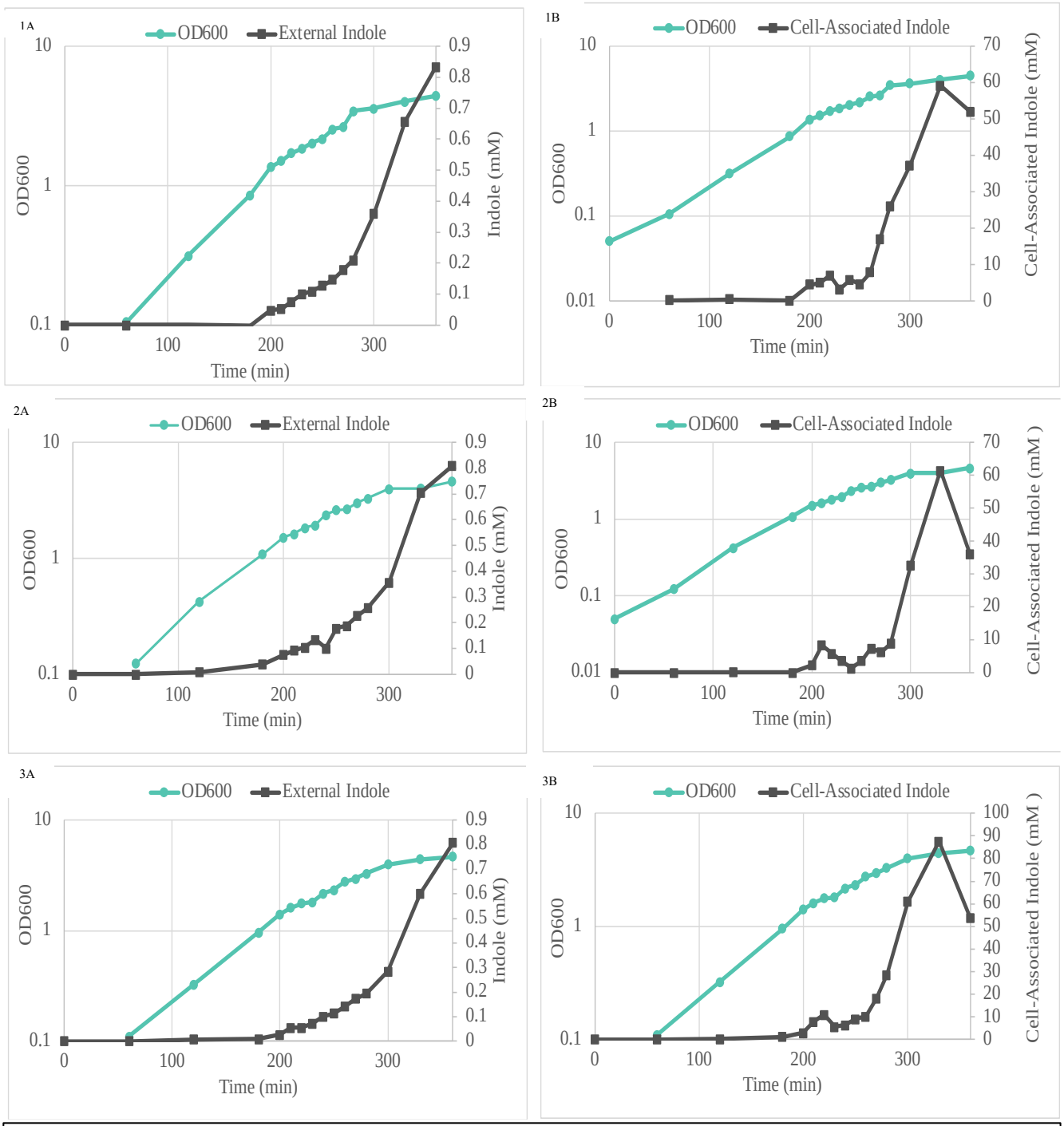


Figure 3.6. Supernatant and cell-associated indole measurements in a *E. coli ΔtnaB kan*^{R} BW25113 culture. Three independent experiments are shown (1-3). Cultures of BW25113 *ΔtnaB* were sampled, OD (600 nm) was measured, and the supernatant was assayed for external indole (A) using Quantichrom™ reagent and cell-associated indole (B) using Kovacs reagent.

The *tnaC* gene in the tryptophanase operon encodes the protein TnaC, which is required for tryptophan-regulated expression of the tryptophanase operon. To ascertain the importance of *tnaC* in the production of an indole pulse, growth and indole production were measured for *E. coli* BW25113*ΔtnaC kan*^{R}. Growth (OD600) measurements are similar to the WT strain (Figure 3.3) demonstrating that *tnaC* does not alter the bacterial growth curve. However, there is no measurable indole production during this experiment, indicating *tnaC* is necessary for production of indole, and most likely tryptophanase (Figure 3.7). When overnight (approx. 16 hour) cultures were tested for the presence of indole using Kovacs reagent, the reaction showed a final supernatant indole concentration of approx. 0.8 mM, consistent with 16-hour WT cultures. This suggests that indole, and therefore, presumably, tryptophanase, is produced at some point past the period monitored in the experiment normal growth phase (approx. 4 hrs). I would hypothesize this is due to leaky expression, where cells gradually produce tryptophanase and indole over an extended period of time, ultimately resulting in the complete conversion of available tryptophan to indole after 16 hours. To test this hypothesis, I would have sampled the external indole production and cell associated indole at regular intervals over the course of 16 hours.

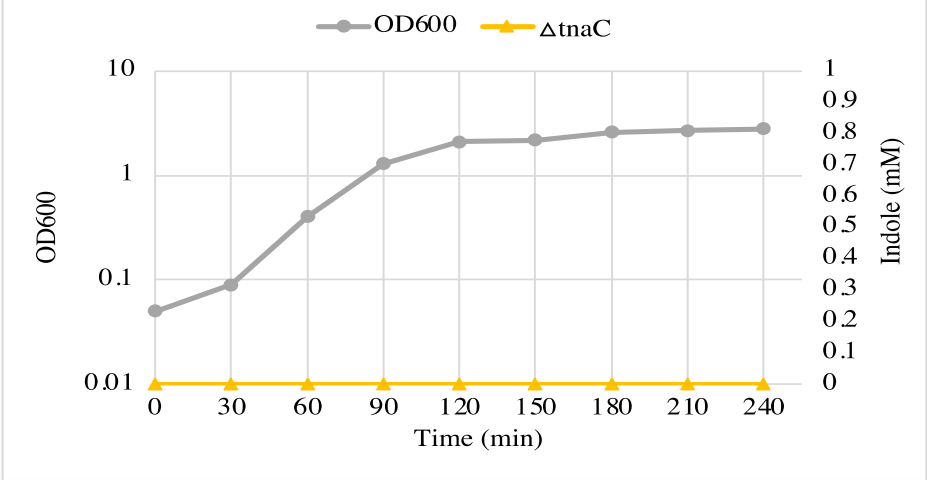


Figure 3.7. Whole indole (cell and supernatant) measurements in *E. coli* BW25113 *ΔtnaC kan*^{R}. BW25113 *ΔtnaC* cells were sampled, OD (600 nm) (grey line) was measured, and the supernatant was assayed for external indole using Quantichrom™ reagent (yellow line).

3.6 Creating an indole pulse using plasmid-encoded tryptophanase

Further investigation of the indole pulse ideally requires an experimental system where the pulse can be generated under the control of the experimenter, rather than under intrinsic cellular regulation. This section describes an attempt to achieve this using a plasmid-borne, arabinose-inducible tryptophanase operon in an *E. coli* BW25113 host with a complete knockout of the chromosomal tryptophanase operon.

A tryptophanase operon knockout was generated by modification of a tryptophanase knock-out strain (BW25113 *ΔtnaA* K^S). The strategy is summarised in Figure 3.8. The kanamycin cassette containing regions of homology flanking the operon was introduced and the DNA was sent for sequencing (see 2.7.3). The validated operon knockout strain was designated *E. coli* BW25113 *ΔtnaCAB* kan^R.

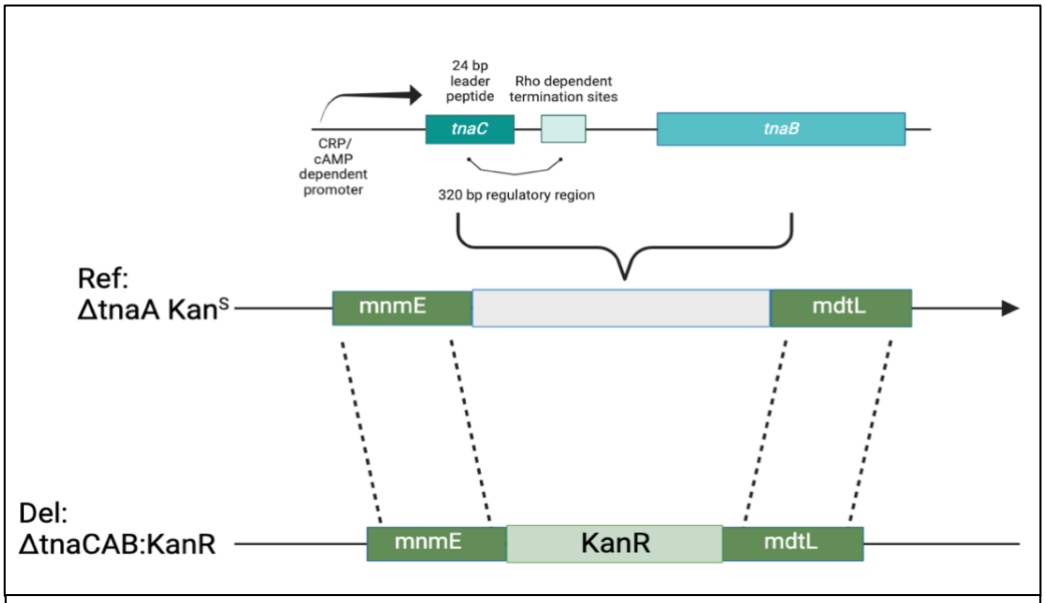


Figure 3.8. Generating an *E. coli* BW25113 *ΔtnaCAB* kan^R knockout by inserting a kanamycin resistance gene (KanR) instead of *tnaCAB* in *E. coli* BW25113 *ΔtnaA:*K^S.

Tryptophanase operon fragments were inserted into the arabinose-inducible pBAD TOPO® plasmid pJMK001. After cleavage of the vector with PmeI and NcoI (cleaved DNA was kindly supplied by C. Croft) to linearize the plasmid and leave the promoter region

for gene expression studies, it was used to clone a series of PCR-generated fragments of the tryptophanase operon generated (see 2.7.4 for complete details). pBAD TOPO® plasmids containing *tnaCAB* (pEK3), *tnaA* (pEK1), and *tnaAB* (pEK2) were constructed using Gibson Assembly. Each of the plasmids was confirmed using gel electrophoresis and DNA sequencing before transformation into the BW25113 *⊿tnaCAB Kan*^{R} strain generated above.

Overnight cultures of *E. coli* BW25113 *⊿tnaCAB Kan*^{R} containing each of the pEK plasmids in turn were inoculated in LB medium in triplicate. Growth and indole production were assayed as described previously, with the following changes. The overnight cultures were grown in the presence of arabinose (0.2%). Where appropriate (i.e. in those cultures that produce indole), this would replicate the pulse that would occur normally during the growth of overnight cultures of *E. coli* BW25113 wild-type. The following morning the cells were inoculate in in fresh LB medium that tryptophan supplementation (0.5 mM) that contain 0.2% arabinose. *E. coli* BW25113 is not capable of metabolizing arabinose, due to a mutation in the *araBAD* operon, which is responsible for the metabolism of arabinose, so the presence of the transcriptional inducer concentration should not affect growth, and this was confirmed by the OD_{600} measurements recorded during the experiment. Controls used in the experiment included non-induced plasmids and the *E. coli* BW25113 *⊿tnaCAB Kan*^{R} strain. Data values from the controls were subtracted from the data to account for any background absorbance or leaky expression.

When the *tnaA*-expressing plasmid pEK1 was present in the *tnaCAB* knockout strain, supernatant indole accumulated with kinetics very similar to the wild-type strain (Compare Figure 3.9 and Figure 3.3). Measurement of cell-associated indole revealed a clear indole pulse during the transition from exponential to stationary phase (Figure 3.9). The timing of the indole pulse is striking since the expression of tryptophanase from plasmid pEK1 was induced throughout the experiment and was under the control of a promoter that does not contain the transciprtional regulation elements in *tnaCAB*. Thus, it appears that a pulse cannot be created on demand but is linked in some way to the exponential to stationary phase transition.

It is possible that the inability to produce an on-demand indole pulse is due to the absence of *tnaB* or *tnaC* from plasmid pEK1. To explore this possibility, the experiment was repeated with pEK 2 (*tnaAB*) and pEK3 (*tnaCAB*). In cultures containing pEK2 (Figure 3.10) cell-associated indole peaked at approximately the same time as in cells containing pEK1 (Figure 3.9). The main difference is the highest cell-associated indole concentration (~80 mM) was greater in the pEK2-containing strain. This is consistent with findings that *tnaB* (present in pKE2 but not pKE1) is important for tryptophan to enter a cell, and therefore more tryptophan would be available to convert into indole for a higher and more sustained indole pulse.

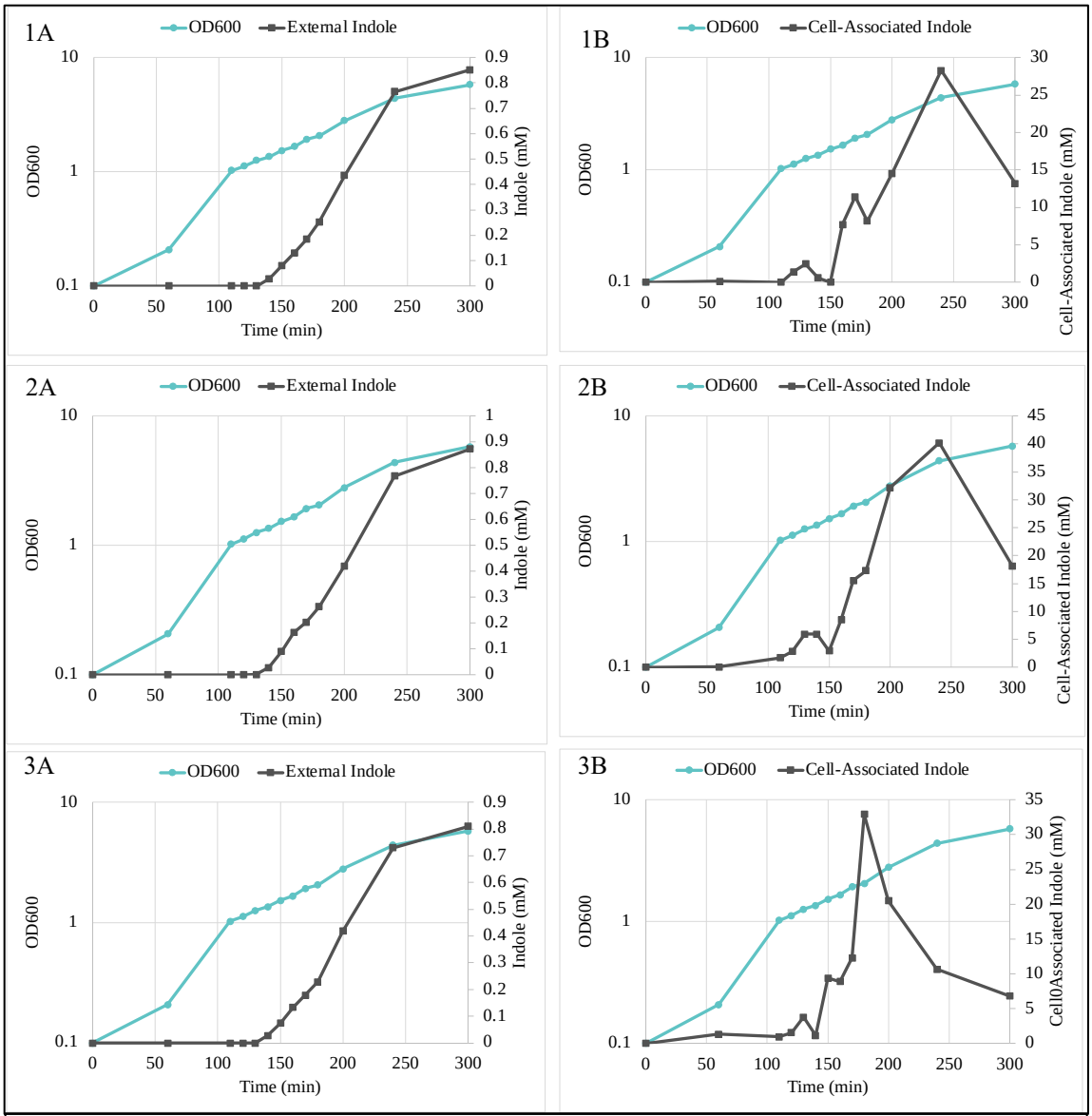


Figure 3.9. Cell-associated and supernatant indole produced by *E. coli* BW25113⊿*tnaCAB kan*^{R} containing pEK1 Representative experiments are shown (1-3). Cultures were grown in LB supplemented with tryptophan (0.5 mM), induced with arabinose (0.2%) at time = 0 min. Samples were taken throughout growth and OD (600 nm) was measured. The supernatant was assayed for external indole (A) using Quantichrom™ reagent and cell associated indole (B) using Kovacs reagent.

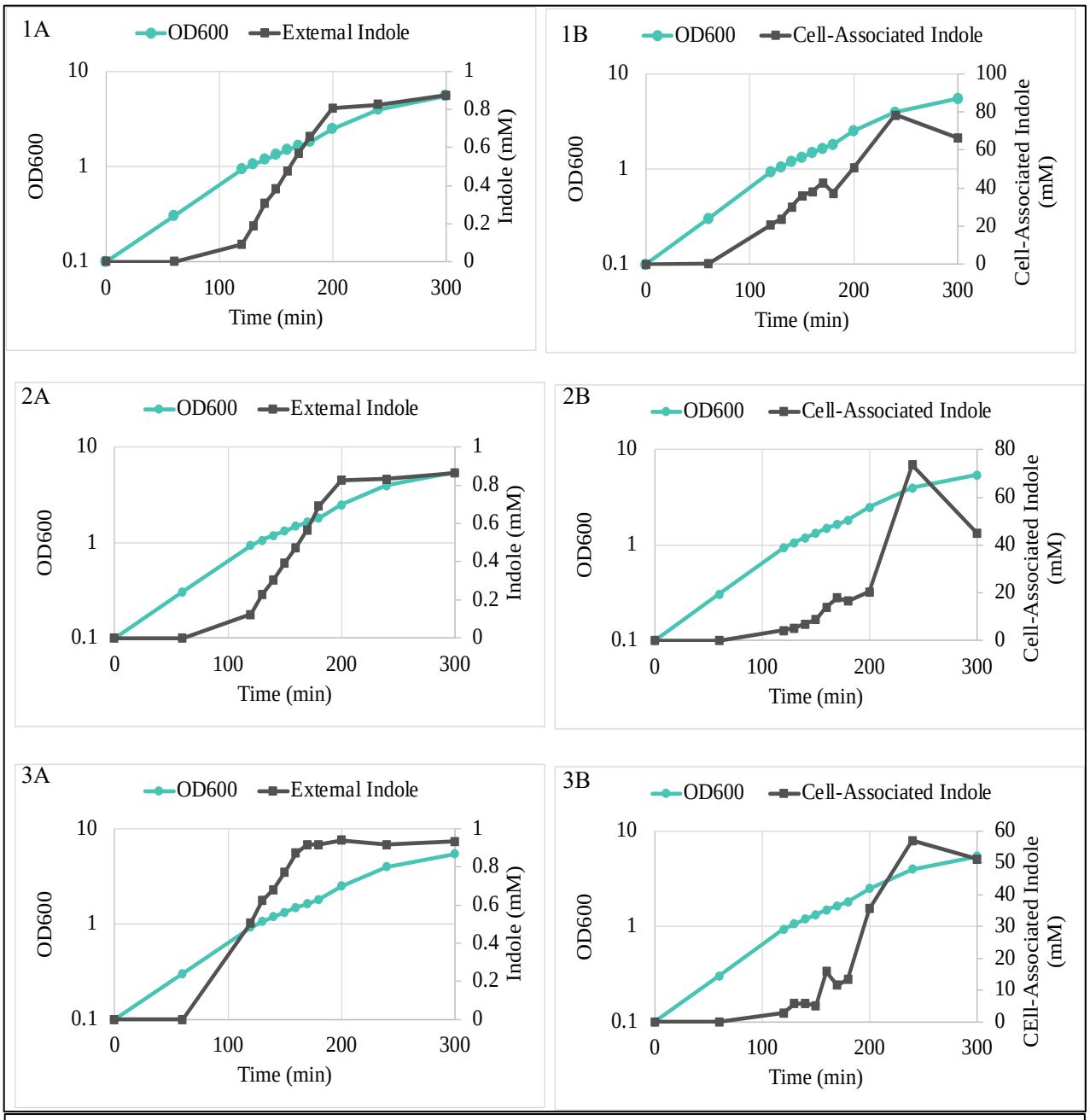


Figure 3.10. Cell-associated and supernatant indole during the transition into stationary phase produced by *E. coli* BW25113 *ΔtnaCAB* containing pEK2. Representative experiments are shown (1-3). Cultures were grown in LB supplemented with tryptophan (0.5 mM) and induced with arabinose (0.2%) at time = 0 min. Samples were taken throughout to monitor growth (OD 600 nm). The supernatant was assayed for indole (A) using Quantichrom™ reagent and cell-associated indole (B) was assayed using Kovacs reagent.

In the plasmid that contains *tnaCAB*, pEK3, the timings of production and concentrations of indole are the most consistent with *E. coli* BW25113 WT experiments (3.1) (Figure 3.11).

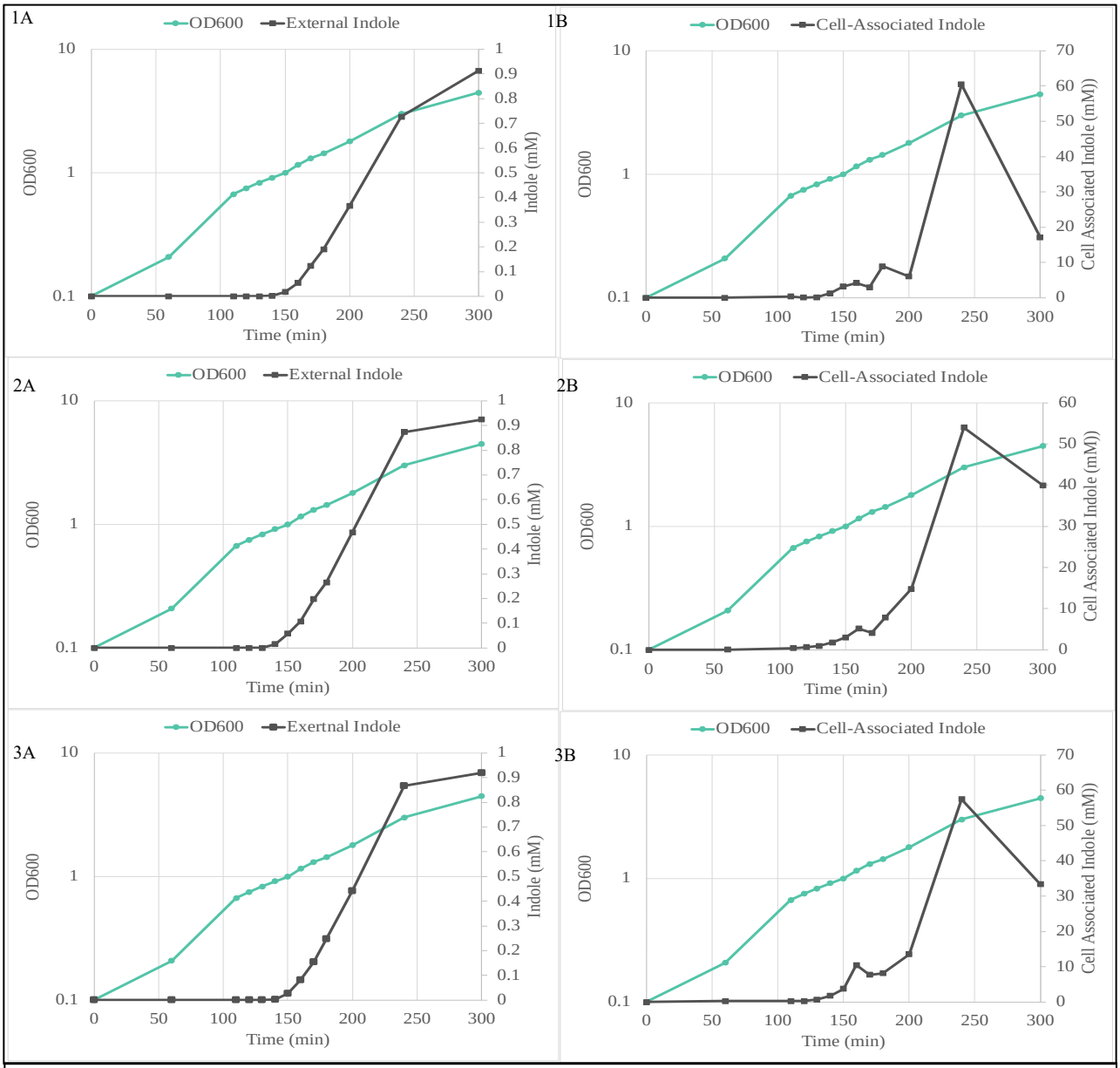


Figure 3.11. Cell-associated indole and supernatant during the transition into stationary phase produced by pEK 3 in *E. coli* BW25113 *ΔtnaCAB*. Representative experiments are shown (1-3), as cultures were grown in LB supplemented with tryptophan (0.5 mM), induced with arabinose (0.2%) at time = 0 min. Samples were taken throughout growth and OD (600 nm) was measured for comparison. The supernatant was assayed for external indole (A) using Quantichrom™ reagent and internal cell associated indole (B) using Kovacs reagent.

3.7 Discussion

The initial hypothesis of this chapter was that inducing tryptophanase production earlier in cell growth will trigger an indole pulse*, leading to a* quiescent like state in cells. The first part of this chapter aimed to reproduce the indole pulse. The initial experiments described in this chapter demonstrated that it was possible to reproduce the previously described phenomenon of the indole pulse, confirming the work of (147). The pulse measured in these experiments were wider than previously seen in by Gaimster and colleagues and lasted for an additional approx. 30 min (147). This difference is assumed to be caused by the difference in experimental technique and variations in LB. The exact levels of tryptophan and different metabolites in the LB medium has significantly changed over time, as tryptophan is now estimated to be approx. 0.3 mM compared to 0.8 mM at the time of the experiments in Gaimster and colleagues (147). Further, this thesis used more technical and biological repeats than previously used, which may introduce heterogeneity from different cultures. Given the inherent difficulty of the assay, sampling errors may contribute to variations in both the maximum cell-associated indole levels and their timings. I hypothesize that with additional repeats, this variation may be reduced.

The OD_{600} measurements in these experiments reached a much higher optical density (600 nm) than was used in the original indole pulse assays. The indole pulse was measured at an OD_{600} of approx. 2.5 in these experiments, compared to an OD_{600} of approx. 1. Typical optical density reaches a higher value of approx. 7, compared to an optical density of approx. 4.5 (147). This indicates that a higher bacterial cell density in LB medium. However, the optical density may change due to bacterial cell size or background noise from the supernatant.

The variation in LB is most likely due to manufacturing differences over time. As LB is not a defined media, the variations in yeast extract and tryptone may mean that the cells are exposed to varying levels of essential nutrients. This would create variations in the growth of the cells, as seen through representative optical density sampling, but also be seen in indole levels, as the initial amino acids may also be different. Further, research has shown that the growth curve of *E. coli* in LB medias differs from other media, with exponential

growth ends at an optical density of 0.3 due to a lack of carbon sources (225). Although LB is necessary for these indole pulse assay because the cells do not undergo catabolite repression associated with the operon, it is important to acknowledge that LB media is not ideal for reproducible assays.

The timing of the pulse consistently occurred after the cells transitioned from exponential growth to the stationary phase plateau. However, significant heterogeneity was observed between individual cells, which would require single-cell techniques, rather than conventional bacterial culture, to confirm. Due to this heterogeneity, pooling data from different experiments made the indole pulse harder to identify. Therefore, data from independent cultures were presented separately to better highlight these individual variations.

The cell-associated indole concentration (approx. 22 mM) that was measured during late stationary (16 hr) is assumed to be the equilibrium point. At this point, the supernatant indole concentration is the same as is measured in early stationary as all available tryptophan has been converted into indole. The calculated equilibrium value of cell-associated indole is consistent with previous experiments that correlated supernatant indole to cell associated indole (147). However, the cell-associated indole concentration (approx. 22 mM) is still much higher than the supernatant indole concentration (approx. 0.8 mM). This can be explained by approx. 90-fold higher affinity of indole for lipids in the cells than water, due to the hydrophobicity of the structure of the aromatic ring of the molecule (147). Thus, given time to equilibrate, the cell-associated indole concentration is still much higher as the indole remains in the pellet.

Initially, it was hypothesized that all elements of the operon were required for an indole pulse in *E. coli*. However, the mutant strains lacking *tnaB* was still capable of pulsing, likely due to other tryptophan importer genes. In contrast, a *ΔtnaC* strain could not produce indole within usual experimental timeframe but did produce indole over 18 hours. From this data, it can be hypothesized that tryptophanase is produced at some point past the typical growth phase (approx. 4-6 hrs). The delayed indole production may result from leaky expression, with cells gradually producing tryptophanase and indole over time, fully

converting available tryptophan to indole after 16 hours. To test this hypothesis, both external and cell-associated indole production would be measured at regular intervals over 16 hours.

In this thesis, two different *E. coli* strains were used to see if the presence of a *tnaCAB* operon would produce an indole pulse in other strain. It was confirmed that the pathogenic *E. coli* serotype, ATCC® 25922™, can produce an indole pulse with similar timing. From these results, we would hypothesize that any *E. coli* serotype with the *tna* operon would undergo indole pulse signalling. To test this, I would select 10+ strains from a variety of settings, including more pathogenic strains, and environmental strains, and conduct the same indole pulse assays.

To test the hypothesis that the indole pulse can be induced by early tryptophanase expression, an experimental system was created that put the timing and generation of the pulse under the control of the experimenter. The pliability of the pulse was attempted by placing tryptophanase production under the control of an inducible promoter however, this failed to yield the desired effect. The data indicate clearly that the control of the indole pulse is more complicated than previously thought. There are several possibilities that may explain why the pulse may not be able to be controlled. This includes the plasmid-encoded tryptophanase enzyme not being produced until the transition from exponential to stationary phase or the enzyme being produced early but not being active until the transition.

Chapter 4: Evidence of post-translational regulation of tryptophanase activity

4.1 Introduction

As described in detail in Chapter 1, the regulation of *tnaA* transcription involves catabolite repression, anti-termination and the sigma factor RpoS (136, 226, 227). However, since LB medium does not contain glucose or other sugars that cause catabolite repression, this mechanism is not pertinent to experiments described in the previous chapter. There is also evidence of post-translational regulation of tryptophanase. This is most clearly illustrated by the plasmid ColE1-encoded transcript Rcd that up-regulates tryptophanase activity, stimulating indole production in response to plasmid multimerization. There is also some evidence for a poorly understood mechanism of metabolic post-translational regulation during the latter stages of stationary phase (158, 172, 207).

In a way that is not clearly understood, the physical distribution of tryptophanase in the cell is thought to correlate with activation or deactivation of the enzyme. This first became apparent during studies of the GFP-fusion derivatives of tryptophanase that are used extensively in this work (206, 228). The *tnaA* gene is fused to a GFP protein domain, as in the strain *BW25113 tnaA-GFP kan*^{R}, which allows the visualization of TnaA-GFP expressed from its native chromosomal locus. A single focus of TnaA-sfGFP seen initially in exponential phase subsequently disperses as the indole concentration in the supernatant increases. Since indole formation implies the presence of active tryptophanase, it has been hypothesised that breakdown of the TnaA-GFP focus somehow causes increasing tryptophanase activity (206). The rise in indole production upon the entry into stationary phase has alternatively been attributed to increased *tnaA* transcription (172). It is possible that the two mechanisms work in concert.

Results described in the previous chapters have demonstrated that the timing of the indole pulse is not easily altered. The failure to produce an early pulse using plasmid-encoded tryptophanase in various mutant backgrounds suggests that either expression of the tryptophanase protein is being impeded earlier in growth or that the protein is

present, but its activation occurs only during the late exponential and early stationary phases of growth.

This chapter delves into the various potential factors that might contribute to the inability to separate the indole pulse from the transition between exponential and stationary phase. First of all, it is important to determine whether tryptophanase protein was being produced successfully from the expression plasmid. Further, did the timing of the indole pulse reflect the timing of *de novo* tryptophanase expression from the plasmid or the activation of pre-existing protein? To investigate this latter possibility, the activity of tryptophanase in different growth phases will be explored to evaluate the possibility of tryptophanase activation being a necessary prerequisite for the indole pulse.

4.2 Quantification of plasmid-mediated tryptophanase expression

A possible explanation for the failure to produce an indole pulse during exponential phase is that plasmid-encoded tryptophanase was not produced during this phase of growth. A different promoter and plasmid were used because the pCA24N *tnaA-GFP* cm^R plasmid was already validated and present in the literature (158). To confirm that transcriptional induction of a plasmid-borne *tnaA* gene results in successful production of tryptophanase protein, plasmid pCA24N *tnaA-GFP* cm^R, was introduced into *E. coli* BW25113 *ΔtnaCAB* kan^R by electroporation. The strain was cultured, in triplicate, in LB medium containing IPTG (0.1 mM) and supplemented with tryptophan (0.5 mM). Aliquots of culture (300 μL) were added to a 96-well plate and GFP fluorescence was measured in a SpectraMax iD3 plate reader. The plate was maintained at 37°C and the plate was shaken throughout growth. A protocol was set to measure fluorescence every 15 min (λ = 508 nm) for approx. 18 hours. A plasmid-containing strain that was not induced by IPTG, and a plasmid-free strain, were used as controls. These are visible in Figure 4.1 and produced minimal fluorescence, most likely due to noise from the cells themselves. This confirms that all fluorescence originates solely from the fusion protein expressed from the plasmid's promoter. However, plate reader studies have limitations, as cells often struggle in this environment due to restricted oxygen and nutrient availability in the enclosed space. The shaking motion of the plate does not facilitate gas exchange as effectively as a shaking

flask culture. Therefore, these results should be validated in a flask culture, which offers a more supportive environment for cell growth.

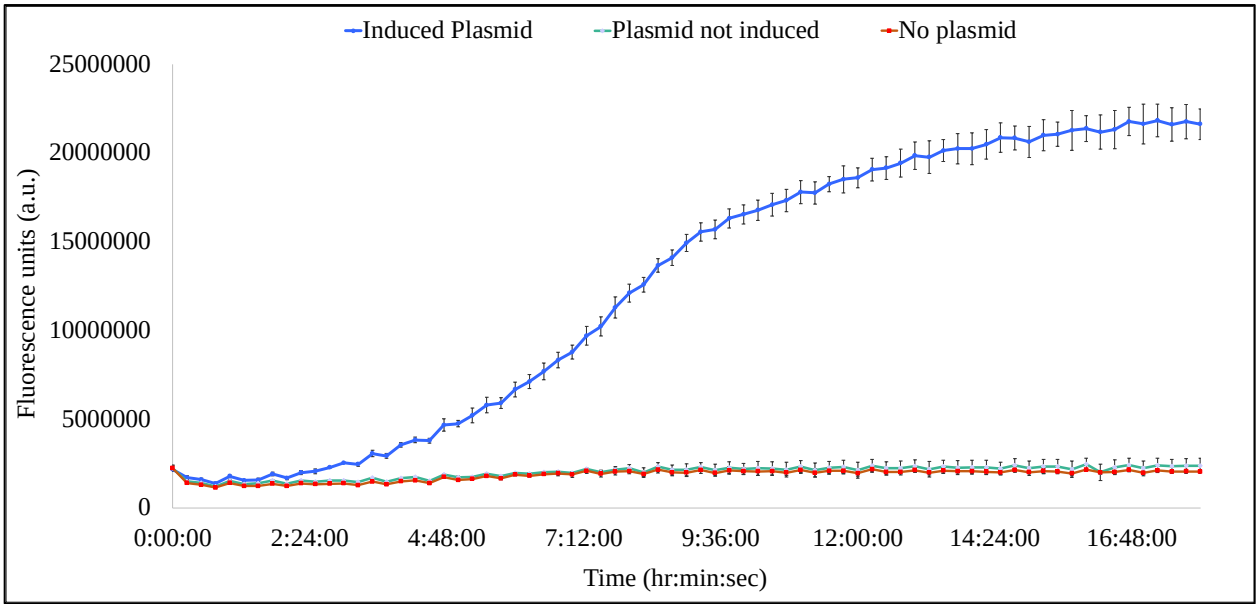


Figure 4.1. Fluorescence of plasmid-encoded TnaA-GFP protein produced by *E. coli* BW25113 *ΔtnaCAB kan^R pCA24N.* The culture was inoculated at OD_{600} = 0.1 (t=0) in LB medium and induced with IPTG (0.1mM). Aliquots of the culture were added to a 96-well plate and grown for 18 hrs at 37°C. Data shown is the average of at least 3 independent experiments. Error bars show standard deviation.

Figure 4.1 confirms good expression of TnaA-GFP from pCA24N *tnaA-GFP cm^R* where tryptophanase is detectable from approx. 3 hr after inoculation. There was a steady increase in TnaA-GFP fluorescence over time until a plateau was reached after 12 h (late stationary phase). This result indicates that tryptophanase production from the IPTG-inducible plasmid occurs throughout growth. This contrasts with tryptophanase synthesis from the native *tna* operon in the *E. coli* chromosome that primarily occurs immediately before the indole pulse (172). However, it is important to note that the experiment depicted in Figure 4.1 does not capture information on the culture density. Consequently, it is not possible to say whether the increase in fluorescence demonstrates an increase in tryptophanase expression *per* cell or a constant rate of production in an increasing number of cells.

4.3. Quantifying per-cell expression of TnaA-GFP from plasmid *pCA24N*

The next experiment was designed to estimate tryptophanase expression per cell, over a range of cell densities. Optical density is used as a proxy for cell number. This should provide a much clearer estimate of when, during growth, individual cells express tryptophanase when under the control of the plasmid promoter.

The experiment is illustrated in Figure 4.2. BW25113 *⊿tnaCAB kan*^{R} *pCA24N* cells (plasmid-mediated expression of tryptophanase from P_{lac}) were inoculated in LB medium containing additional tryptophan (0.5 mM) and induced with 0.1 mM IPTG where appropriate. Cells were inoculated at an OD_{600} = 0.1 and grown at 37°C, with shaking for approx. 300 min. Samples (1 mL) were taken every 30 min, the OD600 was recorded, and they were centrifuged at 13 000 rpm for 1 min. The cells and supernatant were separated, the supernatant was assayed for indole using Quantichrom^{TM} reagent and the cells were washed twice with PBS and resuspended in 1 mL of PBS. The washed cells were measured for GFP fluorescence in a fluorimeter at an excitation wavelength of 488 nm. To give a per-cell estimate of tryptophanase protein, the fluorescence measurement at each time point was divided by the optical density (600 nm).

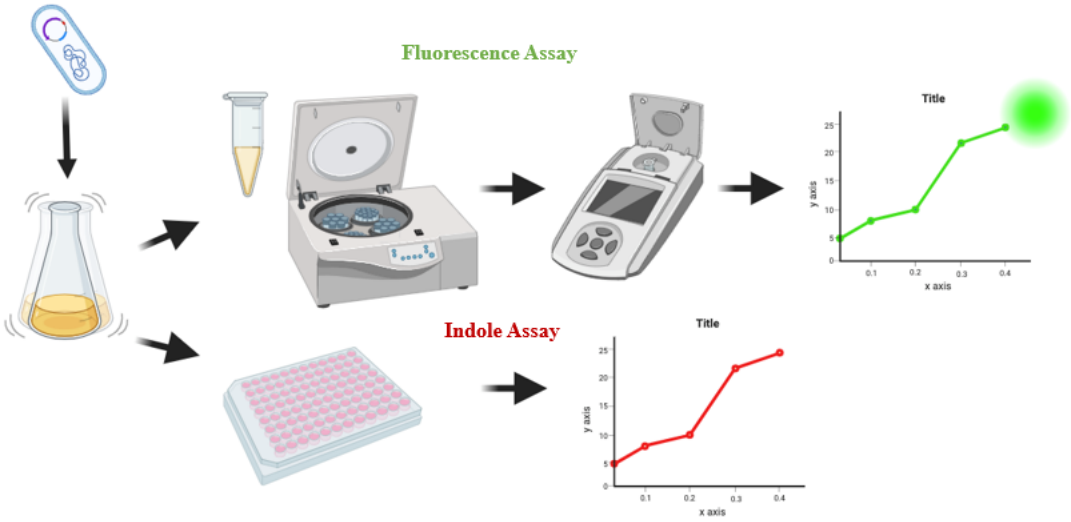


Figure 4.2. Experimental procedure for estimating the per-cell fluorescence of TnaA-GFP in *E. coli* cells grown in LB medium. The cells and supernatant were separated, the supernatant was assayed for indole using Quantichrom™ reagent. The cells were measured for GFP fluorescence in the fluorimeter which was subsequently divided by the optical density (600 nm).

In Figure 4.3, the TnaA-GFP fluorescence/OD_{600} of BW25113Δ*tnaCAB kan*^{R} *pCA24N* cells is initially high while the cells are in lag phase (OD_{600} = 0.1) showing an average of approx. 20 fluorescence units per optical density unit (600 nm). The cultures reach a steady state (approx. 7 a.u./OD_{600}) during exponential phase that is maintained for approx. 90 min before fluorescence per cell increases again as OD_{600} approaches 1, and the cells start the transition into stationary phase. As the cells enter stationary phase the fluorescence reaches approx. 15-20 a.u./OD_{600} which is very similar to the overnight culture used to inoculate the experiment at t=0. Indole production is low during exponential phase but increases during the transition to stationary phase.

Talk about the variation in the growth curve of Figure 4.3, why was it different etc. Under the control of the lac operon, might be different to what we would expect etc.

The two unambiguous conclusions from these experiments are (1) that plasmid pCA24N is capable of producing active tryptophanase and (2) that the TnaA content of pCA24N-containing cells varies considerably during growth, remaining low during exponential

phase and increasing strongly towards stationary phase. Indole production follows a similar pattern, and it is tempting to treat this as evidence that, as has been suggested previously, indole production is regulated primarily by the availability of TnaA protein. However, these data give only limited information about the regulation of tryptophanase expression under normal conditions. Plasmids such as pCA24N show strong copy number amplification when growth slows during the transition to stationary phase (229). It is therefore entirely plausible that in this experimental system tryptophanase production, and hence indole production, is being regulated by a copy number artifact. The results tell us little about what might happen when the tryptophanase gene is chromosomally located and under native regulation. It was therefore decided that further investigations should use GFP fused to *tnaA* in its normal location.

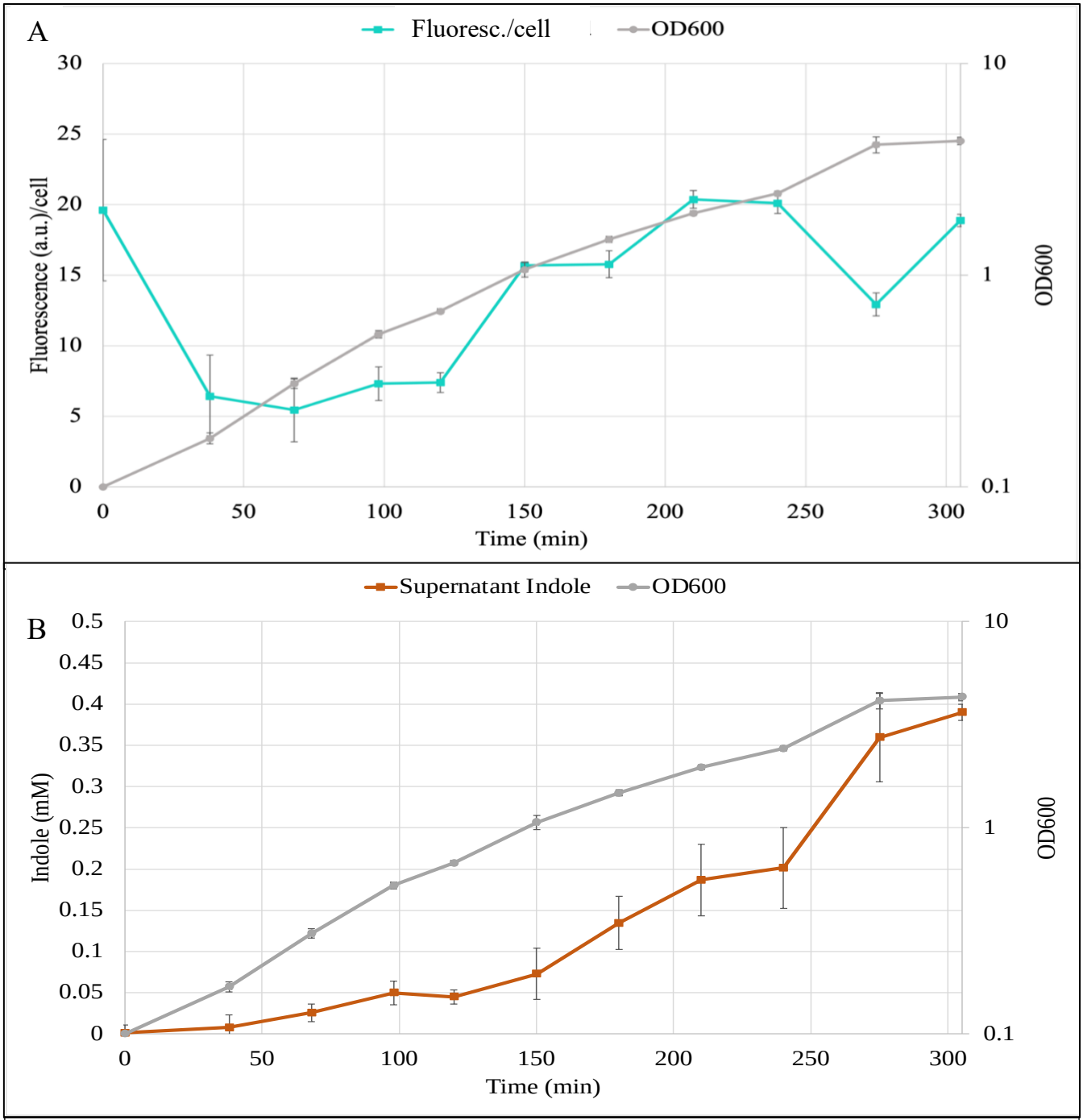


Figure 4.3. Fluorescence of TnaA-GFP over time in BW25113 *DtnaCAB pCA24N* cultures with corresponding supernatant indole concentrations. (A) Fluorescence of TnaA-GFP (at l=510 nm) in BW25113 *DtnaCAB* pCA24N cultures divided by the optical density (600 nm) to normalise for cell number. Cultures were inoculated and grown in LB medium plus additional tryptophan (0.5 mM) and sampled at 30 min intervals. The cultures were induced with 0.2% IPTG at time = 0 min. (B) The supernatant was assayed for supernatant indole using Quantichrom™ reagent. Data shown are the mean values for ± standard deviation for three independent repeats.

4.4 Tryptophanase-GFP expression from the chromosome

Previous studies concluded that the majority of TnaA-GFP was produced just before the transition to stationary phase, and that tryptophanase availability was the key determinant of the timing of the indole pulse (147). However, data presented in Chapter 3, and in the first part of this chapter, seemed to suggest that the earlier initiation of tryptophanase production does not yield an earlier indole pulse. To avoid the possibility of experimental artifacts arising from the tryptophanase gene being located on a plasmid, this section switches its focus to the regulation of tryptophanase under its native location on the *E. coli* chromosome. By using a chromosomal promoter, the experiments become more physiologically relevant, as tryptophanase expression is under the control of its native promoter. This approach also eliminates the variability caused by plasmid copy number, leading to more consistent and reliable data.

To test the hypothesis that less tryptophanase is expressed in exponential phase compared to the stationary phase, as previously reported in the literature (172), cell density was carefully controlled and the quantity of tryptophanase was measured. This approach ensures accurate comparisons between different growth phases, allowing for a clearer understanding of the regulation of tryptophanase expression. In a preliminary experiment (Figure 4.4) a culture of the chromosomal fusion strain *E. coli* BW25113 *tnaA-GFP* kan^R was sampled and scanned for tryptophanase-GFP fluorescence during exponential (OD_{600} = 0.5) and early stationary (OD_{600} = 2.0) phase. The exponential phase sample was concentrated four-fold so that both samples were scanned at the same cell density. Both cultures showed peak fluorescence at λ=510nm.; approx. 270 a.u. in exponential phase and approx. 240 a.u. in early stationary phase. The control strain used in this experiment, BW25113 Δ*tnaCAB* kan^R, showed low basal fluorescence at this wavelength. The striking conclusion from this experiment is that mid-exponential and stationary phase cells contain very similar levels of tryptophanase protein. This contrasts with previous assertions (147, 172) that substantial TnaA-GFP expression is delayed until the late exponential to stationary phase transition. However, GFP fluorescence only provides an estimate of the concentration of TnaA-GFP protein monomers in cells. The

enzymatic activity of tryptophanase in these two growth phases remains unknown and will be addressed in subsequent experiments.

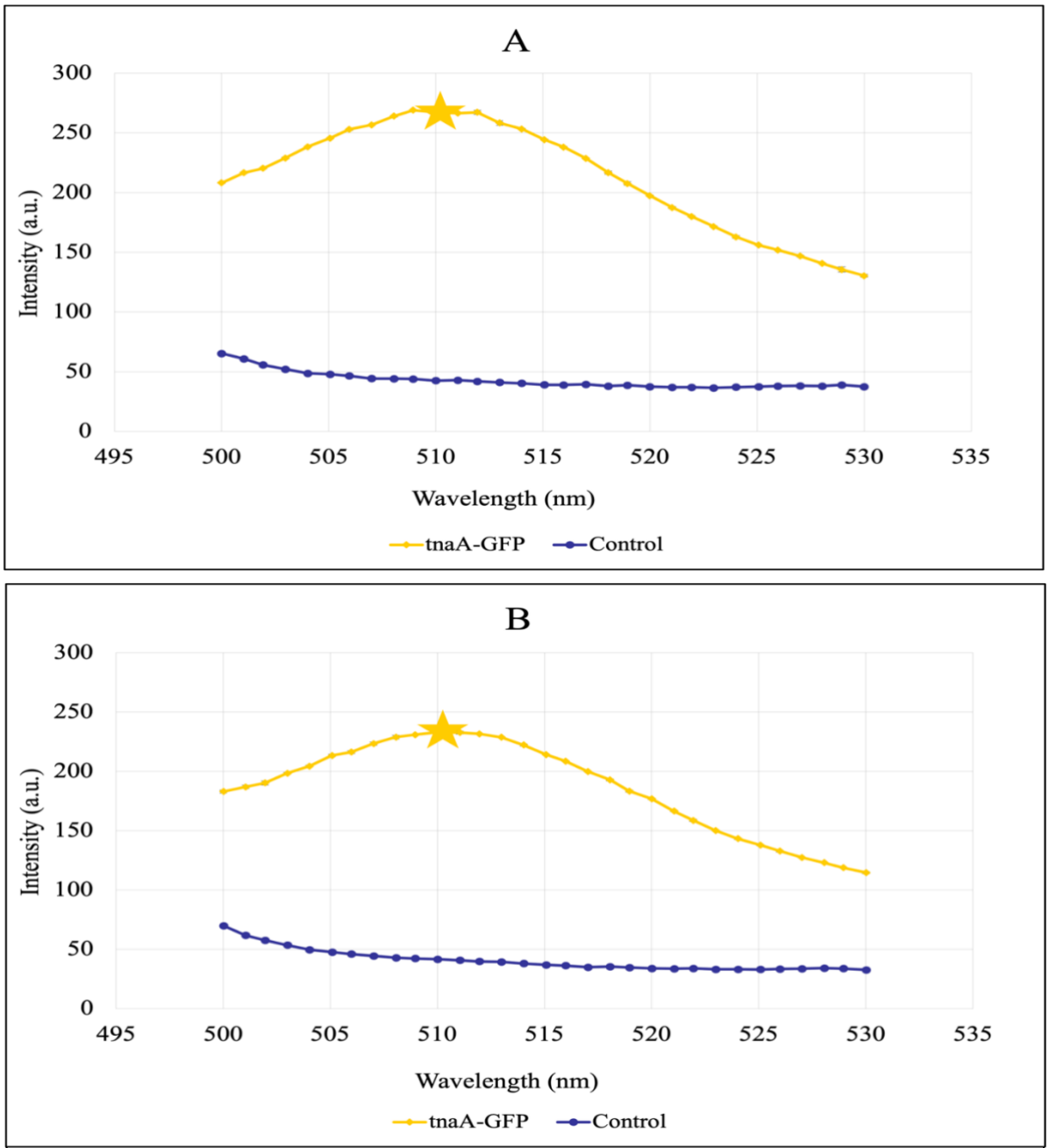


Figure 4.4. TnaA-GFP fluorescence measurements of *E. coli* cultures sampled during exponential or stationary phase. An overnight culture of BW25113 *tnaA-GFP* was inoculated at an OD_{600} = 0.1 in LB medium and grown at 37°C. (A) Exponential phase: a sample was taken at OD_{600} = 0.5 and concentrated to a final OD_{600} = 2.0. (B) Early stationary phase: a sample was taken at OD_{600} = 2.0. Samples were washed in PBS twice and the final density was OD_{600} = 2.0 in 1 mL of PBS. The cells were scanned for fluorescence in a fluorimeter at wavelengths λ=400-600nm. Stars indicate the peak fluorescence values. A culture of BW25113 *ΔtnaCAB kan*^{R} was used as a control. The data are the result of three independent repeats.

4.5 Tryptophanase-GFP activity in the chromosomal fusion throughout growth

Unexpectedly, the data presented in Fig 4.4 indicate that the amount of tryptophanase per cell may not vary substantially between exponential and early stationary phase in the chromosomal fusion strain. However, in this experiment, the tryptophanase content of a culture was evaluated at just two points in the growth curve (OD_{600} = 0.5 and 2.0). To obtain a more fine-grained measure of TnaA-GFP levels in the chromosomal fusion strain, fluorescence measurements were taken throughout growth. This chromosomal fusion experiment follows the same experimental protocol used for cells containing the TnaA-GFP plasmid (Figure 4.3). Data are presented in Figure 4.5.

Throughout the experiment, the value of GFP-fluorescence/OD_{600} was substantially higher than seen when TnaA was expressed from a plasmid (Figure 4.3). The fluorescence per cell (initially approx. 90 a.u./OD_{600}) decreased as the cultures entered exponential phase (OD_{600} = approx. 0.4), indicating that cells were dividing faster than TnaA-GFP was being produced. During late exponential phase and the transition to stationary phase (OD_{600} = approx. 0.8-1.5), the GFP-fluorescence/OD_{600} value remained constant (approx. 13 a.u./OD_{600}) indicating that the rate of production of TnaA-GFP had become the same as the rate of cell division. Subsequently, at an OD_{600} of approx. 2, GFP-fluorescence/OD_{600} began an increase that persisted until the end of the experiment, culminating in a fluorescence of 45 a.u./OD_{600}. This corresponds to approximately 70% of the fluorescence exhibited by late stationary phase overnight cultures. This is consistent with previous flow cytometry studies on BW25113 *tnaA-GFP* kan^R that demonstrated tryptophanase expression is greatest in late stationary phase (158).

This fine-grained study explains the previous surprising similarity of TnaA-GFP/OD_{600} at OD values 0.5 and 2.0 (Figure 4.4). The vertical red lines indicating these OD values in Figure 4.5, show that, despite considerable variation throughout the experiment, TnaA-GFP/OD_{600} was almost identical at these culture densities (approx. 10).

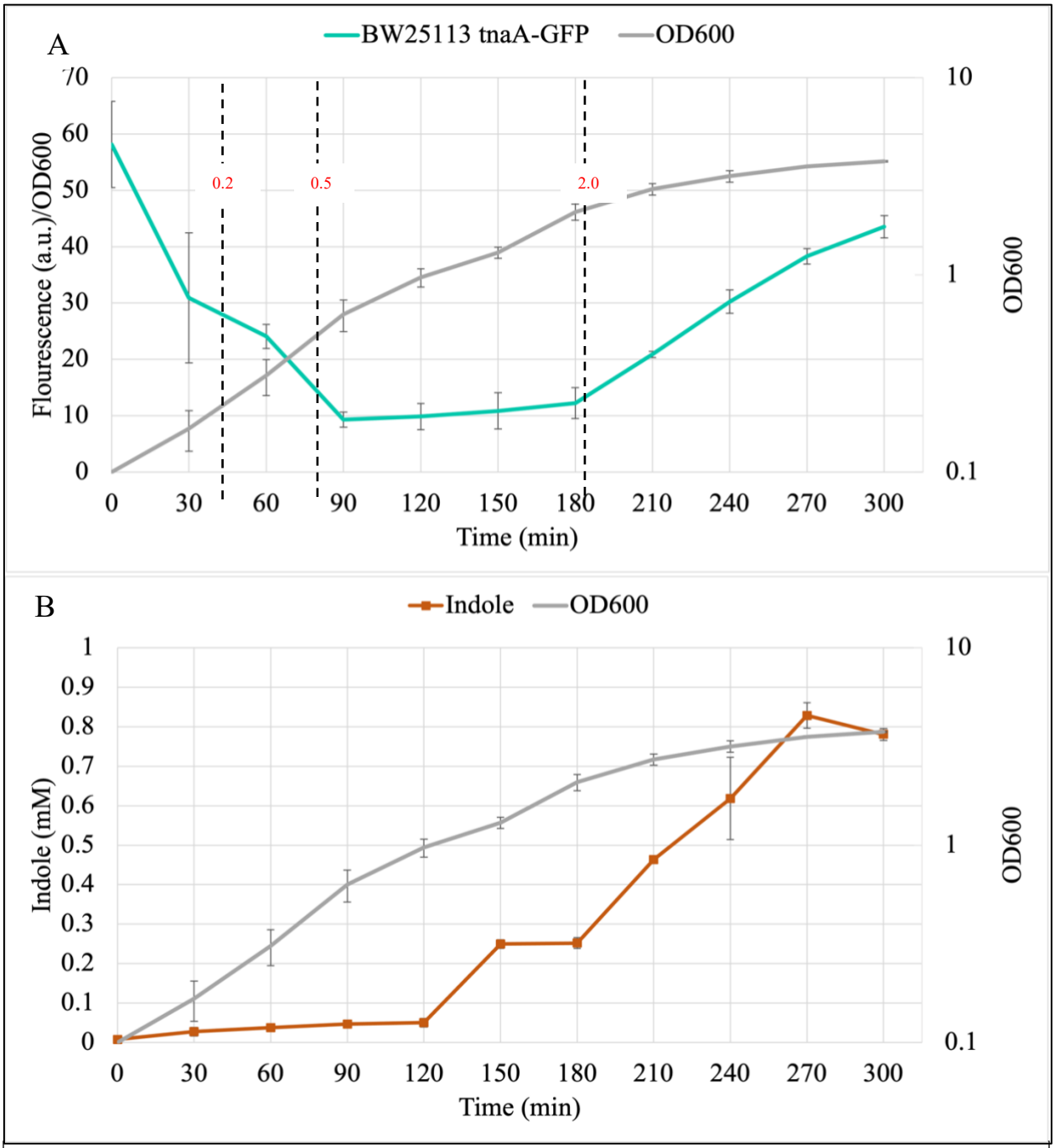


Figure 4.5. Fluorescence of TnaA-GFP per cell in BW25113 *tnaA-GFP kan^R* cultures over a period of 5 hours. Cultures were inoculated and grown in LB medium plus additional tryptophan (0.5 mM) and sampled at 30 min intervals. (A) The fluorescence (a.u.) of TnaA-GFP a λ=510 nm was measured and divided by OD_{600}. (B) The supernatant indole concentration was assayed using Quantichrom™ reagent. Data shown are the mean ± standard deviation for three independent repeats. Significant optical density values (600 nm) are depicted in red with a corresponding dotted line to show where they lie on the graph.

A significant feature emerges when TnaA-GFP/OD_{600} is compared with the indole concentration in the supernatant. Specifically, the indole concentration begins sharply to rise when the TnaA-GFP/OD_{600} is at a minimum. Indeed, indole continues to rise for 60 min before an increase in TnaA-GFP/OD_{600} becomes apparent. Since active tryptophanase is required for indole production, this clearly contradicts the previous view that up-regulation of *tnaA* transcription is responsible for initiating indole production and the intra-cellular indole pulse.

These data imply that the indole pulse could be triggered by a combination of factors, including both *de novo* synthesis and activation of pre-existing tryptophanase enzyme. To determine the details of post-translational tryptophanase regulation, it was necessary to conduct an enzymatic activity assay, rather than an assay that simply reports the amount of TnaA protein.

4.6 Tryptophanase activity measurement in exponential and stationary phase cells

Preliminary initial-rate assays of indole production were conducted to compare the enzymatic activity of tryptophanase in the TnaA-GFP chromosomal fusion strain at different stages of growth. Indole production rate was used as an indicator of enzymatic activity. The assay was performed according to the following methodology. *E. coli* BW25113 *tnaA-GFP* kan^R cells were inoculated (OD_{600} = 0.1) in LB medium containing an additional 0.5 mM of tryptophan. The cells were grown at 37°C, with shaking, for 8 hours. Two time points were chosen for measurement of the enzymatic activity of tryptophanase. Samples were collected at OD_{600} = approx. 0.2 and OD_{600} = approx. 2.0. In both cases, cells were washed twice with PBS and resuspended in PBS to an OD_{600} of 2.0. One hundred μl aliquots of resuspended cells were added to a 96-well plate and analysed in a SpectraMax iD3 plate reader. The plate reader was set to 37°C and the plate was shaken between measurements. A protocol was set up to add 3 μl of tryptophan (10 mM) to each well. The cells were shaken for 3, 8, 15, or 22 min, before Quantichrom™ reagent was added to the wells. After 1 minute of incubation, the mixture was measured at λ=565nm to ascertain how much tryptophan had been converted into indole (as a proxy for enzymatic activity). The concentration of indole in the sample was calculated using a

calibration curve. As an experimental control, *E. coli* BW25113 Δ*tnaCAB kan*^{R} cultures were subjected to identical treatment, and the measurements were subtracted from those of the wild-type cells. The rate reaction for the first time point for exponential and stationary phase cells was calculated as the change in concentration of indole (M) per unit of time (min). Only the first time point was used to calculate the reaction rate, as it was the most accurate due to the rate not following a linear curve, possible because of enzyme saturation or the enzyme existing in variable activation states. Each experiment was completed in biological triplicate.

The results of the assay are shown in Figure 4.6. From the rate of indole production, it appears the enzymatic activity of tryptophanase in *E. coli* cells during stationary phase surpasses that observed in the cells during exponential phase. Focusing on the initial rate of indole production, the stationary phase cell suspension reaches a reaction rate of 5.76 x 10^{-5} M/min, while exponential phase cell suspension reaches a reaction rate of 7.1 x 10^{-6} M/min (Figure 4.6).

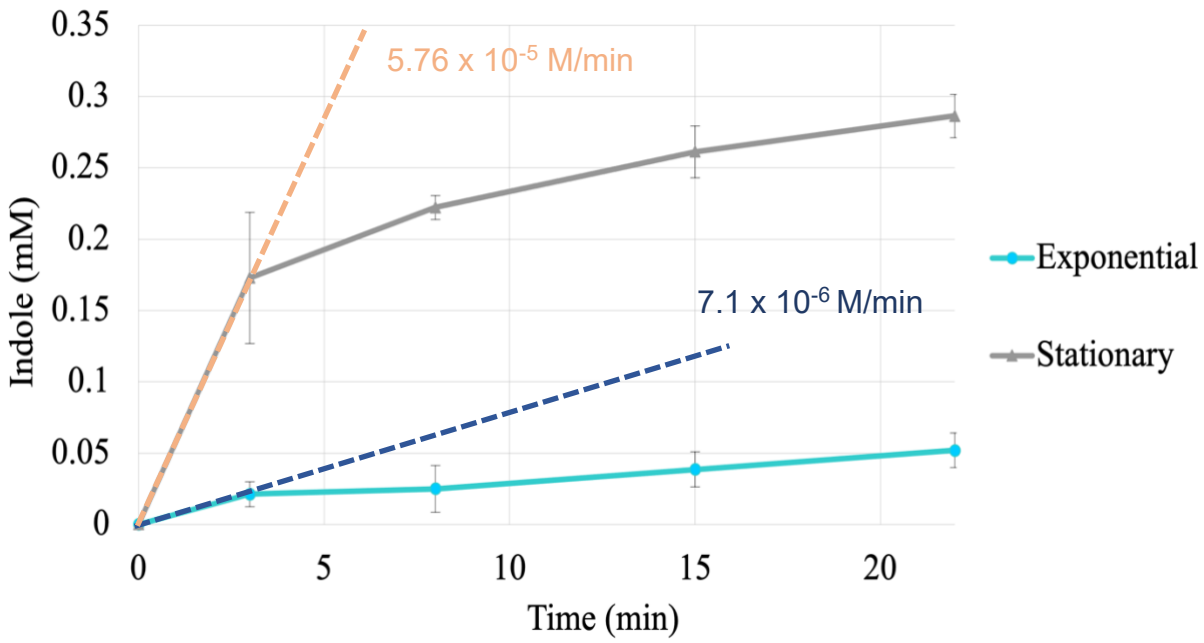


Figure 4.6. Indole production from tryptophan in *E. coli* BW25113 *tnaA-GFP kan*^{R} cells harvested at two different growth phases: exponential (OD_{600} = approx. 0.2) and stationary (OD_{600} = approx. 2.0). The indole production rate is compared at the two different growth phases as a proxy for the activity of tryptophanase. Cells were inoculated in LB medium with additional tryptophan (0.5 mM) and grown at 37°C until the relevant optical density was reached. Samples were adjusted to an OD_{600} of 2.0. The cells were added to a 96-well plate and tryptophan (0.3 mM) was added. Quantichrom™ reagent was used to assay the indole produced after each time point. The rate reaction line with its associated slope is shown for the first time point in exponential orange and for stationary phase in dark blue. Error bars show standard deviation for three independent experiments.

The previous experiment (Figure 4.6) used OD_{600} normalisation in an attempt to ensure that tryptophanase activity was being compared in equal numbers of exponential and stationary phase cells. This provides insight into relative enzyme activity only as long as the amount of tryptophanase in exponential and stationary phase cells is the same. Although the data of Figure 4.4 suggests that this is not an unreasonable assumption, a more direct comparison of enzyme activity would be achieved if the sample densities were normalised using TnaA-GFP fluorescence, rather than optical density.

An additional experiment was therefore conducted to compare the activity of tryptophanase in exponential and stationary phase, normalising the assay samples for GFP fluorescence, which more directly reflects enzyme content. An overnight culture of BW25113 *tnaA-GFP* kan^R was inoculated at an OD_{600} = 0.05 in LB medium with additional tryptophan (0.5 mM). Cultures were grown at 37°C, with shaking, until the desired optical density (600 nm) was reached. Samples were collected at OD_{600} = 0.2, 0.5, and 2.0. Samples were washed twice in PBS to remove any indole and resuspended in PBS. The samples were measured for GFP fluorescence and adjusted to a final value of 20 a.u/mL to ensure there were equivalent amounts of tryptophanase protein in each sample. Aliquots of cell suspension (100 μL) were added to a 96-well plate and tryptophan (0.3 mM) was introduced into each well. Quantichrom™ reagent was used to assay the indole produced after the cells had been shaken for 3, 8, 15, or 22 min.

Figure 4.7 shows little to no capacity for indole production by tryptophanase in early and mid-exponential growth phase. The rate reaction at the initial time points remained fairly consistent throughout the experiment. Therefore, the rate rection was calculated for the duration of the experiment. The reaction rates were as follows: 0 M/min for early exponential phase cells, 7.0×10^{-7} M/min for mid-exponential phase, and 6.7×10^{-5} M/min for early stationary phase cells. This indicates that the reaction rate for early stationary phase cells is 104.5 times faster than that of early exponential phase cells. This marked variation in enzyme activity indicates activation of the enzyme at some point during the transition from exponential to stationary phase.

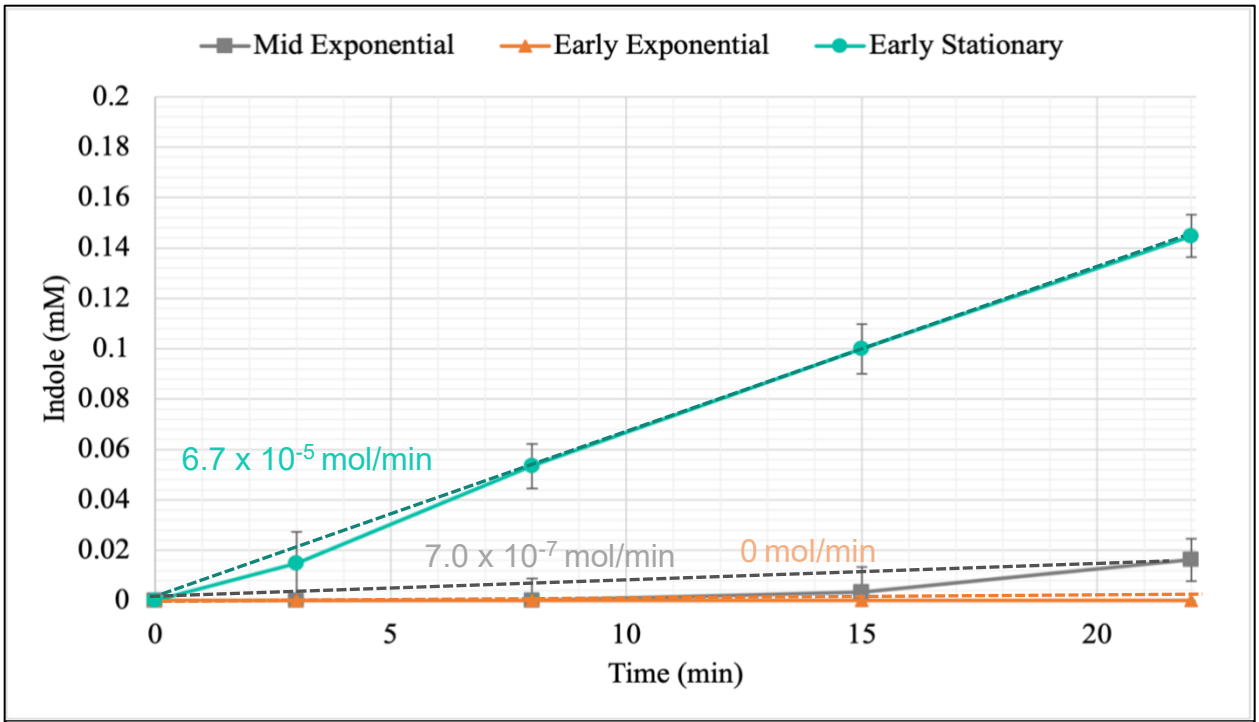


Figure 4.7. Indole production from tryptophan in *E. coli BW25113 tnaA-GFP kan^R* samples taken at three different growth phases, equalised for TnaA-GFP fluorescence. An overnight culture of BW25113 *tnaA-GFP* was inoculated in LB medium (with additional 0.5 mM tryptophan) at OD_{600} = 0.05 and grown at 37°C, with shaking. Samples were taken in early exponential phase (OD_{600} = approx. 0.2; orange), mid exponential phase (OD_{600} = approx. 0.5; grey) and early stationary phase (OD_{600} = approx. 2.0; green). The cells were washed twice in PBS and resuspended in PBS. The samples were adjusted to a final fluorescence reading of 20 a.u./1mL. Aliquots of the samples were added to a 96-well plate and tryptophan (0.4 mM) was introduced into each well. Quantichrom™ reagent was used to assay the indole produced. The rate reaction values are for early exponential, mid exponential, and early stationary phase in shown in turquoise, grey and orange, respectively. Error bars show standard deviation for three independent experiments.

4.7 Tryptophanase activity in rich vs poor environments

Although the mechanism of activation is unknown it may relate to metabolic changes associated with the transition from exponential to stationary phase. Evidence from past studies indicate that the depletion of nutrients results in the conversion of inactive tryptophanase into an active state (158, 206). A preliminary experiment to test the hypothesis that in the presence of glucose, enzyme will be less active and convert

tryptophan into indole at a slower rate was conducted. Late stationary BW25113 *tnaA-GFP kan*^{R} cells were washed twice with PBS. The washed cells were resuspended in PBS with or without 2% glucose. This higher concentration of glucose was used to test the effect of nutrient abundance on indole production and its relationship to cellular metabolism, gene expression, and stress responses. Aliquots of these samples (100 μL) were added to a 96-well plate in a SpectraMax iD3 plate reader. The plate reader was set to 37°C and the plate was shaken between measurements. A protocol was set up to add 3 μl of tryptophan (10 mM) to each well. The samples were shaken to for 3, 8, 15, or 22 min, and Quantichrom™ reagent was added to the wells. After 1 minute of incubation, the mixture was measured at λ = 565 nm to ascertain the how much tryptophan was converted into indole as a proxy for enzymatic activity. *E. coli* BW25113 *ΔtnaA kan*^{R} cultures were subjected to identical treatment as the wild-type cultures, and the recorded measurements were subtracted from those of the tryptophanase-producing cells, serving as an experimental control. The concentration of indole in each sample was calculated using a calibration curve. The rate reaction coefficient is calculated by the change in concentration (indole) over time (min). The experiment was performed in biological triplicate (Figure 4.8).

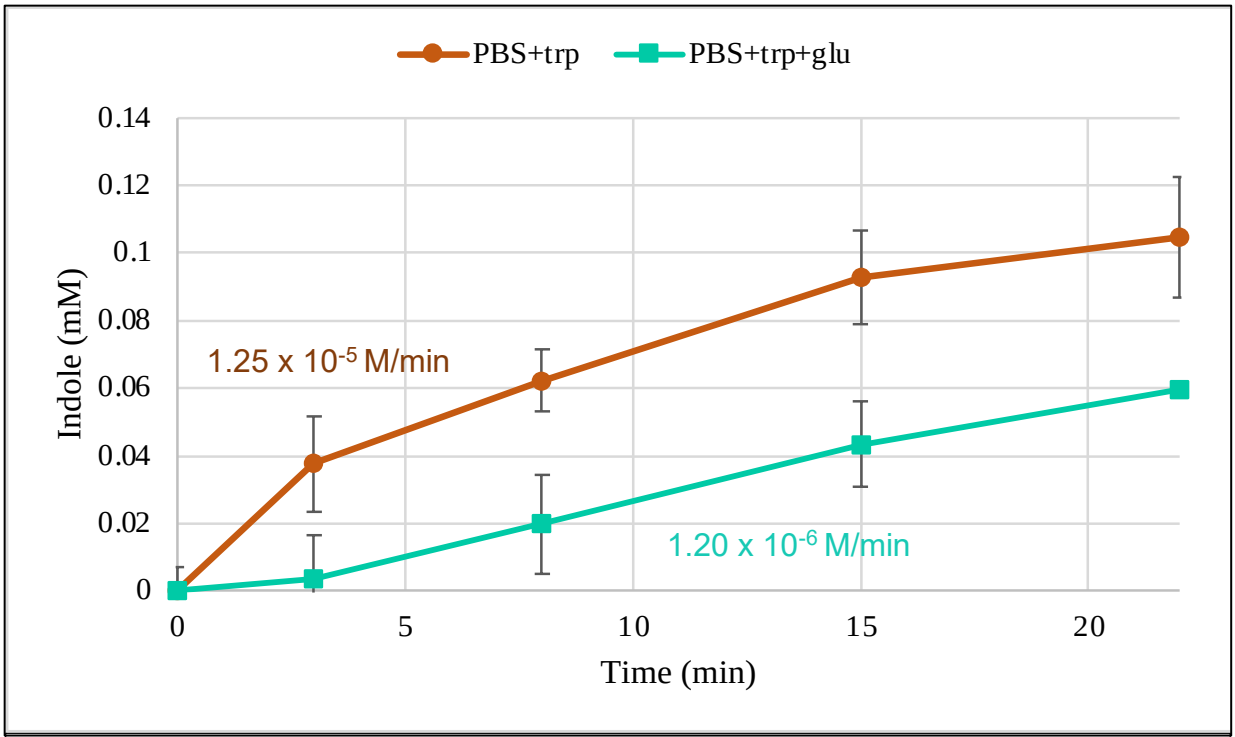


Figure 4.8. Indole production from tryptophan (0.3 mM) in *E. coli* BW25113 WT cultures in the presence (teal) and absence (orange) of glucose. Late stationary cells (OD_{600} = approx. 6.0) were washed in PBS twice and resuspended in PBS in the presence or absence of glucose (2%). Quantichrom™ reagent was added to the cells to assay indole production. Data are the average of at least 3 independent experiments. Error bars show standard deviation.

Indole was produced both with and without glucose, although the initial rate was lower in its presence. In the absence of glucose, the rate reaction was 1.25 x 10^{-5} M/min and in its presence the rate slowed down to 1.20 x10^{-6} M/min. Samples lacking glucose produced tenfold more indole in the first 3 minutes. After 22 minutes, the difference was less pronounced but samples without glucose had still produced twice as much indole as glucose-containing samples. This experiment supports the hypothesis suggesting post-translational regulation involving catabolite metabolism. In the presence of glucose, the enzyme shows reduced activity, leading to a diminished capacity for indole production from tryptophan. Conversely, when glucose is removed, the cells demonstrate an enhanced ability to produce indole. Notably, the reaction rate of tryptophanase within glucose-rich environments eventually increases, indicating possible scenarios where either an inhibitor is not able to fully supress enzyme activity, or the enzyme can operate

independently of an activator. To test this hypothesis further, I propose conducting this experiment again with different carbon sources like arabinose, glycerol, galactose and pyruvate, among others. This approach would aim to discern variations in activity influenced by different carbon sources and their subsequent metabolism.

4.8 Discussion

Previous studies, and experiments detailed in the preceding chapter of this thesis, imply that the indole pulse is subject to complex control. The data presented in the current chapter indicate that the timing of the pulse is a largely a result of post-translational regulation of the tryptophanase enzyme.

An early objective of this chapter was to determine the efficacy of tryptophanase production from a plasmid that had been used originally in an attempt to alter the timing of the pulse. The first experiment (Figure 4.1) confirmed that tryptophanase was produced from the plasmid but was relatively uninformative as, in the absence of cell density information, distinguishing an increase of tryptophanase within individual cells from an increase in the number of cells with a constant tryptophanase content was difficult. Furthermore, the cells were cultured in the non-optimal environment of a 96-well plate. This environment presents several challenges. The limited volume in each well restricts the availability of growth medium and nutrients, likely resulting in lower cell density. Additionally, compared to shaking flask cultures, there is limited gas exchange and oxygen availability, which can impact growth. Despite these limitations, this method allows for continuous monitoring and an ability to process a larger number of samples simultaneously.

Consequently, subsequent experiments took account of cell density and were conducted in a more conventional shake-flask apparatus (Figure 4.3). This provided insight into the concentration of plasmid-encoded tryptophanase in individual cells during lag, exponential and stationary phases. It showed that the greatest amount of tryptophanase per cell was present in lag phase, immediately after inoculation, followed by a rapid decrease as the culture entered exponential phase and an increase as the next stationary phase approached. The high concentration in lag phase reflects the state of the stationary phase cells used to inoculate the culture and is consistent with findings by Liu and Summers (2017) that tryptophanase continues to be produced throughout stationary phase. In flow cytometry experiments they found that cells in an

overnight culture contained the greatest amount of tryptophanase, as well as exhibiting high heterogeneity among cells.

However, despite similarities to previous reports, it is very important to remember that in Figure 4.3 tryptophanase is being produced from a plasmid, under the control of the *lac* promoter. Furthermore, this plasmid lacks *tnaB*, which is important for tryptophan import into the cell, as well as *tnaC* which is an important transcriptional regulatory element of the operon. Hence it is difficult to draw valid conclusions about the normal regulation of tryptophanase from this experiment.

In the light of these concerns, subsequent experiments were conducted in the TnaA-GFP chromosomal fusion strain. The data in Figure 4.5 depict a pattern of high tryptophanase per cell in lag phase (reflecting the state of the inoculating culture), a decline and plateau during exponential phase followed by an increase towards the next stationary phase. Strikingly, although the tryptophanase content per cell remains constant between 90 and 180 min, indole production increases strongly during this period. This finding counters a prior hypotheses that the rapid accumulation of indole during the transition from exponential to stationary phase was a result of increased tryptophanase production at that time (Gaimster and Summers 2015). These new data raise the possibility that the activation of pre-existing tryptophanase is important in triggering the start of indole production upon the transition into stationary phase. Having said this, it is important to note that in Figure 4.5 there is a clear increase in tryptophanase per cell after 180 min (OD_{600} = approx. 2.0) so it seems reasonable to suggest that both activation and increased production of tryptophanase may play their parts in triggering the pulse of indole production.

My evidence for post-translational regulation of tryptophanase aligns with previous experimental findings by Liu and colleagues, wherein it was shown that late stationary phase cells will produce indole when resuspended in tryptophan-containing PBS, but not in nutrient-rich LB medium (158). This observation suggests that the deactivation of tryptophanase is somehow caused by metabolic changes associated with a nutrient-rich medium. This is consistent with the timing of the indole pulse in LB medium. The

initial deactivation of tryptophanase caused by the nutrient-rich environment in fresh LB medium would be reversed as cells begin to exhaust the nutrients, leading to rapid indole production.

To probe deeper into the hypothesis of post-translational regulation of tryptophanase, experiments were designed to compare indole production in different growth phases, normalised for cell number or TnaA-GFP protein content (Figures 4.6 and 4.7). The data provide strong evidence that the enzymatic activity of tryptophanase is much higher in stationary phase than during exponential phase. However, despite some suggestion that tryptophanase activation is driven by low nutrients, the actual mechanism underpinning activation is not known. Although this evidence appears convincing, a potential caveat is the possibility of fluorescence not equating to quantity of enzyme in different growth phases, perhaps because of changes including GFP folding and maturation.

Foci was first associated with enzyme activity by researchers when they observed that GFP-tagged tryptophanase sometimes formed bright foci localised at one pole of the cell and was diffused throughout the cell (206). The researchers determined that enzyme activity is low when foci are formed, and the activity increases as the protein becomes more diffuse. They proposed that this might give a clue about the mechanism of post-translational regulation. They speculated that focus formation during exponential phase leads to occlusion of the enzyme's active site. A metabolic change associated with nutrient depletion triggers the dispersion of these foci, thereby creating an active enzyme. The final experiment described in this chapter (Fig 4.8) does give some tentative support to this metabolic switch idea. However, without extensive investigation of individual cells, it is difficult to say whether or not this mechanism is correct.

The work described in this chapter raises many questions. Why is the enzyme inactive in lag and exponential phase? What triggers the enzyme to become active upon the transition into stationary phase? How is glucose involved? How does the switch between diffuse tryptophanase and condensed foci come into this story? Is the indole

pulse a population-wide phenomenon and does every cell in a culture regulate indole production in the same way and at the same time? Exploring single cell tryptophanase production and regulation, alongside indole production, should provide answers to at least some of these outstanding questions.

Chapter 5: Mother machine experiments: single cell analysis

5.1 Introduction

So far in this thesis, tryptophanase activity and indole production have been studied exclusively in bulk culture. In this conventional microbiological approach, data is averaged across a population of cells. This is convenient within a laboratory setting and in previous chapters it has revealed heterogeneity between cultures. However, it fails to capture potential heterogeneity within an individual culture. For example, it is possible that the level and activity of tryptophanase varies from cell to cell. Indeed, it is unclear whether every individual cell experiences an indole pulse, or whether the timing is the same in every pulsing cell. In this chapter, by analyzing the concentration and activity of tryptophanase in single cells, cell-to-cell heterogeneity can begin to be explored.

Studies based on an averaging over all the cells within a culture also fail to capture important physiological outliers. Persister cells are a prime example of a rare cell type that is important to study due to their considerable clinical importance (174). The ideal approach to investigate individual cell heterogeneity and the presence of physiological outliers would be to analyse individual bacterial cells grown under the same conditions used previously to study bulk cultures. The approach developed in this chapter represents an attempt to approach this ideal scenario.

The mother machine (Figure 5.1) is a microfluidic platform specifically designed for high-throughput study of defined bacterial cell lineages over time in a dense and changing culture environment. The device is made from polydimethylsiloxane (PDMS) and fabricated from a patterned silicon wafer. The channels are made using electron beam lithography, which allows the fabrication of small channels, closed at one end, that are just wide enough to fit a single *E. coli* cell and long enough to accommodate approximately 10 cells, depending upon the stage of growth.

The “mother cells” (the cells at the closed end of each channel) can divide for many generations and their physiology and gene expression can be monitored. As the cells divide, daughter cells are ejected from the open end of their respective channels. Once

out of their original channel, these daughter cells join a wider, perpendicular channel where they are flushed away, eventually joining the main culture reservoir growing under conventional shake-flask conditions. In a closed loop, culture from the flask is returned to the chip through the wide feeding channel. The growth medium reaches the cells in each side channel by diffusion. This ensures that the cells under observation in the mother machine experience the same environment as cells in the culture flask, including the gradual depletion of nutrients, as they exit lag phase, divide in exponential phase, and finally re-enter stationary phase. The cell-containing channels located on both sides of the feeding tube are monitored by time-lapse microscopy where images are acquired every 3 minutes. The consistent flow of the medium and the physical constraints of the microchannel ensure that the mother cell remains in the same focal plane which allows for clear imaging. The cells are observed at a magnification of 40X and imaged for TnaA-GFP fluorescence in real time. The shake-flask culture vessel is separate from the microscope to prevent vibrations affecting image acquisition and is maintained at a constant temperature of 37°C. Tubing running from the culture reservoir to the outside of the incubator allows sampling of the culture for OD_{600} measurements and indole assay. The data shown in this chapter originate from a representative experiment, illustrating typical data and timings, consistent with the findings of three separate and independent experiments.

To meet the requirements of the particular experiment, novel features were introduced into the mother machine methodology (214). While typical mother machine experiments employ a one-way medium system, in this experiment a closed loop was developed to enable the continuous observation of cells throughout their natural growth phases. Additionally, an Erlenmeyer flask containing the bulk culture medium and a stirring bar was placed inside a 37°C incubator, in order to replicate as closely as possible, the environmental conditions that cells had experienced in the previous experiments described in this work.

The intracellular distribution of tryptophanase (foci *vs* dispersed) has been flagged as a possible cause of the activity or inactivity of the enzyme. It was therefore deemed important to observe this both qualitatively and quantitatively in the mother machine.

Learning models called SyMBac created by Georgeos Hardo from the Department of Engineering were used to allow the relevant software to identify the formation of foci within cells (216). To train the system, images of cells within the side channels of the mother machine were chosen at random, and foci were identified by the author of this thesis (see 2.9.4). This allowed the quantification and distribution of TnaA-GFP within the cells to be automated and large numbers of cells to be analysed rapidly. Furthermore, it provides information of the formation and dispersion of foci during the various growth stages of the culture. The lag time between the event (indole production) occurring in the reservoir flask and the time of the same event occurring in the mother machine was taken into account within the data (approx. 30 min). The lag time observed between events can be attributed to the length of the tubing connecting the chip to the flask machine. This represents a difference in the experimental setup between flask culture and the Mother Machine setup and should be considered when interpreting results.

5.1 Strain construction for mother machine experiments.

Strains used previously in this thesis were not entirely suitable for the mother machine, as they did not include an RFP reporter that is necessary to determine cell segmentation in the Mother Machine chip. *E. coli* strain *MI13* (MG1655 7740 *ΔmotA*) was modified to optimize it for this purpose. This strain expresses a mCherry RFP reporter under the control of the *rpsL* promoter. The RFP reporter is optimal for segmentation of the cells during division. The RFP serves as a marker to delineate the boundaries of individual cells, enabling for precise segmentation of the cells in each microfluidic channel. Further, it opens avenues for the use of other fluorescent markers, such as GFP in this thesis, whether in conjunction or separate form RFP, facilitating the study of diverse cellular processes. *E. coli MI13* was obtained from the laboratory of Dr Somenath Bakshi at the Department of Engineering. Primers described earlier in this thesis were used to PCR-amplify a *tnaA-GFP* kan^R cassette, tailed with sequences flanking *tnaA* from *E. coli* MG1655. To integrate the *tnaA-GFP* kan^R amplicon into MI13, the helper plasmid, pSLTS, was first introduced into *E. coli MI13* using electroporation and selecting for carbenicillin resistance. The cells containing the plasmid were induced with

arabinose, as pSLTS expresses the lambda Red recombinase under the control of the arabinose-inducible P_{araB} promoter. After induction, the cells were made electrocompetent and the tryptophanase-GFP cassette was introduced by electroporation. The cells were recovered in LB medium at 30°C, as pSLTS replication is temperature sensitive, and sufficient time was allowed for recombination between the cassette and the bacterial chromosome. The culture was plated on kanamycin and carbenicillin at 30°C, and surviving colonies were re-streaked onto plates containing either antibiotic to look for MI13 colonies that had kanamycin resistance without the carbenicillin plasmid. Genomic DNA was extracted from kanamycin resistant and carbenicillin sensitive (pSLTS$^-$) colonies, and the region-specific DNA was sent for sequencing. The validated strain was designated *MI13 tnaA-GFP kan^R*.

5.2 The mother machine experiment

The mother machine microfluidic system is shown in Figure 5.1. The microscope, bulk culture reservoir and microfluidic device are isolated in a 37°C degree chamber, and the flask contains a magnetic spinner to oxygenate the medium. A colony of MI13 *tnaA-GFK kan^R* was cultured overnight in 5 mL LB with additional 0.5 mM tryptophan and 0.085% pluronic fluid, and grown for 16 hours, with shaking, at 37°C. A 1 mL aliquot of the overnight culture was centrifuged at 4000 rpm for 5 min and the majority of supernatant (900 μL) was discarded. The cells were resuspended in the remaining 100 μL of liquid to give a viscous mixture for feeding the cells into the microchip and channels. Using a micropipette tip, the cells were introduced into the microfluidic chip and the parallel side-channels using osmotic pressure. An Erlenmeyer flask containing 100 mL of LB with additional tryptophan (0.5 mM) was inoculated with the overnight culture to a final OD_{600} = approx. 0.05 and incubated at 37°C, agitated with a stir bar. Medium was pumped between the bulk culture and the mother machine at a constant flow rate of 15 μL/min. The cells in the microfluidic chip were photographed approx. every 3 minutes using phase contrast magnification 1.49NA Plan Apo phage contrast at 100X magnification with an oil objective (as set-up by Georgeos Hardo).

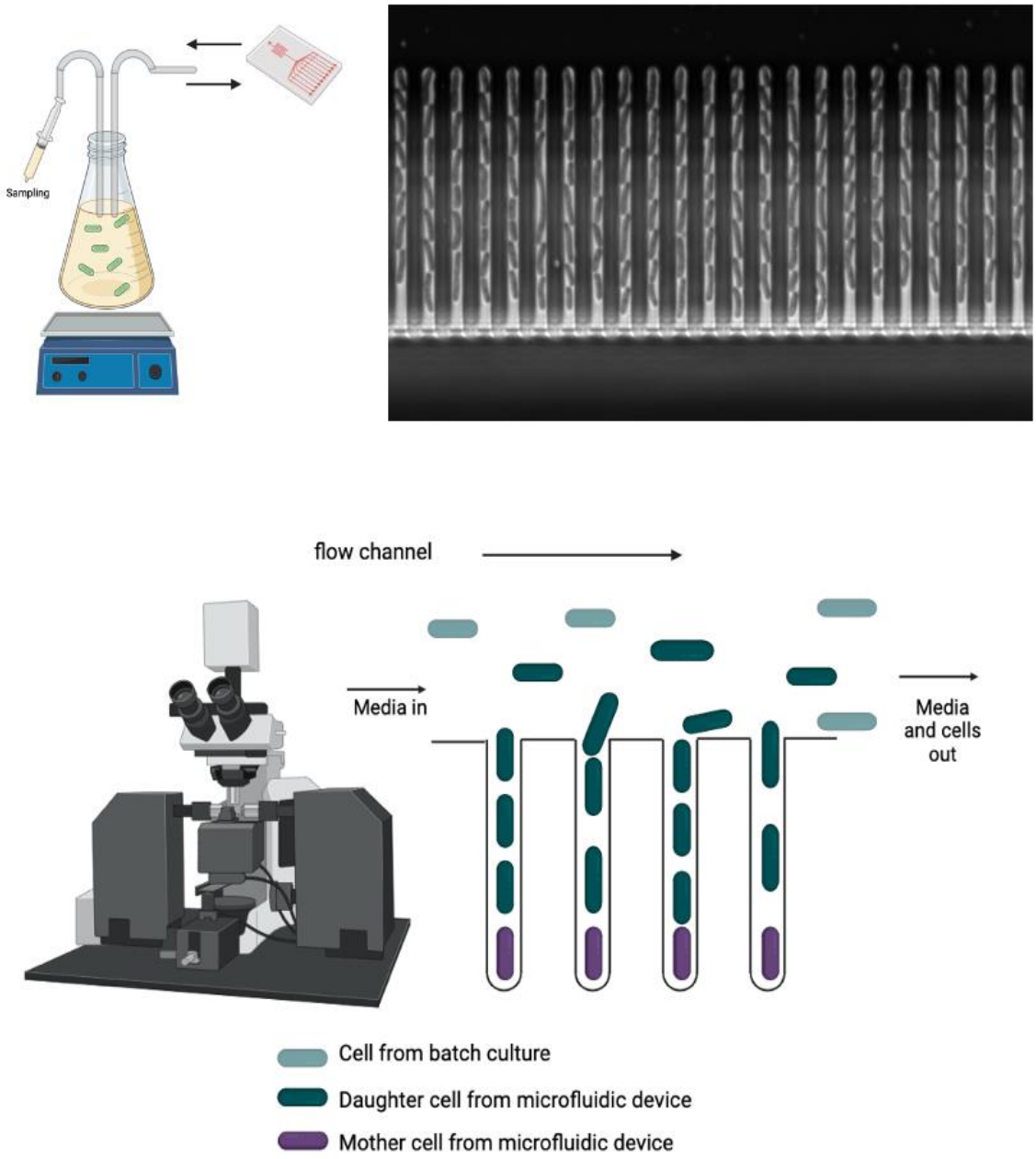


Figure 5.1 The mother machine experimental platform. Overnight cultures are fed into a microfluidic chip where osmotic pressure pushes bacterial cells into individual side channels. The cell at the end of the channel becomes the mother cell (purple), and this cell is monitored for physiology and gene expression throughout the experiment. The displaced daughter cells (dark turquoise) are fed back into the culture reservoir *via* the main flow channel. A flow channel also feeds culture from the reservoir back into the microfluidic device. The apparatus is therefore a closed loop system and the same medium conditions are experienced by cells in the reservoir and in the side channels of the microfluidic device. *Figure constructed using BioRender.*

They were also photographed and quantified for GFP fluorescence at its wavelength 485/510 nm (ex/em). The cells are monitored for approx. 6 hours until they reached mid stationary phase, at which point the cultures were too dense to get readable data.

5.3 Visualizing single cells throughout growth

In addition to collecting quantitative data, the mother machine allows visualization of cellular growth, tryptophanase content and distribution, offering insights which have previously been unattainable in this thesis and most indole research. Figure 5.2 illustrates the morphology and TnaA-GFP fluorescence of the cells in the mother machine throughout various growth phases.

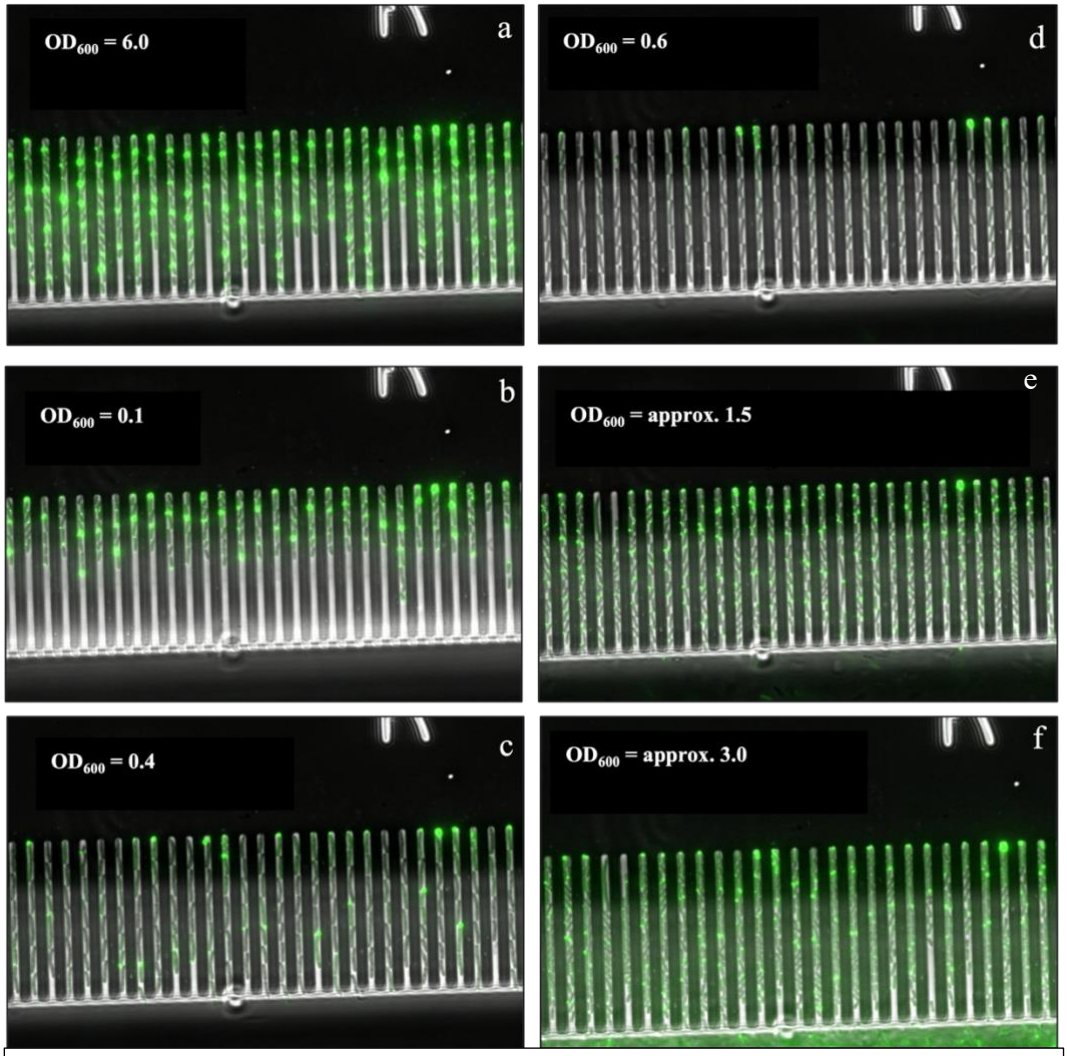


Figure 5.2. MI13 *tnaA-GFP kan^R* cells are shown in channels at significant optical densities (panels a-f) throughout growth. Green represents TnaA-GFP fluorescence within the cells.

Panel a in Figure 5.2 portrays the variation in fluorescence among the deep stationary phase cells that were introduced into the mother machine. The majority of the cells are bright green reflecting the accumulation of TnaA-GFP over time in the overnight culture. Once new medium is flushed through the mother machine, the cells enter lag phase. This is shown in Figure 5.2b. Several cells were flushed out of the channels when the new medium was introduced. There remains significant heterogeneity among the cells. Discrete, concentrated spots of green fluorescence are visible, along with more evenly distributed fluorescent green. The difference between TnaA-GFP in focussed and dispersed states is illustrated more clearly in Figure 5.3. Foci can be a variety of different sizes but always localize at one end of the cell (Figure 5.3a). Dispersed TnaA-GFP is more evenly distributed throughout the cell (Figure 5.3b).

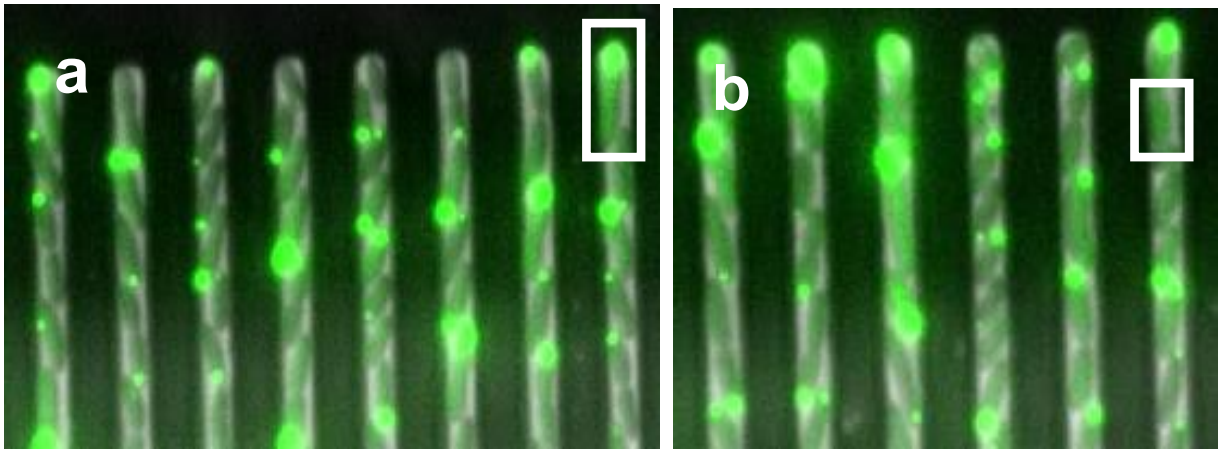


Figure 5.3 The panels show MI13 *tnaA-GFP kan*^{R} cells in deep stationary phase that have been loaded into the mother machine microfluidic chip. In panel a, the white rectangle is a cell that contains concentrated fluorescence in a spot, called a focus. One the panel b, the white rectangle shows a cell with dispersed TnaA-GFP in the absence of a focus.

In Figure 5.2c, the cells are in early exponential phase. There is a notable change in morphology, as the cells elongate and double in size before dividing. During this period the cells do not appear to produce more TnaA-GFP, so the TnaA-GFP content is successively diluted. The accompanying dilution of the foci is further illustrated in Figure 5.4. A focus generated during overnight culture before loading of the mother machine

is visible in lag phase, but as this cell divides, the focus shrinks. After several generations, many of the original mother cell foci are no longer apparent.

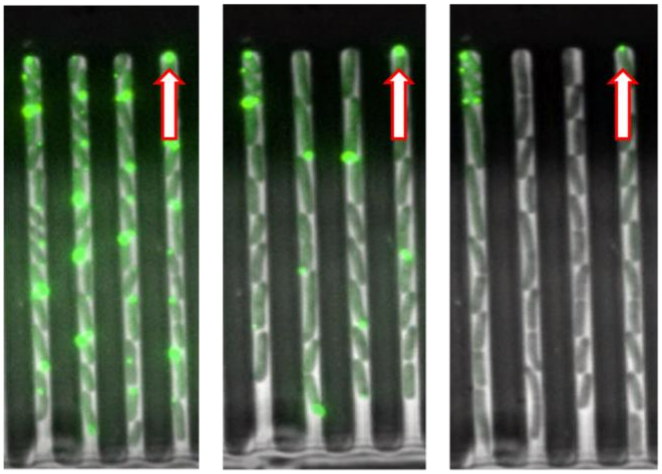


Figure 5.4. A mother cell with a foci containing TnaA-GFP that appears to dilute over time from lag phase to mid-exponential. The chosen mother cell is identified by a red arrow.

As growth proceeds, cells show reduced TnaA-GFP fluorescence and lack foci. This is especially prevalent in mid-exponential phase, as seen in Figure 5.2d. The few foci that are present are seen in original mother cells which appear not to have divided or decreased TnaA-GFP fluorescence since the experiment began.

As cells approach stationary phase TnaA-GFP fluorescence greatly increased and many of the cells develop foci. This is seen in Figure 5.2e. In any given channel, 80-90% of the cells contain foci that vary in size. As the cells enter stationary phase (Figure 5.2f), there is a reduction in cell size leading to increased crowding within the channels. Further, numerous foci have dispersed, and the majority of the cells show diffused TnaA-GFP.

Another significant benefit of visualizing cells throughout their growth, is the ability to uncover behaviours that might otherwise be obscured by the averaging of quantitative data. There were a few mother cells that were observed never to have divided or lost their TnaA-GFP. One such cell is outlined in Figure 5.5. This mother cell is constant in

size and remains bright green throughout the growth experiment over the course of 6 hours. This may be a persister cell, as it contains TnaA protein, but remains quiescent for a long period of time.

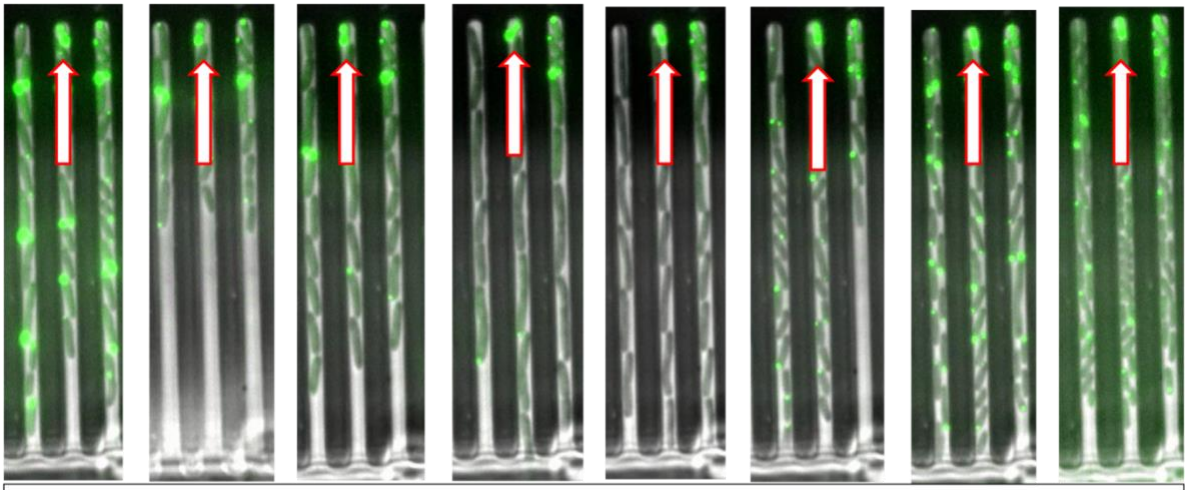


Figure 5.5. A mother cell observed throughout a growth cycle that appears to not divide or lose any of its TnaA-GFP. The chosen mother cell is identified by a red arrow.

5.4 Population dynamics from single cell data

To provide a more comprehensive exploration of trends in TnaA-GFP production and distribution in microscope images captured from the mother machine during the growth cycle, I examined the quantitative data.

Georgeos Hardo extracted the data using custom python scripts with images covering cell size, time to division, and fluorescence intensity. The collected data was then pooled from all the channels and several statistical tests were conducted. This data is presented graphically below and analysed to provide a more comprehensive understanding of TnaA-GFP production and distribution.

Panel A of Figure 5.6 displays optical density measurements for the flask (reservoir) culture. Given the constant circulation of culture between the flask and the chip, this also represents the growth phase of cell culture within the microfluidic device. The kinetics correspond closely with data from conventional shake-flask culture seen in Chapters 3 and 4 of this study. The growth curve follows the same shape, with a

prolonged transition period from approx. 100 min to 250 min, during which the cells plateau in stationary phase. Additionally, indole production is delayed, beginning to increase around 225 min and plateauing by the end of the experiment. Despite the final indole concentrations differing (approximately 0.8 mM compared to approximately 0.65 mM), the indole concentration in the mother machine experiment has not fully plateaued and is likely to continue rising slowly until reaching 0.8 mM. This gives confidence in the validity of results obtained from the mother machine and suggests that the apparatus shown in Figure 5.1 can effectively reproduce the conditions of the conventional shake-flask assays. The flask culture, inoculated at an initial OD_{600} of 0.05, is monitored until the final measurement is taken at an OD_{600} of approx. 4.5. The growth trajectory adheres to a normal pattern: lag phase occurs from 0-30 min, exponential growth between approx. 30 – 180 min and, after approximately 230 min, the culture transitions into stationary phase. The concentration of indole in the supernatant of the flask culture is presented alongside the optical density measurements. The supernatant indole concentration starts to rise in late exponential and early stationary phase until it peaks around 0.6 mM. This mirrors the pattern of indole production seen in the supernatant of conventional shake-flask culture and described in Chapter 3.

To display the indole production rate in the flask culture, the derivative of the supernatant indole concentration was plotted with respect to time (Panel B of Figure 5.6). This gives a good indication of the intra-cellular indole production rate at any given time since the rate of accumulation of indole in the supernatant depends upon the intracellular indole concentration (147). When the indole production rate reaches a plateau, it is likely that the maximum rate of indole export from the cells has been reached. This plateau, at approx. 250-300 min, therefore, corresponds to the timing of the intra-cellular indole pulse. This is a similar timing to direct measurements of the pulse described in Chapter 3. Subsequently, from approx. 300 min, the indole production rate drops substantially, most likely due to the depletion of tryptophan in the medium, as proposed by Gaimster and colleagues (147).

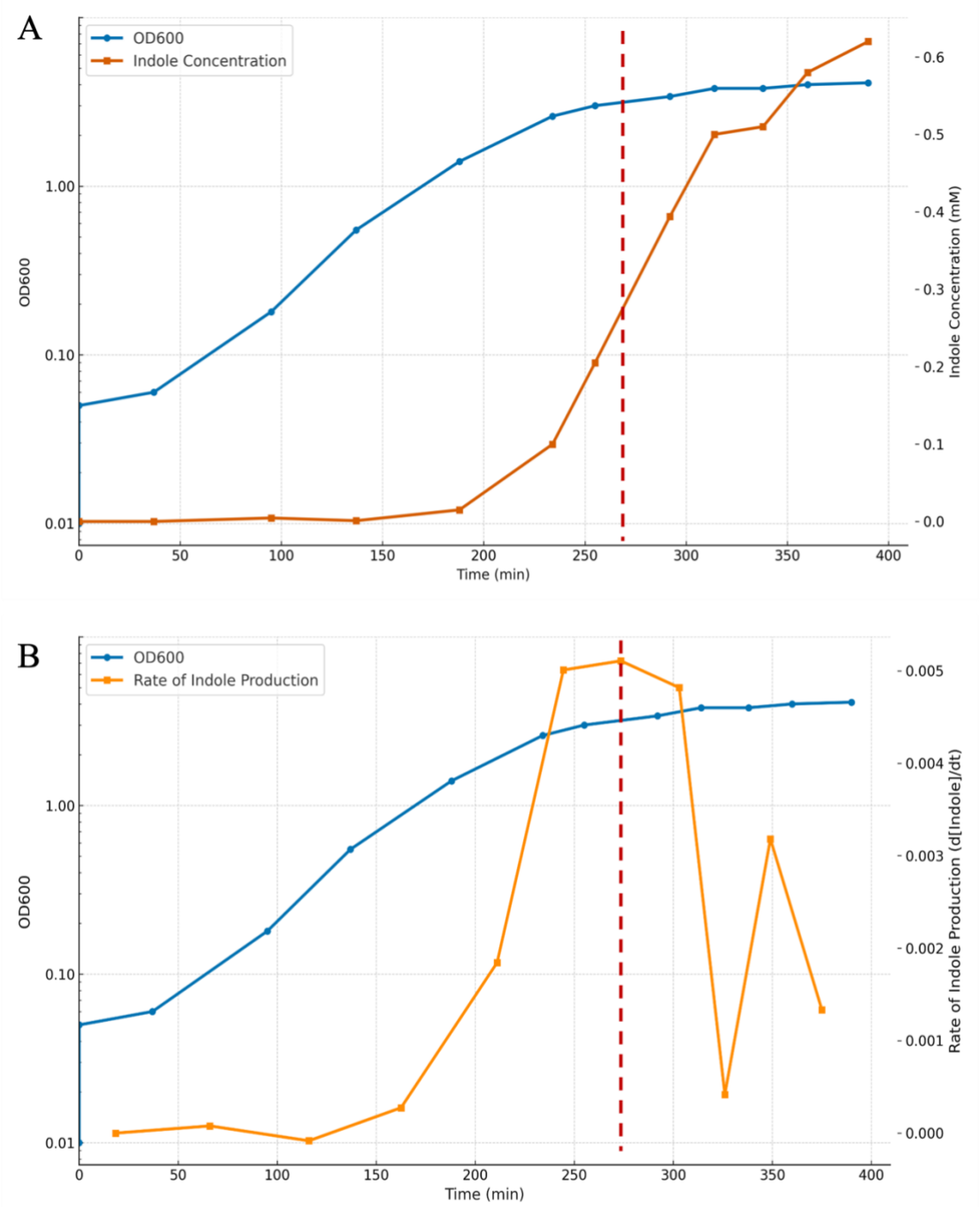


Figure 5.6. Cells from an overnight culture of *E. coli* MI13 *tnaA-GFP kan*^{R} were added to the mother machine microfluidic chip and the associated Erlenmeyer flask containing LB medium with an additional 0.5 mM tryptophan and pluronic fluid (0.085%). The flask culture was inoculated at an optical density (600 nm) of 0.05. The medium in the culture flask was stirred and incubated at 37C for approx. 380 min. Samples were taken from the Erlenmeyer flask every 30 minutes to measure optical density and indole production. A: Growth curve of culture in the Mother Machine apparatus and indole concentration of the flask culture over time. B: Growth curve of the same culture with the rate of indole production plotted at each time point. The dotted red line indicates the peak rate of indole production in the culture at an optical density = approx. 3.0. Data shown is a representative experiment of three independent experiments.

In Figure 5.7, the TnaA-GFP fluorescence in the mother machine (mean ± standard deviation, n = approx. 15 000 cells) at each time point has been added to the growth and indole production rate data presented already in Figure 5.6B. The trends in the single-cell data aligns with TnaA-GFP fluorescence measurements in conventional shake-flask culture experiments (Chapter 4). The mean fluorescence of the cells is high in lag phase, at approximately 1000 a.u. The fluorescence intensity then drops off and tryptophanase *per* cell declines as the cells divide during early exponential phase. It reaches a plateau at approx. 250 a.u. where it remains constant during mid-exponential phase (100-170 min). In late exponential phase (200-270 min), as cells transition into stationary phase, there is an increase in fluorescence to approx. 600 a.u. While the general trend in mean TnaA-GFP fluorescence matches data seen previously, it is important to note the substantial heterogeneity in the data, most likely due to phenotypic variation among the cells. The standard deviation in fluorescence intensity is largest in lag phase but remains high throughout the experiment.

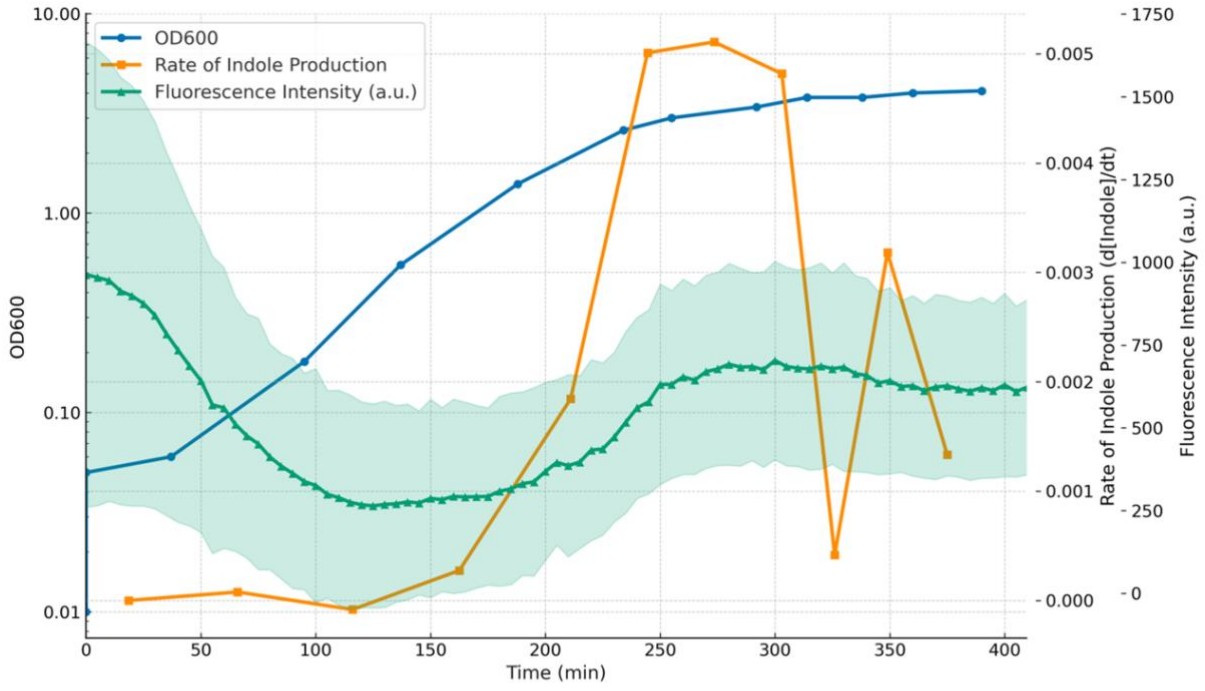


Figure 5.7. Mean fluorescence intensity of GFP in *E. coli* MI13 *tnaA-GFP kan*^{R} cells plotted over time in the mother machine microfluidic device. The green line shows the mean fluorescence intensity (a.u.) of the mother machine cells throughout time with the standard deviation at each point (shading). OD_{600} and indole production rate are shown in blue and orange, respectively. Data shown is a representative experiment of three independent experiments.

The heterogeneity of tryptophanase fluorescence, while large, is not representative of most of the population but, rather, indicates the presence within the culture of a relatively small sub-population of highly atypical cells. When the estimate of variation is changed from standard deviation to a 95% confidence interval using standard error, much of the heterogeneity is removed (Figure 5.7). Some 95% of the population now lies within the green shading. The general trend of mean fluorescence is unaffected, but the data implies that the majority of heterogeneity arises from a small subset of outlier cells, which exhibit strongly elevated or depressed tryptophanase production in contrast to the majority of the population. These outliers represent less than 5% of the data, but show the subpopulations of cells with different tryptophanase quantities (230).

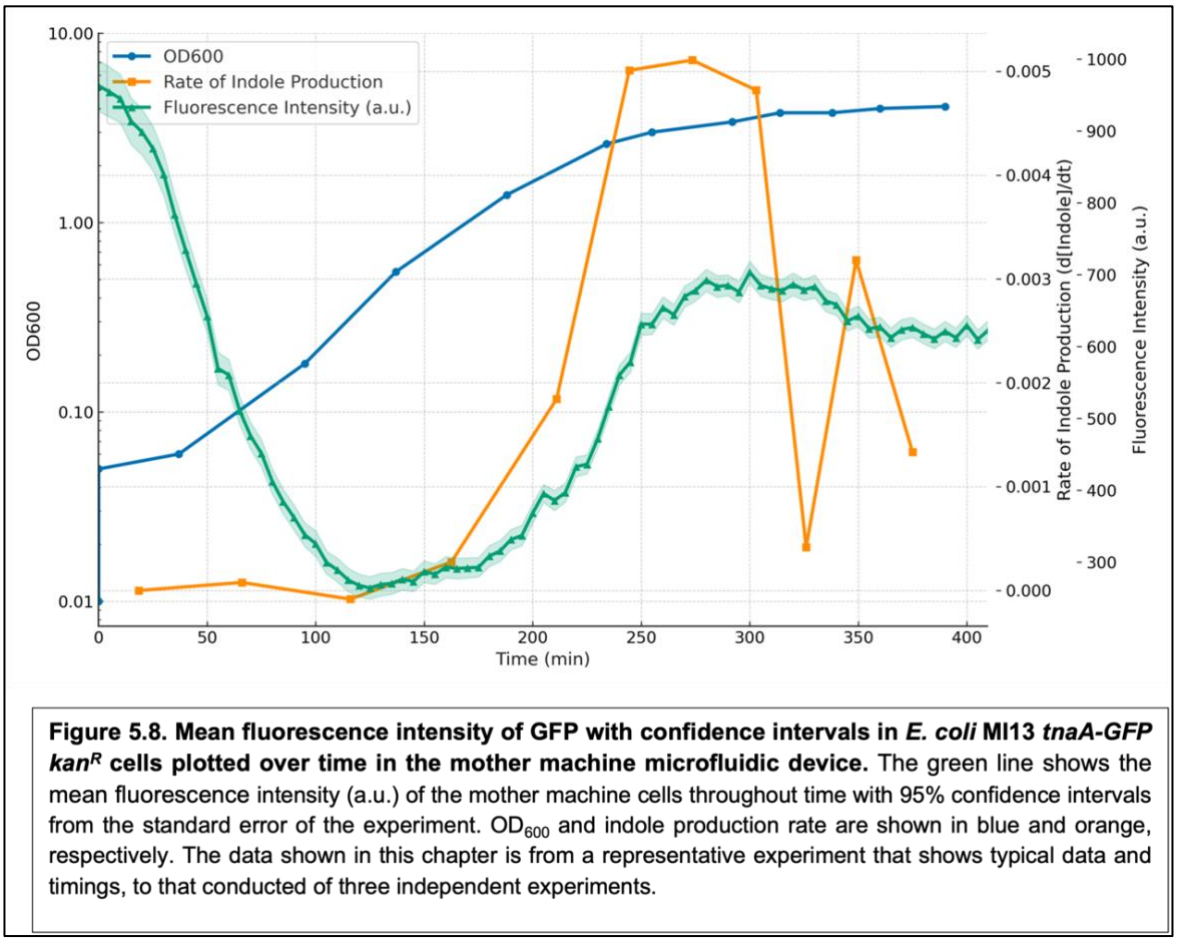


Figure 5.8. Mean fluorescence intensity of GFP with confidence intervals in *E. coli* MI13 *tnaA-GFP kan^R* cells plotted over time in the mother machine microfluidic device. The green line shows the mean fluorescence intensity (a.u.) of the mother machine cells throughout time with 95% confidence intervals from the standard error of the experiment. OD_{600} and indole production rate are shown in blue and orange, respectively. The data shown in this chapter is from a representative experiment that shows typical data and timings, to that conducted of three independent experiments.

Inspecting this clearer representation of the dataset, it is evident that the rate of indole production rises prior to the increase of TnaA-GFP that is seen during the transition from exponential to stationary phase. Up-regulation of tryptophanase lags indole production by approx. 30 min. Previous analysis of indole production dynamics (147, 172) suggested that a rise in tryptophanase production was directly responsible for the indole pulse. However, the present data support the view that there may be more than one trigger for the indole pulse. In Chapter (Figures 4.5 and 4.6), I proposed that activation of pre-existing tryptophanase enzyme might explain the sudden increase in indole production. Figure 5.8 supports the theory that enzyme activation, as well up-regulation of tryptophanase expression during the transition from exponential to stationary phase, may play a part in triggering the indole pulse.

Previous studies suggested that the dispersion of tryptophanase foci within *E. coli* cells was possibly associated with enzyme activation (206). To evaluate changes in the cellular distribution (dispersed *vs* foci) of tryptophanase in *E. coli MI13 tnaA-GFP*, images of individual cells were evaluated throughout the experiment. A training platform was used to generate synthetic data phase contrast and fluorescence data to train an Omnipose deep learning model that could detect the segmentation of cells and presence and absence of foci (215, 216).

Figure 5.9 shows the percentage of focus-containing cells throughout the growth experiment (red line). These data are superimposed upon the data for TnaA-GFP fluorescence (green), optical density (600nm) (blue) and indole production rate (orange). Immediately after culture inoculation (t=0), about 20% of the cells contain a focus and the mean cellular fluorescence is high. This reflects the situation in the late stationary phase culture used for inoculation. As the culture moves into exponential growth, both the mean fluorescence and the percentage of cells with foci decline sharply and remain low during exponential phase.

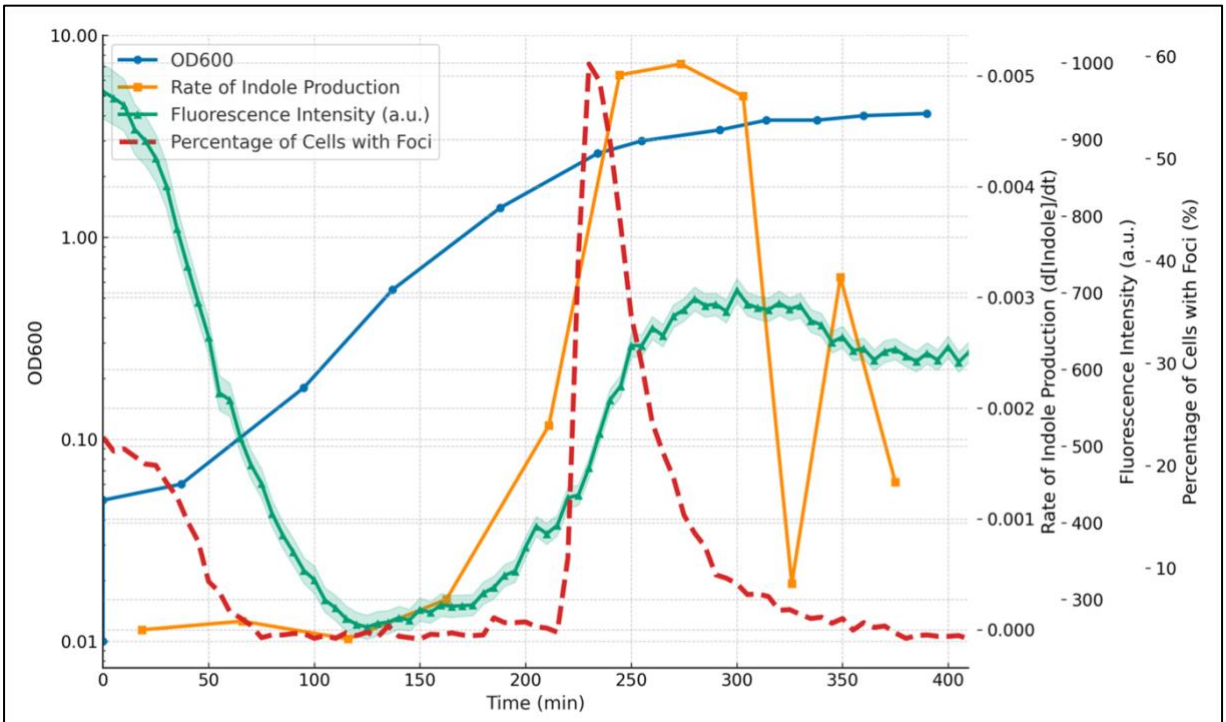


Figure 5.9. Percentage of *E. coli* MI13 *tnaA-GFP kan*^{R} cells that contains a focus at each time point. Data shown is a representative result selected from 3 experiments.

During the transition to stationary phase (OD_{600} = approx. 2.0), both the mean TnaA-GFP fluorescence and the percentage of cells with a focus climb strongly. At the peak, over 60% of cells contain a focus. From approx. 230 min, while the increase in TnaA-GFP fluorescence continues to rise, there is a very steep decline in the proportion of cells with foci. This coincides with a strong increase in the rate of indole production. At the time of peak indole production (approx. 270 min) approx. 10-20% of the cells contain a focus, while 90% display dispersed TnaA-GFP. It is striking that the appearance and subsequent dispersion of foci correlates closely with the time of the indole pulse. The activity of TnaA-GFP (indicated by the rate of indole production) increases as the foci disappear, suggesting that the foci may represent inactive, or less active, forms of the enzyme. However, this analysis has limitations due to the mother machine's inability to distinguish foci smaller than 0.06 microns, making it difficult to differentiate between small foci and noise in the data. According to the images in Figure 5.10, several foci around 150 min are not detected by the mother machine. Qualitative data suggests that the foci begin to appear between 150-175 minutes, increase until peaking at 225

minutes, and the diffuse rapidly. To validate these findings, a camera with smaller pixels or a super-resolution system would need to be used throughout the growth period.

When interpreting correlations among the various parameters depicted in Figure 5.9, it is important to recognise the possibility of a systematic error in counting the percentage of cells with foci. Specifically, the algorithm that identifies the foci is likely to miss the formation of small foci. This is evident in images from the experiment, that show the formation of very small foci during earlier logarithmic stage (Figure 5.10). However, it is likely to be able to identify cells without foci and would not affect the sudden dispersion of foci that is associated with maximum indole production.

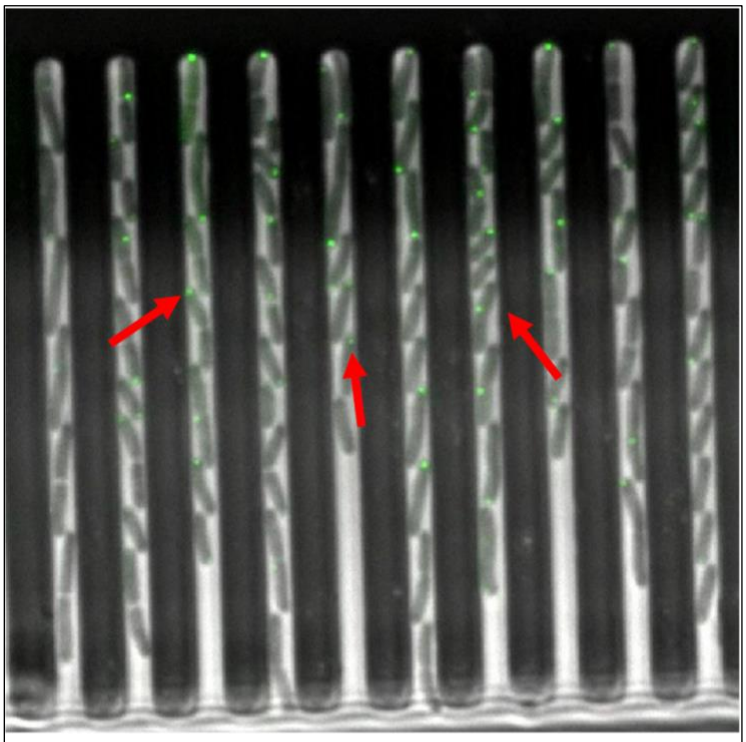


Figure 5.10. MI13 *tnaA-GFP kan*^{R} cells are shown in channels during exponential phase, small foci appear in numerous cells, as identified by the red

5.5 Single cell data analysis in late stationary and lag phase

Inspection of the images captured during the mother machine experiment demonstrates that the cells within the broth culture exhibit significant heterogeneity in the amount of

TnaA-GFP and its cellular location during both late stationary phase (t = 0) and lag phase (0- approx. 30 min). Maximum heterogeneity is seen immediately following inoculation. To quantify the variability of tryptophanase content during this early phase of the experiment, analyses were conducted of the distribution of TnaA-GFP among the cells, as well as the presence or absence of foci in individual cells.

Figure 5.11 illustrates the heterogeneity of tryptophanase content among the cells. The y-axis shows the frequency of the fluorescence (a.u.) value specified by the x-axis. The KDE (green line) gives a clear indication of the broad distribution of these data. The data show a peak at approx. 600 a.u. with a tail skewed towards high fluorescence values. There are low-frequency outliers with very high quantities of tryptophanase (3000-4000 a.u.). Further, there is an additional, minor peak in the data at approx. 1400 a.u.

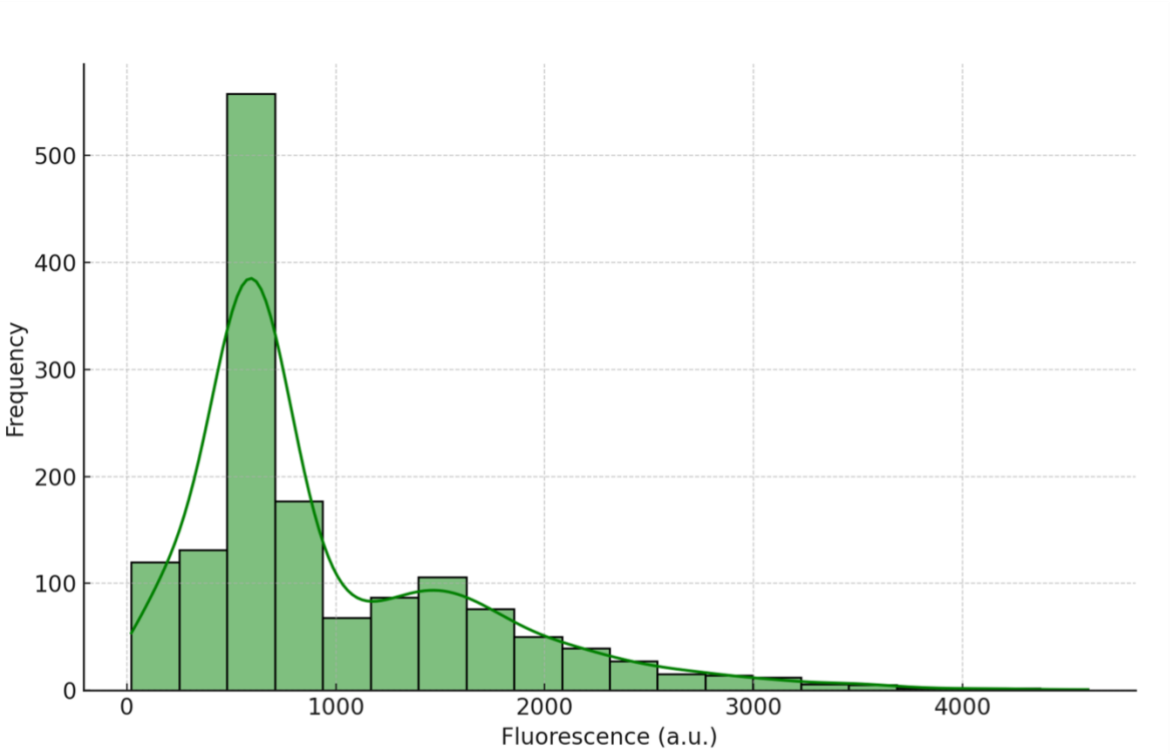


Figure 5.11. The distribution of TnaA-GFP fluorescence at time = 0. *E. coli* MI13 *tnaA-GFP kan*^{R} cells were inoculated at OD_{600} = 0.05 in LB medium with additional 0.5 mM tryptophan and seeded into channels to the Mother Machine device. This graph represents the variability of TnaA-GFP fluorescence among the cells. The smooth curve superimposed on the graph is a Kernel Density Estimation (KDE). Data is from a representative experiment.

Figure 5.12 demonstrates the distribution of foci within the cells after that were seeded into the Mother Machine microfluidic device at t=0. These data illustrate another aspect of heterogeneity within individual cells in late stationary phase and lag phase. The majority of cells (n = 13,777) do not contain a focus, while far fewer have a focus (n = 1,959).

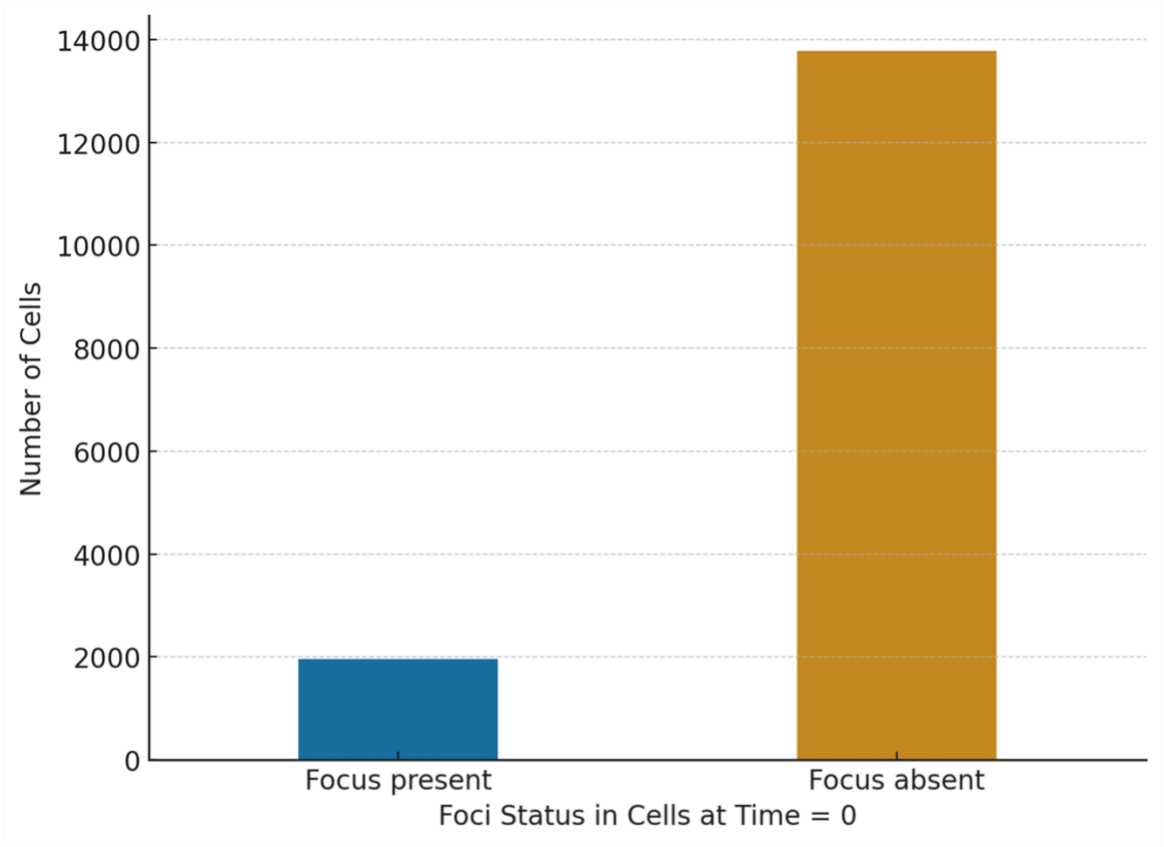


Figure 5.12. The distribution of TnaA-GFP (presence or absence of foci) within 15,736 individual cells in the Mother Machine microfluidic device at time = 0, following the introduction of cells into the microfluidic chip.

Figs. 5.11 and 5.12 do not show how TnaA-GFP content correlates with the presence or absence of foci. This is explored in Figure 5.13 where these two classes are distinguished on the TnaA-GFP content histogram with frequency density represented by the orange and purple lines. There are more cells without a focus (orange) than with one (purple), and most of these cells have fluorescence in the range 250-400 a.u. The cells with a focus (purple) have a similar distribution of TnaA-GFP but show a greater tendency for outliers with a high amount of TnaA-GFP. As the data is not normally distributed, a Mann-Whitney U test was performed for the independent samples in Figure 5.13. The test revealed a significant difference of TnaA-GFP fluorescence in cells with and without a focus, with focus-containing cells having significantly more fluorescence (Mann-Whitney U test, U-statistic = 342919.0, p-value = 5.64e-96).

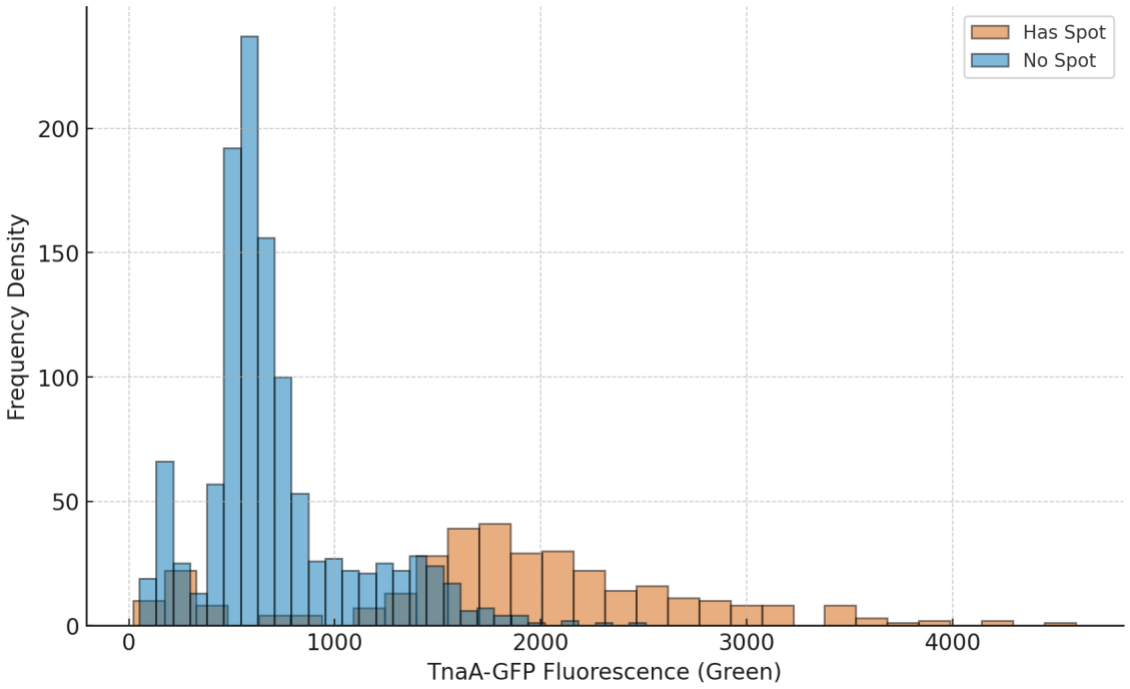


Figure 5.13. Frequency distribution of TnaA-GFP Fluorescence in cells containing a focus (orange) and cells without a focus (blue).

Having investigated the composition of the late-stationary phase culture used to inoculate the mother machine, I next explored the lag phase that immediately followed inoculation. *E. coli* cells sub-cultured into fresh LB growth medium produce indole during lag phase but synthesis ceases when the cells enter exponential phase (180). It has been suggested that producing indole may delay the escape of individual cells from lag phase (unpublished data Summers laboratory). In this thesis it has been argued that tryptophanase foci contain inactive enzyme while dispersed tryptophanase is active, consistent with findings from Li and Young (206). If this hypothesis is correct, cells with dispersed tryptophanase should produce more indole in lag phase and should therefore take longer to escape than their focus-containing siblings. I therefore compared the time to the first cell division (a measure of escape from lag phase) for cells with or without a focus across over 15 000 cells (Figure 5.14). When cells contain a focus, the average time to first division is 54.73 min, in comparison to the cells with dispersed TnaA-GFP, where the average time to first division is 62.07 min. These two division times are significantly different when an independent t-test and Welch's test is performed (t-test, t= -8.07, $p<0.001$). Although only tentative, this result is at least

consistent with the idea that cells without foci have active tryptophanase and hence produce more indole.

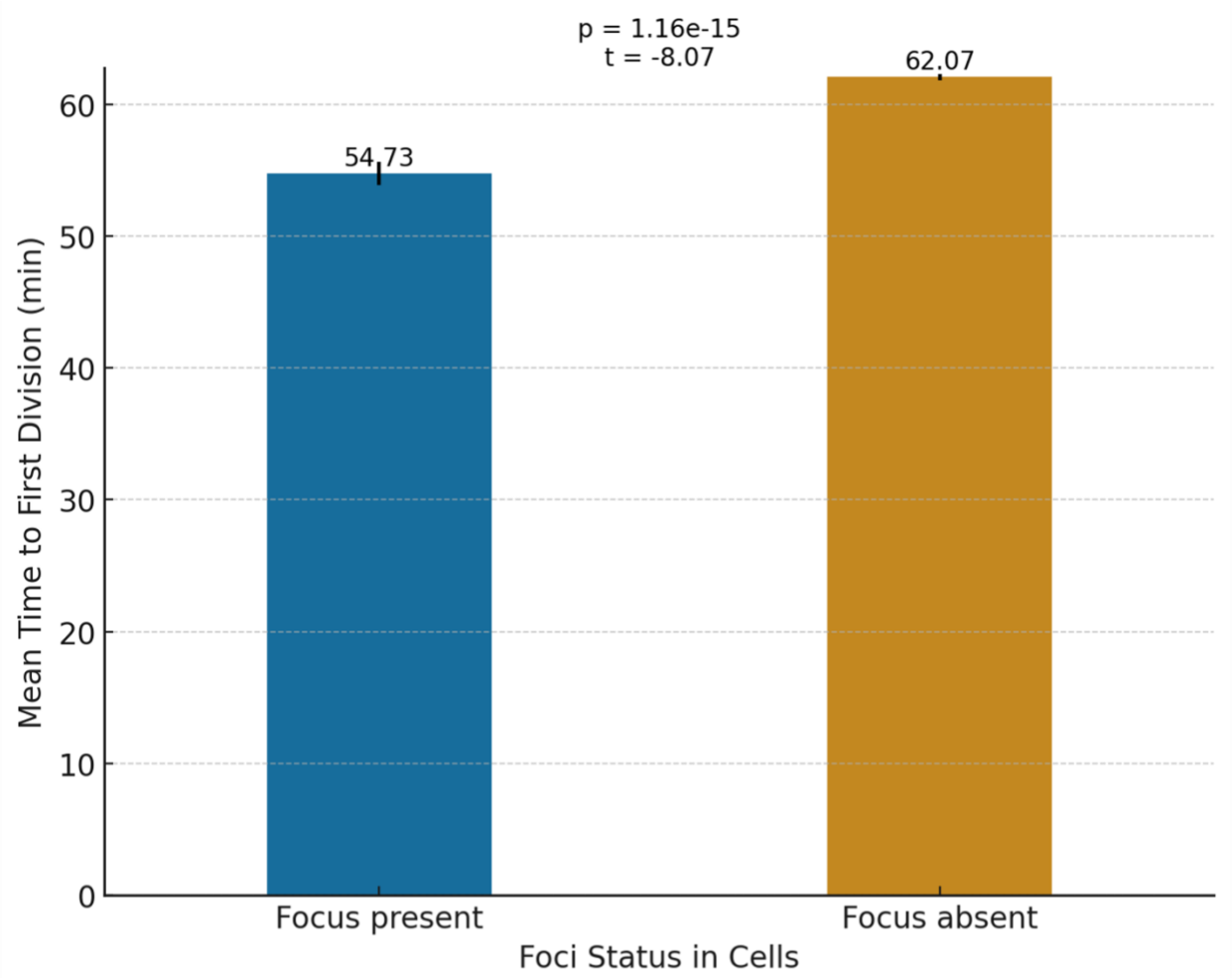


Figure 5.14. The mean time to division (min) for cells that have a focus versus cells that do not. After cells have been loaded into the mother machine chip, the cells are flushed with fresh medium and the time until their first division is recorded. The presence of absence of a focus is determined through learning models and a computer algorithm.

5.6 Discussion

The mother machine has been employed to extend our tryptophanase studies to include single-cell analysis. An important aspect of the approach used here is that it provides a

link between conventional shake-flask experiments and pure single-cell assays (e.g. flow cytometry). Until now, these distinct areas of study have been separated, allowing for only limited comparison. To establish the validity of this approach it was important to compare its data with more conventional studies. This includes work described in earlier chapters of this thesis, published work by Gaimster and colleagues (147, 172), flow cytometry data (158) and tryptophanase distribution/activity data (206).

The mother machine TnaA-GFP data in Figure 5.5 look very similar to shake-flask data reported in Chapter 4. Cells have the most TnaA-GFP in lag phase, followed by a period of dilution as they enter exponential phase. In mid-exponential phase (110 to 160 min) cells produce sufficient TnaA-GFP to balance the dilution effect of growth. This is consistent with a previous study that concluded that TnaA synthesis resumed during mid-exponential phase, as the mean cellular fluorescence remained unchanged despite the increase in culture density (158). Subsequently in the mother machine (200 to 260 min), there is a period of rapid increase which then levels out as the culture enters stationary phase. This pattern of tryptophanase production in the mother machine not only matches that seen in previous chapters, but it is also consistent with published research. Gaimster and colleagues studied tryptophanase expression immediately prior to the indole pulse and saw a rise in TnaA-GFP. This corresponds to the rise from 200 to 260 minutes in the mother machine experiment (172).

Previous researchers looked at tryptophanase production in a culture at different phases of growth (158). Using the single-cell technique of flow cytometry, it was determined that TnaA-GFP content is highly heterogeneous in late stationary phase, and this persists until late exponential phase. This is consistent with data presented in Figures. 5.4 and 5.7-5.9 where the greatest heterogeneity is seen in deep stationary phase (i.e. when cells are first added into the mother machine). Heterogeneity reduces as the culture enters exponential phase, but substantial heterogeneity remains throughout the experiment. The flow cytometry analysis further explored early stationary phase and found distinct subpopulations of cells with high and low TnaA-GFP. This may possibly correlate with cells that contain either foci or dispersed tryptophanase. Figure 5.6 demonstrates that while most cells have a focus during the

transition to stationary phase, there is still heterogeneity in the cells, as approx. 40% of cells have either lost their focus or never formed one.

A key concept that emerges from the mother machine experiment is the relationship between the localization or dispersion of TnaA-GFP foci and the enzymatic activity of tryptophanase. Foci are primarily formed in late exponential phase and disappear on the onset of stationary phase. From Figure 5.6 it is clear that the appearance and subsequent dispersion of foci is tightly linked to the production of indole. The presence of the focus could indicate the inhibition of tryptophanase enzymatic activity, possibly associated with occlusion of the active site, as proposed in previous studies (206). Where this thesis differentiates itself, is the heterogeneity of tryptophanase production and activity throughout growth, that was visualized in the Mother Machine experiments. This heterogeneity has been examined in small snapshots of growth, but never fully visualized at every stage, informing us on the dilution of tryptophanase in exponential phase, the formation of small bright foci during exponential phase, where the proportion of foci compared to diffused is very high. Further, the relationship between diffused foci in lag phase and the production of a mini-indole peak is shown. The cells with large quantities of diffused tryptophanase are delayed to their first cell division, indicating a small lag indole peak may prevent cell division timing. As well, the connection between diffusion of foci and the indole pulse, pulling together indole signalling mechanisms and post-translational regulation is shown in this thesis.

Although there is an overlap between indole production and the presence of foci, all the foci have been dispersed when indole production is at its highest rate. It has previously been asserted that the trigger for the indole pulse is an increase in tryptophanase production in late exponential phase (147, 172)). This hypothesis should be expanded to recognise the likely existence of two triggers: the production of more tryptophanase and the conversion of tryptophanase from an inactive to an active form.

A concept under preliminary investigation in the Summers laboratory is the possibility of a lag phase indole pulse. This mirrors the established indole pulse that occurs during the transition into stationary phase. In both cases the key idea is that, if indole is being

produced more rapidly than it can leave the cell, very high indole concentrations can exist transiently in the producer cell. This may be sufficient to inhibit cell growth and division, even when the external indole concentration is low. While indole is undetectable immediately after subculture, assays have revealed that supernatant indole rises rapidly to 30 ± 10 μM during lag phase (180). Lag phase indole is likely to be produced by the sub-set of inoculating cells that contain dispersed tryptophanase. Consistent with this proposal is the observation (Figure 5.14) that cells with dispersed tryptophanase are slower to exit the lag phase. Given the substantial variation in tryptophanase content within this sub-population (Figure 5.9), individual cells are likely to produce significantly different amounts of indole and be subject to different degrees of growth inhibition. It is even possible that the small number of cells with the highest amount of dispersed tryptophanase may fail to exit lag phase and become dormant persister cells (231).

This chapter concludes with a summary of observations and hypotheses. A study of tryptophanase distribution in lag phase has assisted investigation into a potential lag phase indole pulse. During lag phase, the culture supernatant accumulates only 0.05 mM indole but, importantly, this is being produced by a relatively small number of cells (OD_{600} = 0.05). It is proposed that the sub-set of cells with dispersed tryptophanase actively convert tryptophan into indole, resulting in a mini-indole pulse that causes these cells to divide later than those with inactive tryptophanase concentrated in a focus. While there is substantial tryptophanase in the cells during lag phase (inherited from the overnight culture), this is diluted by cell division as the culture enters exponential phase. During mid-exponential phase, cells start to produce tryptophanase at a sufficient rate to compensate for dilution (Figure 5.5). In the exponential to stationary phase transition, there is a sharp increase in the production of TnaA-GFP, as well as a rapid appearance and subsequent dispersion of foci. Tryptophanase production is up-regulated earlier than hypothesized by previous studies (147, 172). Furthermore, study of the transition to stationary phase has hence identified a new trigger for the indole pulse; the activation of tryptophanase through the dispersion of foci. While a previous study theorized that diffused TnaA was more active that foci, it did not identify the

dispersion of foci as the cause of pulse timing (206). During Mother Machine experiments, the formation of foci in the exponential phase, following dilution of the TnaA from the previous night, was particularly noticeable and represented the most visually homogenous distribution of TnaA during the growth phase. This observation aligns with previous flow cytometry experiments showing homogeneity before the transition to stationary phase (158).

It has been suggested that the localization of foci to the end of cells corresponds to the MinCDE system. In experiments where this system was deleted, the foci did not localize to one end of the cell. The *ΔMinCDE* cells did not divide correctly at the correct mid-cell position and these cells did also not form discrete singular foci per cell. This observation, along with the timing of diffused TnaA in our experiments, suggests that the sequestration of TnaA in localized foci is involved in the regulation of these foci. The disruption of the foci likely triggers TnaA activation.

The homogeneity of foci seen in mid-exponential phase disappears after the transition to stationary phase, correlating with the timing of the pulse. Additionally, the production of indole by diffused TnaA during lag phase was observed to delay cell growth, showing that a smaller indole pulse may affect a limited number of cells in lag phase. This represents a novel hypothesis that requires further validation. Moreover, the discovery that cells vary in their indole production, potentially not all pulsing simultaneously, may be a significant new finding that has not been previously identified.

.

Chapter 6: Investigation of *tnaCAB* presence and evolution in *E. coli* and the microbiome

6.0 Context

During the COVID19 pandemic, laboratories were shut across the University of Cambridge. At this time, it was not possible to conduct wet lab work. As a result, I resolved to continue my research in the dry lab by exploring tryptophanase and indole signalling through bioinformatics. This work, initiated under lockdown, continued after the pandemic with two undergraduate students who were supervised by Dr. Marta Matuszewska and myself (see Table 6.1).

Table 6.1 Workflow of bioinformatics analysis and the lead researcher for each task.

Work completed	Researcher
E. coli downloaded and database created	EK and HL
Pathogen Watch information obtained; ST type MLST tree	MM
Blast of *tnaA*, *tnaB,* and *tnaCAB*	EK, HL
Neighbour joining tree for *tnaA*	EK
MLST tree for *E. coli*	HL, MM
SNP analysis	HL
Compilation of TnaA query sequences	EK, SH
Collection of database and analysis of composition	EK
BLASTp	EK, SW
TnaA vs TPL analysis	EK
RaxML analysis, phylogenetic tree generated	SW, MM

6.1 Introduction

The majority of research into tryptophanase and indole signalling in *E. coli* has relied on conventional microbiological methods, including bulk-culture assays utilizing monocultures in nutrient-rich media. While this provides valuable insights into signalling behaviour in the laboratory, there is limited research into the presence of the tryptophanase operon in other *E. coli* strains or the subgroup, *Shigella.* Information is also limited on how widespread is the tryptophanase operon beyond *E. coli* and in what other species it might be found in an environment like the human gut microbiome. The presence of the tryptophanase operon could give insight into aspects of microbial physiology, host-microbe interactions and even potential applications in medicine and biotechnology (see section 1.1.3.2).

Crosstalk between the host and the gut microbiota is an important interaction that has been widely studied for decades. There is an increasing interest in research focussed on the role of TnaA and indole in modulating and maintaining the human gut microbiome (122, 232-235). Eukaryotes cannot encode TnaA, so indole and its derivatives in the human gut are produced by gut microbials (108). A common assumption in the literature is that *E. coli* is the main producer of indole within the gut (122, 236). However, which species within the microbiota are capable of producing indole *via* TnaA remains to be fully elucidated. While the majority of the microorganisms within the microbiome cannot be cultured, recent advances in Whole Genome Sequencing (WGS) allow researchers to gain insight into the metabolites produced by different phyla and species. Despite the increasing availability of sequence data and WGS, the existing body of research on *tnaA* and *tnaCAB* from the perspective of bioinformatic analysis is relatively small. Learning about the phylogenetic and evolutionary history of the tryptophanase operon within the gut microbiome promises to enhance our comprehension of tryptophanase's role and significance.

6.2 *E. coli* genomes database

A comprehensive collection of 1974 complete *E. coli* genomes were obtained and downloaded from NCBI Assembly, as well as the corresponding metadata (full details

provided in 2.10, 5 February, 2021). Any strain duplications and plasmids were removed and a final BLAST database was created containing 1251 *E. coli* genomes with corresponding metadata that was used for further analysis. Strains were isolated over an extended period (1920-2020) with most samples collected since 2010 (Figure 6.1A). They were collected from 58 countries around the world (Fig 6.1B) of which the most sampled were the USA and China (42%). As seen in Figure 6.1c, the majority of the *E. coli* genomes were sampled from humans (34%), followed by pigs (5%), cattle (4%) and chicken (2%).

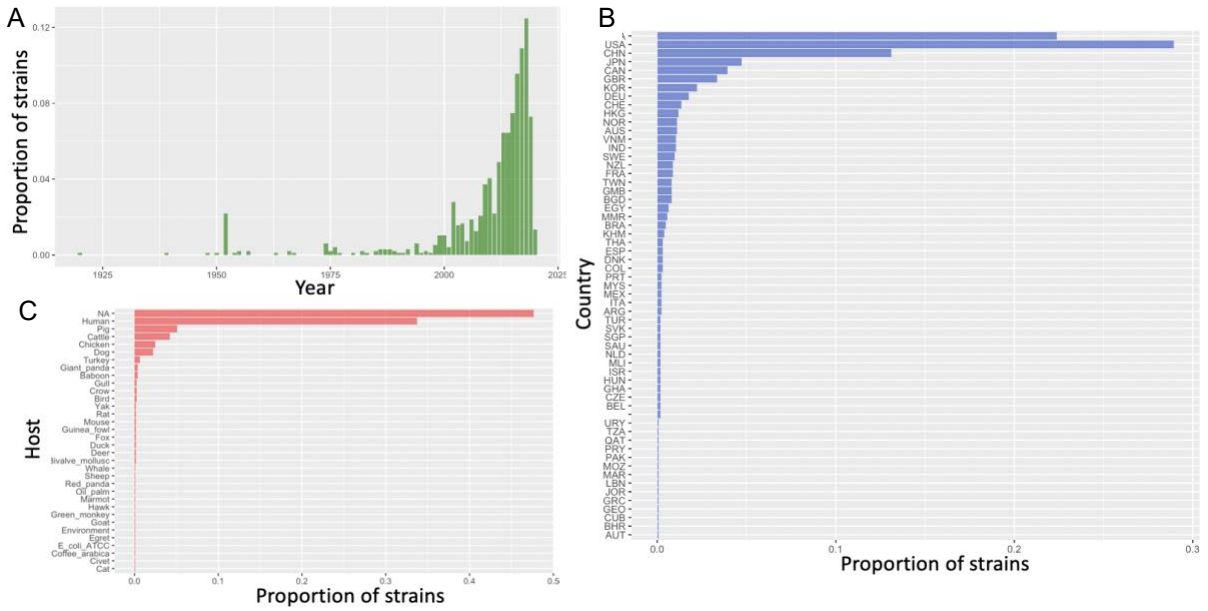


Figure 6.1. Metadata represented in bar plots obtained from NCBI BioSample from the 1251 *E. coli* genomes in the dataset. A. The proportion of samples that were collected in each year. B. The proportion of samples that were retrieved from each country. C. The proportion of strains that were collected from different hosts, there were 33 different sources in total. This figure was generated by Hannah Lin under the guidance if EK and MM.

The species information from each of the sequences, which was determined by the program PathogenWatch, is shown in Table 6.2. Most of the sequences in the BLAST database were identified as *E. coli* (86.8%), with a small portion of the strains coming

from a variety of *Shigella* species (5.2%). The *Shigella* genomes were retained in the database for further analysis as the two species show genetic similarity and are phylogenetically intertwined, reflecting a shared evolutionary history. *Escherichia marmotae* was identified but was removed from further analysis as it was poorly annotated at the time of study (237).

Table 6.2. Species information from the 1251 genomes in the BLAST database. The table shows the number and percentage of the genomes that was represented by each species which was determined by Pathogen Watch.

Species	Number of Strains	Percentage of total (%)
Escherichia coli	1086	86.8
Shigella sonnei	33	2.6
Shigella sp. PAMC 28760	22	1.8
Shigella dysenteriae	7	0.6
Shigella boydi	2	0.2
Escherichia marmotae	1	0.1
Unidentified	1	0.1
No species Information	98	7.8
Total	**1251**	**100**

The sequence type (ST) type is a unique numerical identifier assigned to bacterial strains based on the sequence of specific gene markers or housekeeping genes. The ST information was unavailable for the *Shigella* genomes and there were 99 strains where no information was found. As a result, 1086 *E. coli* strains were used to create an MLST tree in ape (v5.4-1) and annotated with FigTree (v1.4.4) and iTol (v6.0) (Figure 6.2).

From the *E. coli* genomes, there were 326 different STs identified. The most common STs in the created database were ST10 (13%), ST11 (10%), and ST131 (5%). There were 33 sequences that were identified as having novel STs. Figure 6.2. shows the *E. coli* genomes as represented by STs with at least 10 sequences in a MLST tree. The pale-yellow colour in Figure 6.2 represents the novel STs. The tree shows clustering based on STs, which is expected as the tree was constructed based on MLST gene similarities.

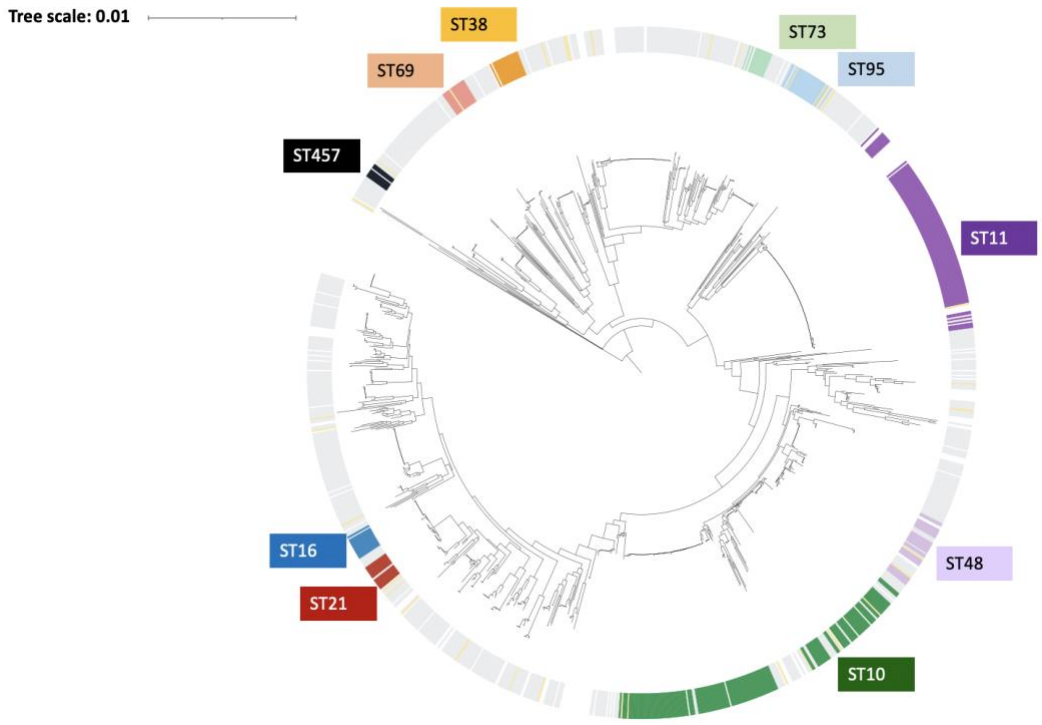


Figure 6.2. A phylogenetic tree of the *E. coli* strains showing genetic relatedness (inner ring) and sequence types (STs) annotated on the outer ring. The center ring shows a MLST tree that represents the *E. coli* population structure. STs are shown in colour on the outer ring and represent STs with 10 or more sequences in the dataset. There are 33 novel STs shown in pale yellow throughout the outer ring. This tree was plotted in iTol (219). Graph generated by MM.

6.3 *E. coli* genome database analysis

To establish the presence of individual genes *tnaA, tnaB, tnaC* and the whole tryptophanase operon (*tnaCAB*) in the *E. coli* and *Shigella* genomes, a BLASTn was run against the established database using the operon and its genes from *E. coli BW25113* as a reference genome. The hits were filtered for alignment length (at least 80% for genes and 75% for operon) and at least 90% identity match. Upon filtering out non-specific matches using the alignment threshold for length (>80%), *tnaA* was found to be present in 99.4% of the genomes (1244/1251). Figure 6.3a-c demonstrates the hits before the alignment thresholds for length, percent identity and e-value are applied for the 1244 *tnaA* matches. This figure reveals that all the matches are over 90% identical to the *tnaA* reference sequence from *E coli BW25113* and there were only a few hits where the

alignment length did not reach 80%. In conclusion, *tnaA* exhibits a high degree of conservation within *E. coli* and *Shigella*.

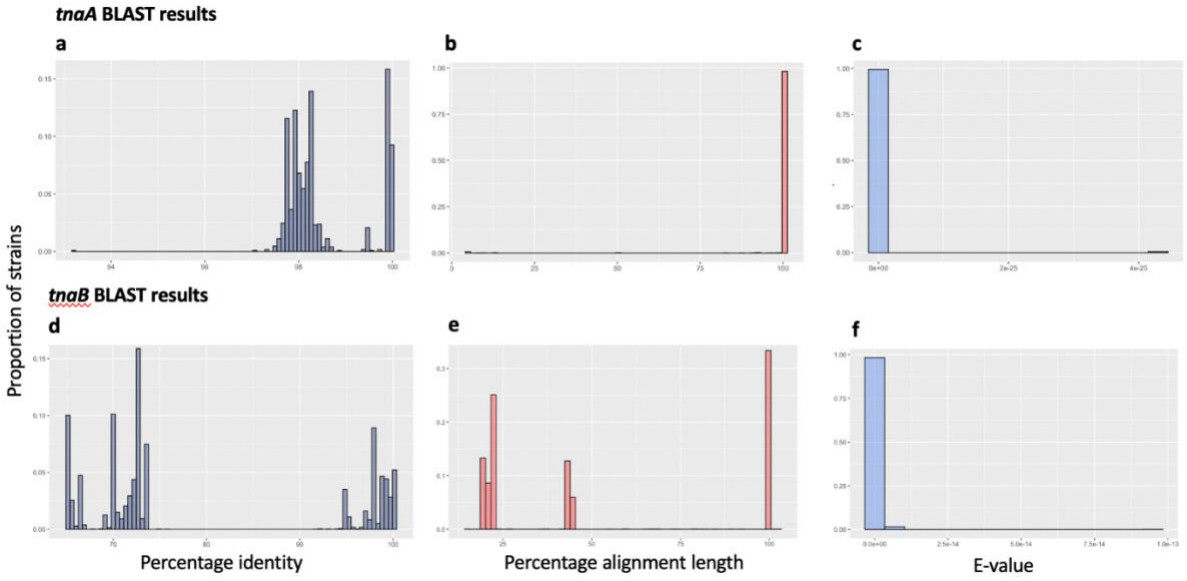


Figure 6.3. Histograms that represent the BLASTn matches from *tnaA* and *tnaB* and their distributions based on the following parameters: percentage identity, alignment length and E-values before alignment thresholds were applied. Graph generated by Hannah Lin under the supervision of EK and MM.

To determine the conservation of *tnaB* and *tnaC*, a similar approach was adopted. It was ascertained that *tnaB* was present in 1232/1251 (98.5%) genomes and *tnaC* was present in 1242/1251 (99.3%) genomes in the database after alignment thresholds were applied (>80% alignment length, >90% percent identity). Figure 6.3d-f shows the distribution of *tnaB* matches before threshold alignments were applied. There are two groups of data in the percentage identity match of *tnaB*, with one cluster matching with about 65-75% similarity and the other cluster representing a >95% identity match. The cluster with 65-75% similarity matched the alignment length of 20-50% of *tnaB*, while the cluster that had >95% similarity also had the same length as the *tnaB* gene. In the BLAST hits for *tnaB*, it was identified that 94.9% (1187/1251) of the sequences showed three BLAST hits for the gene. Matches that showed <40% alignment length of *tnaB* were removed. The sequences that displayed fewer specific matches (approx. 70%) that had an alignment

length approx. 45% of the *tnaB* gene were identified as the gene *mtr.* The gene, *mtr*, is an alternative tryptophan-specific permease found in *E. coli* (230, 238).

To determine whether the tryptophanase gene, *tnaA*, always resides within the tryptophanase operon, a BLASTn search was performed for *tnaCAB*, which included the intergenic and promoter regions. Once the alignment length threshold of 75% was applied (more diversity over a longer alignment length), it was determined that 1232/1251 (98.5%) sequences contained the entire operon. When comparing the start and end positions of the alignments, *tnaA* was always located within the *tnaCAB* operon when both were present. Further, in strains where *tnaA* was duplicated, all the *tnaA* genes still resided in the operons.

6.4 Comparing sequence diversity within *tnaA* and *tnaCAB* to diversity across the *E coli* genome.

This section outlines an investigation into the diversity within *tnaA* and *tnaCAB*. A neighbour-joining tree was created from the *tnaA* alignments for the 1244 *E. coli* genomes where the gene was present (using RAxML; default settings and annotated with FigTree (v1.4.4) and iTol (v6.0). Each branch represents a sequence from a single strain (Figure 6.4a). The constructed tree was used to visualize branches and assign strains to one of five distinct clades (A-E) based on the similarity in the *tnaA* gene.

The MLST tree was annotated with the *tnaA* clades to compare diversity to the rest of the *E. coli* genomes (Figure 6.4b). It appears that the clustering of the *tnaA* genes largely matches the clustering of the seven housekeeping genes used in the MLST tree (239). In areas where they do not match, it is possible that there is recombination or functional specialization. Recombination events can occur when segments of DNA are exchanged between different bacterial strains, leading to genetic variation (240). Functional specialization refers to the adaptation of genes for specific role sin the environments (241). Both of these processes can alter the genetic makeup of bacteria and create clustering patterns, indicating *tnaA* genes may have undergone different evolutionary pressures or genetic exchanging compared to the house keeping genetics in these identified clusters.

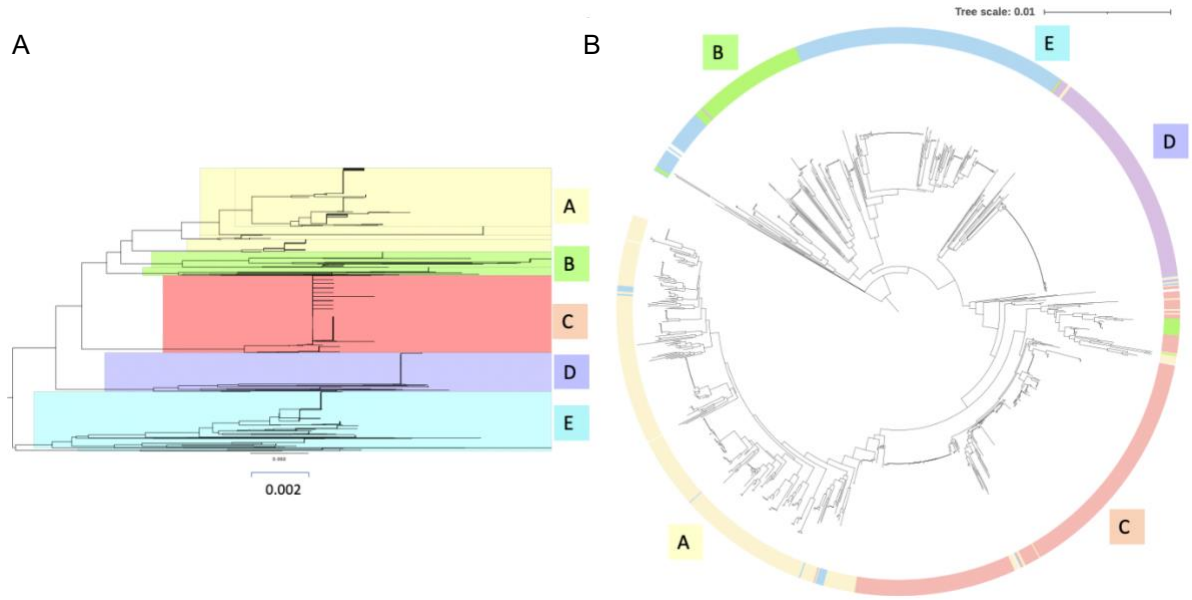


Figure 6.4. Phylogenetic trees depicting *tnaA* diversity in *E. coli*. A. Neighbour-joining tree of *tnaA* alignments based on the unique *E. coli* database. Clusters were visually assigned and divided into 5 separate clusters, A-E. The tree is midpoint-rooted and rendered through FigTree (v1.4.4). The scale bar provides a measurement of distance in SNP/site. B. MLST-based phylogenetic depicting the population structure of *E. coli*, with *tnaA* clades displayed in the outer ring. Clustering according to *tnaA* shows conservation in the overall *E. coli* population. This tree was illustrated in iTol (Letunic and Bork, 2019). (Panel A constructed by author, Panel B constructed by HL and MM).

To measure the evolutionary distance between *tnaA* sequences, *tnaCAB* sequences and the seven housekeeping genes used for the MLST analysis, pairwise distance matrices were created showing the number of SNPs between sequences in the dataset (Table 6.3). The maximum and average number of SNPs per site (number of SNPs present in the given sequence divided by the number total number of sites analyzed) was calculated to measure nucleotide diversity and genetic variation. The minimum number of SNPs per site if the sequences were identical was 0.

Table 6.3. A measure of nucleotide diversity determined by the maximum and mean number of single nucleotide polymorphisms (SNPs) per site among *tnaA, tnaCAB* and MLST genes (HL).

	tnaA	***tnaCAB* operon**	**MLST genes**
Maximum	0.035	0.038	0035
Mean	0.016	0.017	0.012

For the *tnaA* gene, there was a maximum of 0.035 SNPs per site and a mean of 0.016 SNPs per site. This indicates there is a small amount of genetic variation between *tnaA* genes in *E. coli*. When this number is compared to the SNPs per site in *tnaCAB* and the seven MLST genes, it is seen that diversity among the *tnaA* gene, operon and MLST genes is very similar. This substantiates the results that show *tnaA* diversity follows similar diversity in the rest of the *E. coli* genome which is likely to mean that they are under similar selective pressures.

6.5 Diversity of *tnaA* in the human gut microbiome

6.5.1 Acquiring protein sequences from the gut

To evaluate the presence of TnaA in the gut microbiome, the initial analysis was conducted by the author of this thesis, and then continued and elaborated by a student under the joint supervision of this author and Dr Matuszewska of Cambridge University Veterinary School.

A compilation of 48 TnaA query sequences from 18 different bacterial phyla from the NCBI Protein Database (https://www.ncbi.nlm.nih.gov/protein/) was assembled to probe the distribution and diversity of tryptophanase in the human gut microbiome. These protein query sequences were chosen from confirmed TnaA sequences provided in the protein database and aim to capture the diversity of TnaA across different phyla. Protein sequences were used instead of nucleotide sequences, as the diversity among species was sufficient to mask possible hits if the reference genome came from a different phylum.

The Unified Human Gastrointestinal Protein (UHGP) was chosen as the source of gut microbiome, as it is a recent, diverse microbiome database containing 4,616 bacterial

species, spanning 22 distinct phyla, with a total of 285,835 genomes (4) (Figure 6.5). The sequences, functional annotations and metadata were retrieved from MGnify (https://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/).

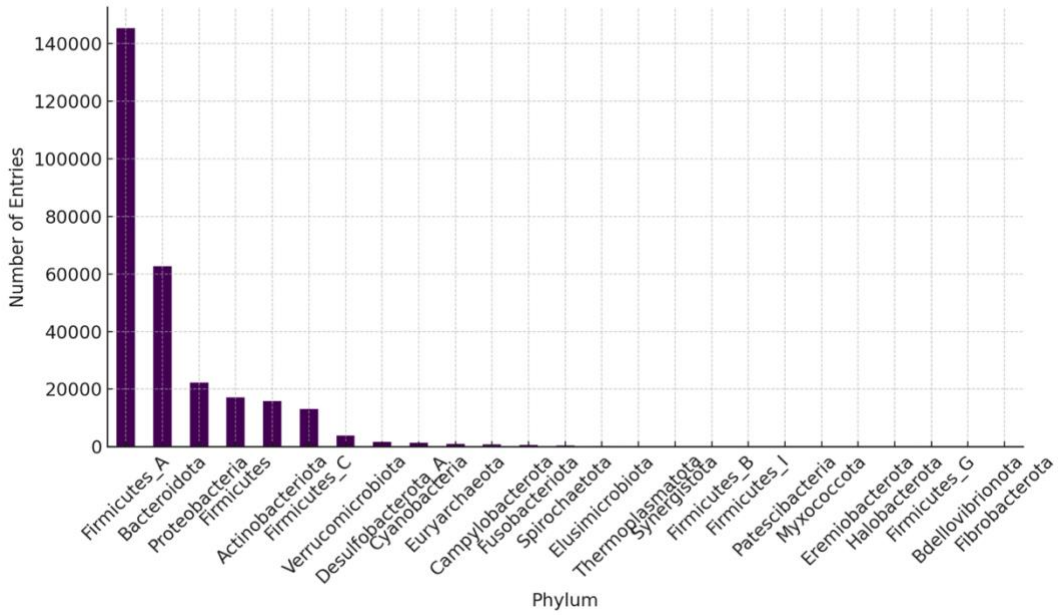


Figure 6.5. The Unified Human Gastrointestinal Protein (UHGP) database represents over 4,616 bacterial species from across the different phyla that were detected from the human gut. Most of the genomes were extracted from the phyla Firmicutes A, Bacteriodiota and Proteobacteria.

6.5.2 BLASTp and Phylogenetic Analysis

The protein sequences were concatenated and a BLASTp database (v 2.12.0+) was created with the following parameters: query start, query end, subject start, subject end, alignment length, percent identity, number of mismatches, e value, sequence ID, aligned query sequence, and aligned subject sequence. The resulting hits from all 48 TnaA alignment searches were filtered for the following alignment thresholds: >50% identity, >80% coverage, E-value <1e-5). The protein matches were aligned and, if sequences contained any significant gaps (25%), these were excluded from the analysis to focus on regions of interest and provide more reliable results. Further, any alignment sites that

were highly conserved (>90% of sites in this analysis) were removed to optimize the presence of phylogenetically informative regions.

6.6 Tryptophanase or tyrosine phenol lyase

As the human gut microbiome was searched for protein sequences with homology to TnaA, there emerged a protein of interest that is genetically similar to TnaA on the amino acid sequence level. Tyrosine phenol lyase (TPL) is an enzyme that converts L-tyrosine to phenol, pyruvate and ammonia employing the cofactor, pyridoxal phosphate. TPL and TnaA follow similar catalytic mechanisms and appear to have similar crystallographic structures (Figure 6.6) (2). During alignment using the BLASTp program, TPL appeared in multiple phyla (Firmicutes, Bacteriodiota, Spirochaetota) and shared a percentage identity with the TnaA query sequences of >25% and in some cases over 90%. TPL also appeared to be a similar length to TnaA (approx. 458 aa) in many of the BLASTp searches. These two enzymes appear to be genetically very similar, but previously published research indicates that they are extremely specific for their respective substrates: tyrosine for tyrosine phenol lyase and tryptophan for tryptophanase. It is important to consider that even genes that appear nominally similar in different species may have distinct functions. Therefore, exercising caution is imperative when conducting searches for TnaA-like sequences.

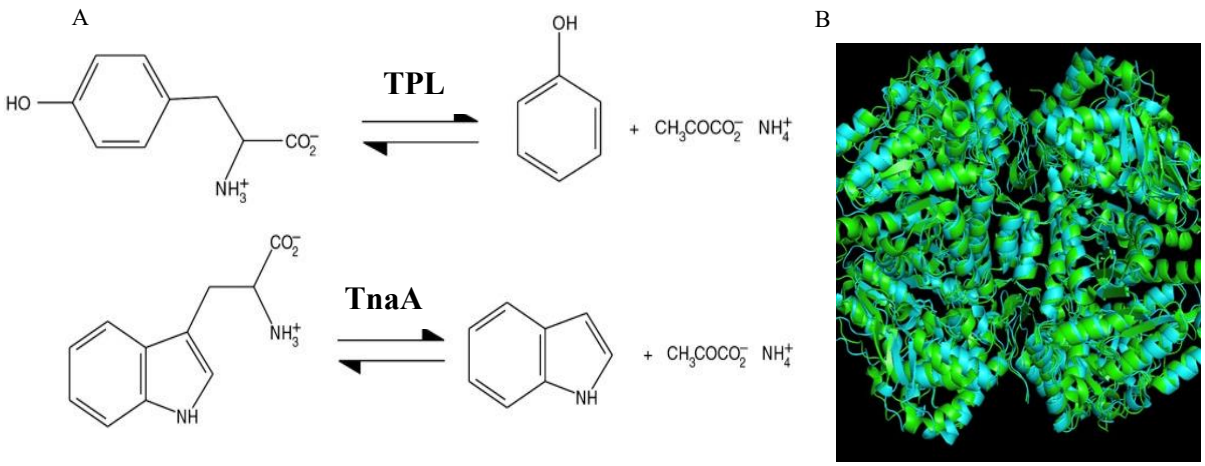


Figure 6.6 Comparison of the enzymatic reactions and crystal structure between the tetrameric enzymes tryptophanase (TnaA) and tyrosine phenol lyase (TPL). A. Enzymatic reaction of TPL and TnaA. Both enzymatic reactions require the cofactor, pyrixodal phosphate and result in the production of ammonia, pyruvate and one other product (1, 2). B. The crystal structures of both enzymes superimposed on each other in PyMOL shows their similar structures.

6.6 The distribution of TnaA in the human gut microbiome

There were 478 TnaA sequences identified in the UGHP when the alignment thresholds were applied. Three phyla accounted for over 90% of the TnaA proteins identified: Firmicutes, Bacteroidota, and Proteobacteria (Figure 6.7a). More than 36% of Bacteroidota in the UHGP catalogue, which is the second most abundant phylum in the gut microbiome, encoded a TnaA protein (Figure 6.7b). The most abundant phylum, Firmicutes, represents 20.7% of the TnaA hits, but only 6.34% of the overall percentage of species in the UHGP. Of the total species represented in the UHGP, approx. 10% of the species contain TnaA.

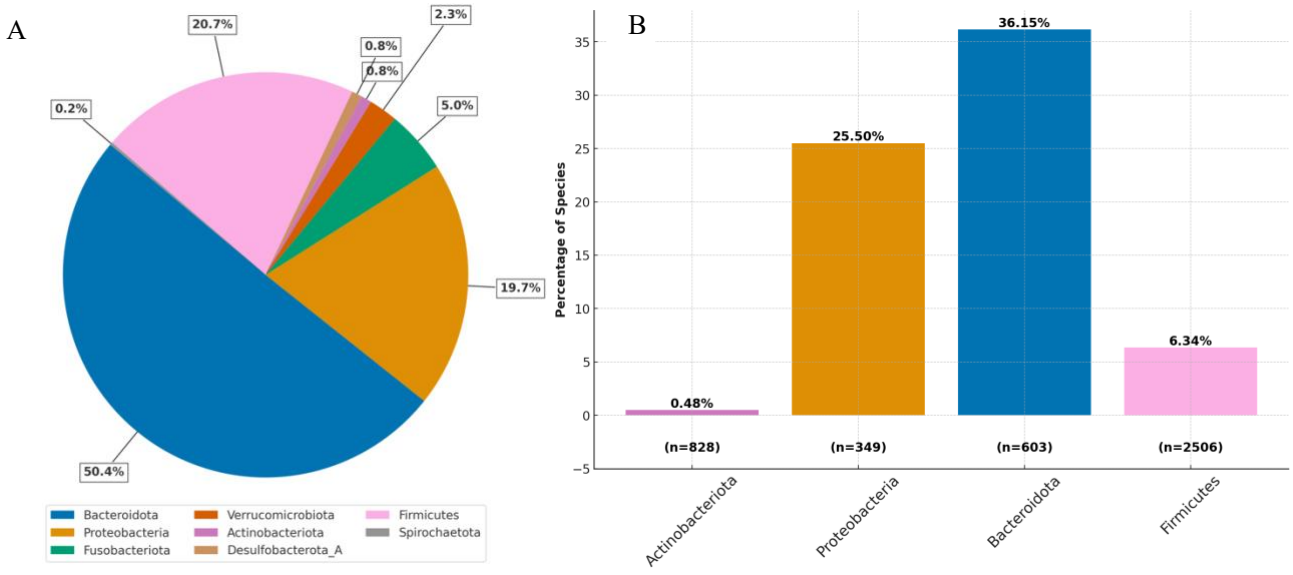


Figure 6.7. Presence and frequency of TnaA in different phyla of the gut microbiome. A. Pie chart showing the percentages from each of the phyla that contains TnaA (n = 478). B. Percentage of the species that contains TnaA that was identified in this analysis. n = total number of species in the phylum. Any phylum that contains less than 100 species in the UHGP was not included in this plot.

To further elucidate TnaA phylogenetic evolution in the human gut microbiome, a phylogenetic tree was constructed. Using the program, RAxML (v 8.2.12), a tree was generated and unsupported branches were removed (<70 bootstraps) (Figure 6.8). Two programs were used for annotating the phylogenetic tree: iTOL (v 6.3) and GrapeTree (v 3). In Figure 6.8A, the inner ring represents the phylum of each of the species within the

UHGP, this was adapted from work by Almeida and colleagues (4). The outer notations on the graph represents where species contained an identified TnaA sequence.

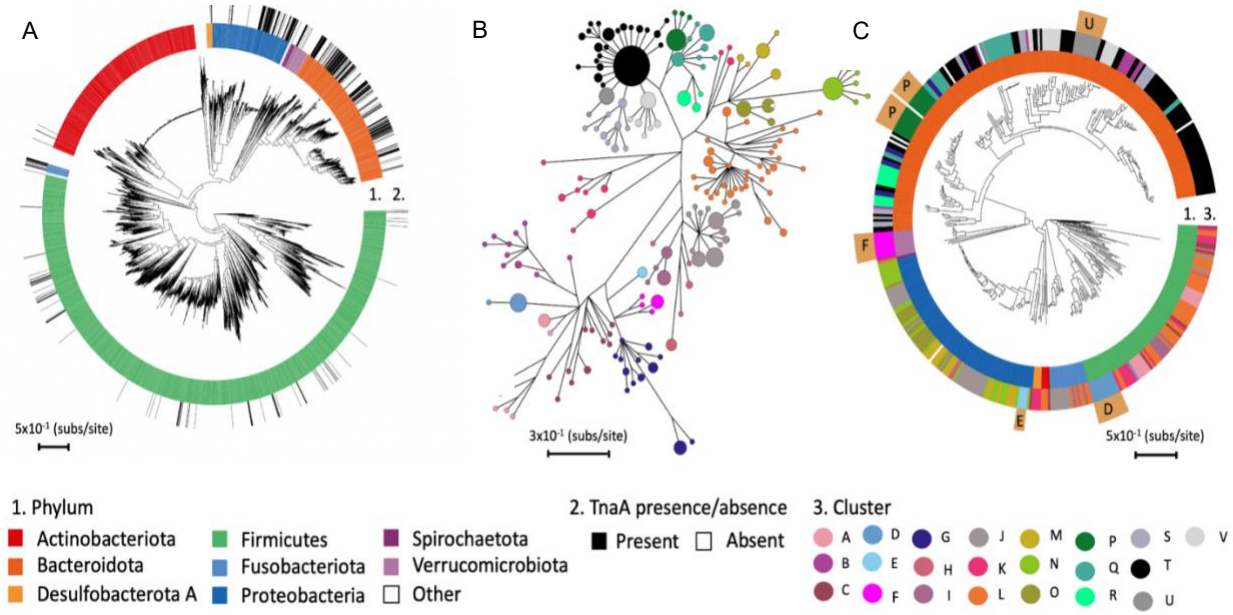


Figure 6.8. Evolutionary history of TnaA in the human gut microbiome. A. 1. Phylum annotation. 2. Species within the UHGP that were identified as containing TnaA. 3. Visually annotated clusters. Other includes phyla that do not have a TnaA (4). B. Minimum-spanning tree derived from a maximum-likelihood tree of TnaA proteins. Branches <0.0494 were collapsed for clarity. C. Same as (A) but this only contains species with a TnaA protein identified in this work. Outer light brown boxes denote clusters with co-phylogeny. Cluster D – Oscillospiraceae, Cluster E – Genus *Oxalobacter*, Cluster F – Class Verrucomicrobiae, Cluster P – Genus *Alistipes,* Cluster U 0 Genus *CAG-485.* Graph generated by Sarah Willich under the supervision of MM and EK.

Clusters were classified visually based on the maximum likelihood tree of the TnaA sequence that was created in the program RaxML (Figure 6.7B). The clusters were annotated onto the phylum-level phylogenetic tree adapted from the UHGP (4).

There was co-phylogeny represented by aggregation of five clusters (D-F, U and P) (6.8B-C). Co-phylogeny refers to the intertwined relationship of microorganisms and their evolutionary history with either their hosts or other organisms that they are closely associated with. This suggests that microbial species or taxa within the gut microbiome

may have co-evolved. The clusters P-V are mostly represented by the phylum Bacteroidota. This provides evidence that TnaA has been maintained through evolutionary history in these clusters. Given the dynamic processes of this enzyme's evolutionary history it is possible that there is the potential for horizontal gene transfer of TnaA in the human gut microbiome. However, it needs to be further investigated on a phylum specific level.

6.7 Conclusions from the bioinformatic analysis of the Tna gene cluster/operon

The work described in this chapter has generated a substantial volume of information about the evolutionary history of *tnaCAB* and *tnaA* within *E. coli*. The enzyme appears to be a part of the core genome and has been maintained in *E. coli,* indicating its significance and the probability that it provides a selective advantage.

Beyond *E. coli,* TnaA is found in many phyla of the gut microbiome including Proteobacteria, Firmicutes, and Bacteroidota. Bacteriodota and Firmicutes are two of the dominant phyla in the gut microbiome. They both have TnaA and are therefore likely to be substantial contributors of indole within the gut. The maintenance of the TnaA in clusters seen in the phylogenetic trees suggests that TnaA plays an important role in many different species within the larger gut environment.

6.8 Discussion

The primary goal of this chapter was to observe the presence or absence of *tnaA* and *tnaCAB* in *E. coli.* In accordance with the data, *tnaA* and the tryptophanase operon are predominantly present in *E. coli* and its related subgroup, *Shigella*. This implies the importance of the proper regulation and function of the gene and enzyme.

The phylogenetic analysis of *tnaA* reveals co-evolution of the enzyme alongside the rest of the core genome. This confirms previous assertions that *tnaA* is a part of the core genome and/or is critical for adaptation to certain environmental conditions. As mutant species lacking the enzyme and operon are capable of growing in a laboratory setting, the conservation of the gene must provide an important role for species surviving in the natural environment.

Further analysis of *tnaA* and *tnaCAB* is needed to substantiate these claims and make further conclusions from the phylogenetic analysis of the enzyme and operon. A whole genome tree for *E. coli* would be critical for comparing the evolution of TnaA with the evolution of the core genome, rather than just the seven MLST genes. More *in vitro* testing of a wide range of *E coli* and *Shigella* species for indole production is needed in order to identify whether the genotypic presence of *tnaA* leads to the phenotypic production of enzymatically active enzyme.

The similarity of TPL and TnaA, both genetically and structurally, suggests that the enzymes may share an evolutionary history. It is also possible that the software used by the UHGP to predict functionality of the protein based on the gene name/identifier/sequence labelled a number of the alignment matches for TPL or TnaA incorrectly. TPL appears to be strain-dependent for *E. coli*, where our analysis suggests *tnaA* is highly conserved in *E. coli*. It is difficult to separate the two genes using genetic techniques alone such as comparison of the alignment scores for nucleotide or amino acid sequences, as the annotations represented within the gut microbiome database may be incorrect. Identifying the areas of each gene that are necessary for binding their highly specific substrates could be a valuable future direction of work.

The results from the microbiome analysis show that approx. 10% of the species that were sequenced within this gut microbiome database appear to contain the *tnaA* gene, as determined by the applied thresholds and protein alignment techniques. However, it is not known whether these enzymes are capable of converting tryptophan into indole and further research would be necessary for determining this. An unexpected result of the microbiome search was the prevalence of TnaA within the phylum Bacteroidota (approx. 36% of species were identified as encoding the enzyme). Since Bacteroidota is one of the dominant phyla in the gut (4) indole is likely to be found in the gut environment of the majority of the world's population. Of the 4,616 representative bacterial species that were sequenced in this dataset, it is noteworthy that the majority of the human gut microbiome is unculturable and so it would be difficult to test these species for *in vitro* production of indole (4). This analysis addresses the prevalence of TnaA within these sequences, however, it remains uncertain whether these genes have the capability to produce the

corresponding protein. This investigation revealed 10% of the individual species of the microbiome possess TnaA, but the precise abundance or distribution of each species within the gut ecosystem is undetermined. The phylogenetic relationship of *tnaA/tnaCAB* and TnaA should be further studied by expanding the query sequences to cover more strains and species. Future studies could focus on qPCR of TnaA in the gut microbiota.

The study of metabolites produced by the gut microbiome is crucial to our comprehensive understanding of human health and disease. The trillions of microorganisms in the GI tract play a significant role digestion, immune function, and mental health. Enhanced knowledge of which bacteria produce tryptophanase brings medicine closer to potential development of targeted therapeutics and interventions.

Chapter 7: Discussion

7.1 Technical challenges inherent in this investigation.

An initial aim of this thesis was to investigate the possibility of experimentally controlling the indole pulse in conventional broth culture using a range of approaches, including genetic modifications to the operon and alterations to the growth medium. However, this work revealed that the control of the indole pulse is more difficult to achieve than previously assumed. It became clear that the tryptophanase operon, and the tryptophanase enzyme itself, are subject to additional regulatory mechanisms beyond those initially recognised.

Subsequently, in order better to understand the cellular heterogeneity inherent in broth culture, the study was extended to focus on single cells using the ‘mother machine’. This provided extensive information about the distribution of tryptophanase, but a significant limitation was that, at present, detecting indole production at the single-cell level is unattainable. Previous investigations in the Summers laboratory experimented with intracellular synthesis of indigo as a surrogate for indole production, but the outcomes proved challenging to reproduce.

In the single-cell studies described in this thesis, indole production in the bulk cell reservoir (the culture flask) was used as a proxy for indole production by the cells under observation in the microfluidic device. However, there are potential difficulties in comparing the environment of a bulk culture assay and a mother machine microfluidic set-up. Despite efforts to standardize conditions such as shaking and ambient temperature, it is important to acknowledge that cells under investigation were subjected to different physical environments. Specifically, the cells confined to channels in the mother machine may experience stress-induced signalling or mobility constraints, which in turn could impact the expression of tryptophanase. However, these limitations, while important to acknowledge, are unavoidable in single-cell studies. The findings emerging from this work are of considerable interest, and future investigations should focus on using a variety of methodologies.

7.1.2 Heterogeneity in bacterial populations.

A prominent theme that emerged from this research was the heterogeneity of tryptophanase content and indole signalling. From the bulk culture assays in Chapter 3 and 4 to the single cell assays in Chapter 5, heterogeneity is seen in all aspects of the experiments. It becomes evident that cells exhibit heterogeneity in the quantity of tryptophanase the cells contain and in the exact timing of its expression and activation. This leads to variation in the timing and kinetics of indole production. In Chapter 3, the cell-associated indole levels peaked anywhere from 40-80 mM depending on the tryptophan content in the growth medium and the duration of the pulse also varied from 30-45 minutes. These observations are in contrast to the initial experiments performed by Gaimster and colleagues, where a concentration of 60 mM for approx. 20 minutes was reported (147).

Heterogeneity can be crucial in enhancing the adaptability and functionality of bacteria, enabling them to effectively respond and adapt to shifting conditions and environments (242). Tryptophanase activity has been identified as having bistability with two distinct populations: one with high tryptophanase activity and one with no/low activity (230). Bistability contributes to phenotype diversity within a clonal population and can be advantageous for the population's survival in fluctuating environments. Further, this bistability allows *E. coli* to rapidly react to changes in tryptophan and glucose availability. Cells with a high activity state can quickly produce indole and respond to their environment, while cells with low activity state conserve resources. This bistability was further confirmed by Figure 5.11, which showed two subpopulations of cells either large or small quantities of TnaA-GFP. The variability in tryptophanase expression among individual cells observed in the mother machine microfluidic device in Chapter 5, may constitute a form of "bet hedging" (243). Phenotypically distinct subpopulations of cells are produced within a bacterial community, and this diversity ensures that at least a few individual cells are better equipped to survive unpredictable environmental threats, such as nutrient depletion, heat stress, or even antibiotic treatment (244). The diversity observed in both the quantity of tryptophanase production and its temporal expression within the population might ensure the presence of a few cells that are capable of

surviving under a wide range of conditions. However, it is also possible that the heterogeneity observed is stochastic fluctuation that creates differential gene expression (245). The maintenance of phenotypically distinct subpopulations is a common property of bacterial populations (245). Perhaps the most notorious phenotypic variants are antibiotic persisters. Distinct sub-populations of slow or non-growing cells can survive treatment with antibiotics, as many of these drugs require metabolic activity in target bacterial cells (174).

Indole production is known to modulate biofilm production (169). Biofilms are crucial for the survival of many bacteria, and the complex, heterogeneous nature of biofilm populations contributes to their resilience and adaptability (246). The heterogenous nature of tryptophanase expression may allow the biofilm population to adapt to strenuous environments such as antibiotic challenge or challenge from the human immune system. Even in planktonic culture, wild-type *E. coli* which can produce indole survive longer term than indole-negative mutant cells (147, 172). As persistent indole signalling is experienced by all the cells in a population, cells that have limited amounts of tryptophanase themselves are still subject to persistent signalling. However, the cells that showed little to no tryptophanase expression in experiments from Chapter 5 would be unable to generate an indole pulse and would succumb more quickly to a period of long-term starvation.

7.2 A proposed role for tryptophanase activation in indole pulse regulation

Initial experiments in this study attempted to induce an indole pulse before the exponential to stationary phase transition (Figure 3.11). The lack of success indicated either that the timing of tryptophanase production could not be altered because it was tightly linked to the growth phase transition or that, even if enzyme production could be advanced, the enzyme did not become active until the transition. Figures 4.3-4.6 provide evidence that tryptophanase activity is much higher in stationary phase. The stationary phase cells convert approximately half of the available tryptophan to indole after 3 minutes, and approximately all the tryptophan after 22 min (Figure 4.6). In contrast, the exponential cells are much less proficient at converting tryptophan into indole (less than 10 percent)

after 22 minutes. Taken together, the data suggest that tryptophanase production precedes the indole pulse, but activation occurs only as the cells transition from exponential to stationary phase. A hypothesis incorporating these observations is proposed in Figure 7.1.

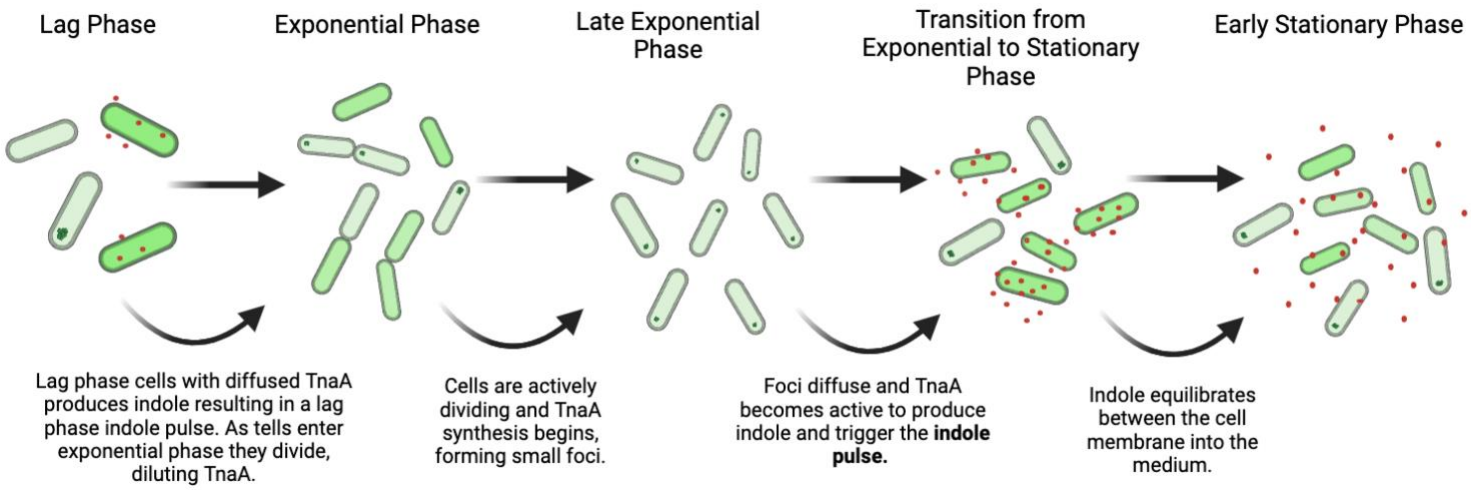


Figure 7.1. A new model for triggering the indole pulse by the dispersal of tryptophanase foci and activation of the enzyme. In lag phase, there is significant heterogeneity in the amount of TnaA-GFP and its location in the cell (foci vs dispersed). In lag phase, the cells with dispersed TnaA-GFP are slower to wake up than their foci counterparts (or cells without TnaA-GFP present), as the active TnaA-GFP converts tryptophan into indole forming a lag phase rise in both intracellular and extracellular indole. As the cells divide in early exponential phase, the TnaA-GFP is diluted and the amount of TnaA-GFP *per* cell declines. During mid-exponential phase many small foci begin to form as TnaA-GFP expression increases. These foci grow larger until the transition to stationary phase, when the TnaA-GFP is dispersed. This dispersion event activates TnaA-GFP, and the available tryptophan in the medium is converted to indole, forming the indole pulse. The indole eventually equilibrates between the cells and medium and TnaA-GFP expression rate slows. MinCDE may play a role in sequestering foci to the poles. The localization and formation of these foci could be mediated by interactions with components of the Min system or by the cellular environment shaped by MinCDE activity. Additionally, the regulation of tryptophanase activity through foci formation might be involved. *Generated in Biorender.com.*

As discussed in Chapters 4 and 5, freshly sub-cultured cells in lag phase may contain substantial amounts of tryptophanase that was produced during the previous stationary phase. Single-cell analysis reveals a broad distribution of TnaA-GFP (Figure 5.7), ranging

from cells with little or no fluorescence to cells with very high (>4000 a.u.) fluorescence. It is proposed that the lag phase cells with substantial amounts of diffuse tryptophanase produce indole from the tryptophan present in the fresh medium. This phenomenon could be called a “lag phase indole pulse”. Cells producing sufficient indole experience depolarisation of the cytoplasmic membrane. These cells take longer to “wake up” and enter exponential phase than the cells with lower amounts of diffuse tryptophanase, or cells with tryptophanase in inactive foci (Figure 5.10). As the cells divide in exponential phase, most of this inactive tryptophanase in foci is diluted into the new daughter cells (Figure 5.2 and 5.5). There is a small amount of inactive tryptophanase produced during early to mid-exponential phase that is likely not picked up by the mother machine’s automated identification of foci (Figure 5.6 and Figure 5.9). As the cells enter late exponential phase, tryptophanase expression is upregulated, and there is a sudden appearance of many foci (Figure 5.2 panel 5 and Figure 5.6). The tryptophanase in these foci is inactive. As the medium is depleted, and the cells make the transition from exponential to stationary phase, the bacteria experience changes in gene expression and decreased metabolic activity. These changes, by an unknown mechanism, initiate the dispersion of foci and the activation of tryptophanase. Once tryptophanase is dispersed and active, indole production is initiated, and the indole pulse occurs (Figure 5.6). The initiation of the pulse is seen by the highest production rate of indole occurring 30 minutes following the peak of foci seen (60% of cells). Following this peak, the cells immediately disperse their foci, revealing <10% of cells with foci remaining. During this period, the rate of indole production increases to its maximum, which is our proxy for the indole pulse. Following the initial surge in indole production, the cells continue to produce tryptophanase, but much more slowly. The exhaustion of tryptophan terminates the pulse and indole comes to equilibrium between cells and the culture medium. As cells progress into later stationary phase, the cells undergo phenotypic variation, resulting in divergent behaviours regarding the production of tryptophanase – some cells exhibit increased tryptophanase production, while others do not. A new cycle begins when these stationary phase cells are introduced into fresh medium.

7.2.1 Cellular distribution as a mechanism of enzyme regulation.

The mechanism outlined above substantiates preliminary research by Li and Young, who found that changes to amino acid residues influencing the formation of a TnaA focus also impacted enzyme activity. This result suggests that TnaA activity is reduced by the localization of TnaA-GFP in foci, possibly by occlusion of its active site (206). Li and Young also reported that mutations that affect the ability of foci to diffuse, delayed tryptophanase production (206). Figure 5.9 shows the production of TnaA-GFP in small foci that do not convert the available tryptophan into indole, and this is consistent with the hypothesis that foci are composed of inactive enzyme. These data are incorporated into the model seen in Figure 7.1, where formation of foci is a way of restricting the availability of active tryptophanase until the transition to stationary phase. It is possible that the enzyme's location and activity is linked to its multimeric state. For instance, tryptophanase may be active when it forms tetramers, which would correspond to its diffused state. Further, MinCDE system's involvement in maintain cellular polarity and regulating proper cellular division likely influences the localization and regulation of various cellular foci, including TnaA. This localization may be crucial for maintaining tryptophanase in its inactive state within the foci, and for dispersing the foci when TnaA activity is needed. Initial evidence for this was seen by Li and researchers, who saw that MinCDE mutants did not form discrete foci (228). However, additional research is required to substantiate such suggestions.

7.2.2 A lag phase indole pulse as a mechanism for regulating wake-up times.

The heterogeneity of TnaA content and distribution in lag phase cells is inherited from the stationary phase parent culture. Thus, cells in lag phase carry a memory of their previous state. A similar example of cellular memory has been proposed previously in the context of the indole pulse (180) and its effect upon the regulation of cytoplasmic pH.

If the TnaA-GFP foci are comprised of inactive enzyme, focus-containing cells will not produce the hypothetical lag phase indole pulse and should be faster to exit lag phase than cells that have dispersed tryptophanase and are capable of a pulse. This is supported by the data in Figure 5.13 where cells with dispersed tryptophanase exhibit a

delay in the time to first division. Lag phase production of indole in shake-flask culture was shown by Zarkan and colleagues, where it was demonstrated that there was a rapid production of indole (30 ± 10 μM) during lag phase (180). The physiological significance of this indole production is unclear; however, considering the delayed division response in these cells, it is plausible to hypothesize that it may be another example of a bet-hedging strategy against environmental stresses.

7.2.3 Indole signalling and persister formation.

A significant benefit of single-cell analysis over conventional bacterial culture is the visualization of outliers. Liu and colleagues used flow cytometry to identify cells that appeared to have much larger amounts of TnaA-GFP than the vast majority in the culture (158). In this work (Figures 5.7 and 5.9), cells on the end of the right tail would have much larger amounts of tryptophanase. Some of these cells did not divide during the growth experiment and this raises the possibility that they are persisters. Previous studies have proposed an indole-dependent mechanism for the formation of quinolone persisters (175) and preliminary evidence suggested that it was the indole pulse that played an important role. Based on the present study, it is hypothesized that the cells in stationary phase with the highest levels of tryptophanase initiate an extended pulse of indole when sub-cultured into fresh medium. These cells stay dormant, possibly due to the loss of trans-membrane potential and the inhibition of cell division. Such cells can be seen in images taken during the experiment (Figure 5.5).

7.3 Future areas of study and concluding remarks.

This work has suggested a new model for the triggering of the indole pulse, based on tryptophanase dispersal and activation. However, the mechanism by which TnaA-GFP foci are dispersed remains unclear. One potential explanation might involve a non-coding RNA similar to Rcd, that binds and activates tryptophanase in response to plasmid ColE1 multimerization (173). Alternatively, dispersion might be linked to metabolic changes associated with stationary phase entry, possibly involving factors such as RpoS (247). Further investigation is warranted to elucidate this mechanism.

A further exciting area for future study is the suggestion that the presence of large quantities of active tryptophanase might be responsible for the generation of persister cells. To address this question, experiments should be designed to identify lag-phase cells with high TnaA-GFP content and assess their ability to survive fluoroquinolone antibiotic treatment. Such experiments should be possible using the mother machine apparatus described in this work.

This thesis presents novel evidence supporting post-translational regulation of tryptophanase and its role in the regulation of indole production. It not only underscores the intricacy of tryptophanase production, but also its remarkably tight post-translational regulation. The importance of understanding tryptophanase production is underlined by its involvement in diverse biological processes, including biofilm formation and maintenance, virulence factor production, antibiotic resistance, persister cell formation, motility, and acid tolerance. Moreover, a deeper insight into tryptophanase production holds the potential to influence human health and disease outcomes, as the inhibition of *E. coli* tryptophanase has been proposed as a novel approach to the treatment of persistent urinary tract infections and disruptions in intestinal microbiota homeostasis.

References

1. Phillips RS, Demidkina TV, Faleev NG. Structure and mechanism of tryptophan indole-lyase and tyrosine phenol-lyase. Biochim Biophys Acta. 2003;1647(1-2):167-72.
2. Phillips RS, Johnson N, Kamath AV. Formation in vitro of hybrid dimers of H463F and Y74F mutant *Escherichia coli* tryptophan indole-lyase rescues activity with L-tryptophan. Biochemistry. 2002;41(12):4012-9.
3. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nature Medicine. 2018;24(4):392-400.
4. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nature Biotechnology. 2021;39(1):105-14.
5. Sender R, Fuchs S, Milo R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. Cell. 2016;164(3):337-40.
6. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, et al. The Human Microbiome Project: a community resource for the healthy human microbiome. PLoS biology. 2012;10(8):e1001377-e.
7. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. Nature. 2007;449(7164):804-10.
8. Moeller AH, Li Y, Mpoudi Ngole E, Ahuka-Mundeke S, Lonsdorf EV, Pusey AE, et al. Rapid changes in the gut microbiome during human evolution. Proceedings of the National Academy of Sciences. 2014;111(46):16431.
9. Mosca A, Leclerc M, Hugot JP. Gut Microbiota Diversity and Human Diseases: Should We Reintroduce Key Predators in Our Ecosystem? Frontiers in microbiology. 2016;7:455-.
10. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Human gut microbes associated with obesity. Nature. 2006;444(7122):1022-3.
11. Severino A, Tohumcu E, Tamai L, Dargenio P, Porcari S, Rondinella D, et al. The microbiome-driven impact of western diet in the development of noncommunicable chronic disorders. Best Practice & Research Clinical Gastroenterology. 2024:101923.
12. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. Gut. 2006;55(2):205-11.
13. Prioult G, Nagler-Anderson C. Mucosal immunity and allergic responses: lack of regulation and/or lack of microbial stimulation? Immunol Rev. 2005;206:204-18.
14. Wang C, Dong D, Strong PJ, Zhu W, Ma Z, Qin Y, Wu W. Microbial phylogeny determines transcriptional response of resistome to dynamic composting processes. Microbiome. 2017;5(1):103.
15. Tilg H, Moschen AR. Food, Immunity, and the Microbiome. Gastroenterology. 2015;148(6):1107-19.
16. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010;464(7285):59-65.
17. Ley RE, Peterson DA, Gordon JI. Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. Cell. 2006;124(4):837-48.

18. Graf D, Di Cagno R, Fåk F, Flint HJ, Nyman M, Saarela M, Watzl B. Contribution of diet to the composition of the human gut microbiota. Microb Ecol Health Dis. 2015;26:26164.
19. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, Mele MC. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. Microorganisms. 2019;7(1):14.
20. Laterza L, Rizzatti G, Gaetani E, Chiusolo P, Gasbarrini A. The Gut Microbiota and Immune System Relationship in Human Graft-versus-Host Disease. Mediterranean journal of hematology and infectious diseases. 2016;8(1):e2016025-e.
21. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473(7346):174-80.
22. Odamaki T, Kato K, Sugahara H, Hashikura N, Takahashi S, Xiao J-z, et al. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. BMC Microbiology. 2016;16(1):90.
23. Smith P, Willemsen D, Popkes M, Metge F, Gandiwa E, Reichard M, Valenzano DR. Regulation of life span by the gut microbiota in the short-lived African turquoise killifish. eLife. 2017;6:e27014.
24. Mariat D, Firmesse O, Levenez F, Guimarăes VD, Sokol H, Doré J, et al. The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. BMC Microbiology. 2009;9(1):123.
25. Ghosh TS, Das M, Jeffery IB, O'Toole PW. Adjusting for age improves identification of gut microbiome alterations in multiple diseases. eLife. 2020;9:e50240.
26. Jeffery IB, Lynch DB, O'Toole PW. Composition and temporal stability of the gut microbiota in older persons. The ISME Journal. 2016;10(1):170-82.
27. Partula V, Mondot S, Torres MJ, Kesse-Guyot E, Deschasaux M, Assmann K, et al. Associations between usual diet and gut microbiota composition: results from the Milieu Interieur cross-sectional study. Am J Clin Nutr. 2019;109(5):1472-83.
28. Albenberg LG, Wu GD. Diet and the intestinal microbiome: associations, functions, and implications for health and disease. Gastroenterology. 2014;146(6):1564-72.
29. Davis SC, Yadav JS, Barrow SD, Robertson BK. Gut microbiome diversity influenced more by the Westernized dietary regime than the body mass index as assessed using effect size statistic. Microbiologyopen. 2017;6(4).
30. Freeman J, Wilcox MH. Antibiotics and Clostridium difficile. Microbes and Infection. 1999;1(5):377-84.
31. Theriot CM, Koenigsknecht MJ, Carlson PE, Jr., Hatton GE, Nelson AM, Li B, et al. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. Nat Commun. 2014;5:3114.
32. Pérez-Cobas AE, Artacho A, Knecht H, Ferrús ML, Friedrichs A, Ott SJ, et al. Differential effects of antibiotic therapy on the structure and function of human gut microbiota. PloS one. 2013;8(11):e80201-e.
33. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc Natl Acad Sci U S A. 2007;104(34):13780-5.
34. Schulz MD, Atay Ç, Heringer J, Romrig FK, Schwitalla S, Aydin B, et al. High-fat-diet-mediated dysbiosis promotes intestinal carcinogenesis independently of obesity. Nature. 2014;514(7523):508-12.

35. Vrieze A, Van Nood E, Holleman F, Salojärvi J, Kootte RS, Bartelsman JF, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology. 2012;143(4):913-6.e7.
36. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444(7122):1027-31.
37. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, Schloss PD. The Gut Microbiome Modulates Colon Tumorigenesis. mBio. 2013;4(6):10.1128/mbio.00692-13.
38. Darkoh C, Plants-Paris K, Bishoff D, DuPont HL. Clostridium difficile Modulates the Gut Microbiota by Inducing the Production of Indole, an Interkingdom Signaling and Antimicrobial Molecule. mSystems. 2019;4(2):e00346-18.
39. Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet. 2022;399(10325):629-55.
40. Caldeira LF, Borba HH, Tonin FS, Wiens A, Fernandez-Llimos F, Pontarolo R. Fecal microbiota transplantation in inflammatory bowel disease patients: A systematic review and meta-analysis. PLoS One. 2020;15(9):e0238910.
41. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science. 2005;308(5728):1635-8.
42. Carabotti M, Scirocco A, Maselli MA, Severi C. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. Ann Gastroenterol. 2015;28(2):203-9.
43. Fuqua WC, Winans SC, Greenberg EP. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. J Bacteriol. 1994;176(2):269-75.
44. Nealson KH, Hastings JW. Bacterial bioluminescence: its control and ecological significance. Microbiol Rev. 1979;43(4):496-518.
45. Novick RP, Geisinger E. Quorum sensing in staphylococci. Annu Rev Genet. 2008;42:541-64.
46. Ng WL, Bassler BL. Bacterial quorum-sensing network architectures. Annu Rev Genet. 2009;43:197-222.
47. Williams P, Cámara M. Quorum sensing and environmental adaptation in *Pseudomonas aeruginosa*: a tale of regulatory networks and multifunctional signal molecules. Curr Opin Microbiol. 2009;12(2):182-91.
48. Dunny GM, Leonard BA. Cell-cell communication in gram-positive bacteria. Annu Rev Microbiol. 1997;51:527-64.
49. Federle MJ, Bassler BL. Interspecies communication in bacteria. J Clin Invest. 2003;112(9):1291-9.
50. Miller MB, Bassler BL. Quorum sensing in bacteria. Annu Rev Microbiol. 2001;55:165-99.
51. Schauder S, Shokat K, Surette MG, Bassler BL. The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. Mol Microbiol. 2001;41(2):463-76.
52. von Bodman SB, Willey JM, Diggle SP. Cell-cell communication in bacteria: united we stand. J Bacteriol. 2008;190(13):4377-91.

53. Waters CM, Bassler BL. Quorum sensing: cell-to-cell communication in bacteria. Annu Rev Cell Dev Biol. 2005;21:319-46.
54. Bassler BL, Wright M, Silverman MR. Multiple signalling systems controlling expression of luminescence in *Vibrio harveyi*: sequence and function of genes encoding a second sensory pathway. Molecular Microbiology. 1994;13(2):273-86.
55. Surette MG, Miller MB, Bassler BL. Quorum sensing in *Escherichia coli*, *Salmonella typhimurium*, and *Vibrio harveyi:* a new family of genes responsible for autoinducer production. Proc Natl Acad Sci U S A. 1999;96(4):1639-44.
56. Bansal T, Jesudhasan P, Pillai S, Wood TK, Jayaraman A. Temporal regulation of enterohemorrhagic Escherichia coli virulence mediated by autoinducer-2. Applied Microbiology and Biotechnology. 2008;78(5):811-9.
57. Sperandio V, Mellies JL, Nguyen W, Shin S, Kaper JB. Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic Escherichia coli. Proc Natl Acad Sci U S A. 1999;96(26):15196-201.
58. González Barrios AF, Zuo R, Hashimoto Y, Yang L, Bentley WE, Wood TK. Autoinducer 2 Controls Biofilm Formation in *Escherichia coli* through a Novel Motility Quorum-Sensing Regulator (MqsR, B3022). Journal of Bacteriology. 2006;188(1):305.
59. Xavier KB, Bassler BL. Regulation of uptake and processing of the quorum-sensing autoinducer AI-2 in Escherichia coli. J Bacteriol. 2005;187(1):238-48.
60. Sperandio V, Torres AG, Girón JA, Kaper JB. Quorum Sensing Is a Global Regulatory Mechanism in Enterohemorrhagic *Escherichia col*i; O157:H7. Journal of Bacteriology. 2001;183(17):5187.
61. Yadav MK, Vidal JE, Go YY, Kim SH, Chae SW, Song JJ. The LuxS/AI-2 Quorum-Sensing System of *Streptococcus pneumoniae* Is Required to Cause Disease, and to Regulate Virulence- and Metabolism-Related Genes in a Rat Model of Middle Ear Infection. Front Cell Infect Microbiol. 2018;8:138.
62. Yang L, Yuan T-j, Wan Y, Li W-w, Liu C, Jiang S, Duan J-a. Quorum sensing: a new perspective to reveal the interaction between gut microbiota and host. Future Microbiology. 2022;17(4):293-309.
63. Bansal T, Alaniz RC, Wood TK, Jayaraman A. The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation. Proc Natl Acad Sci U S A. 2010;107(1):228-33.
64. Coyte KZ, Rakoff-Nahoum S. Understanding Competition and Cooperation within the Mammalian Gut Microbiome. Curr Biol. 2019;29(11):R538-r44.
65. Rakoff-Nahoum S, Foster KR, Comstock LE. The evolution of cooperation within the gut microbiota. Nature. 2016;533(7602):255-9.
66. Germerodt S, Bohl K, Lück A, Pande S, Schröter A, Kaleta C, et al. Pervasive Selection for Cooperative Cross-Feeding in Bacterial Communities. PLoS Comput Biol. 2016;12(6):e1004986.
67. Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiology Letters. 2009;294(1):1-8.
68. Duncan SH, Louis P, Flint HJ. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. Appl Environ Microbiol. 2004;70(10):5810-7.

69. Shin JH, Tillotson G, MacKenzie TN, Warren CA, Wexler HM, Goldstein EJC. Bacteroides and related species: The keystone taxa of the human gut microbiota. Anaerobe. 2024;85:102819.
70. Fisher CK, Mehta P. Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. PLoS One. 2014;9(7):e102451.
71. Gutiérrez N, Garrido D. Species Deletions from Microbiome Consortia Reveal Key Metabolic Interactions between Gut Microbes. mSystems. 2019;4(4).
72. Srinivasan R, Kumawat DK, Kumar S, Saxena AK. Purification and characterization of a bacteriocin from *Lactobacillus rhamnosus* L34. Annals of Microbiology. 2013;63(1):387-92.
73. Theriot CM, Young VB. Interactions Between the Gastrointestinal Microbiome and Clostridium difficile. Annual Review of Microbiology. 2015;69(Volume 69, 2015):445-61.
74. Sockett RE. Predatory lifestyle of *Bdellovibrio bacteriovorus*. Annu Rev Microbiol. 2009;63:523-39.
75. Rotem O, Pasternak Z, Jurkevitch E. Bdellovibrio and Like Organisms. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, editors. The Prokaryotes: *Deltaproteobacteria* and *Epsilonproteobacteria*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 3-17.
76. Negus D, Moore C, Baker M, Raghunathan D, Tyson J, Sockett RE. Predator Versus Pathogen: How Does Predatory *Bdellovibrio bacteriovorus* Interface with the Challenges of Killing Gram-Negative Pathogens in a Host Setting? Annu Rev Microbiol. 2017;71:441-57.
77. Iebba V, Totino V, Santangelo F, Gagliardi A, Ciotoli L, Virga A, et al. *Bdellovibrio bacteriovorus* directly attacks *Pseudomonas aeruginosa* and *Staphylococcus aureus* Cystic fibrosis isolates. Front Microbiol. 2014;5:280.
78. Hsiao A, Ahmed AM, Subramanian S, Griffin NW, Drewry LL, Petri WA, Jr., et al. Members of the human gut microbiota involved in recovery from Vibrio cholerae infection. Nature. 2014;515(7527):423-6.
79. Mafra D, Barros AF, Fouque D. Dietary protein metabolism by gut microbiota and its consequences for chronic kidney disease patients. Future microbiology. 2013;8(10):1317-23.
80. Zelante T, Iannitti RG, Cunha C, De Luca A, Giovannini G, Pieraccini G, et al. Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. Immunity. 2013;39(2):372-85.
81. Lamas B, Richard ML, Leducq V, Pham HP, Michel ML, Da Costa G, et al. CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. Nat Med. 2016;22(6):598-605.
82. Proctor L. Priorities for the next 10 years of human microbiome research. Nature. 2019;569(7758):623-5.
83. Schroeder BO, Bäckhed F. Signals from the gut microbiota to distant organs in physiology and disease. Nature Medicine. 2016;22(10):1079-89.
84. Wu WK, Hsu CC, Sheen LY, Wu MS. Measurement of gut microbial metabolites in cardiometabolic health and translational research. Rapid Commun Mass Spectrom. 2020;34 Suppl 1:e8537.

85. Perino A, Demagny H, Velazquez-Villegas L, Schoonjans K. Molecular physiology of bile acid signaling in health, disease, and aging. Physiological Reviews. 2021;101(2):683-731.
86. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. Cell Research. 2020;30(6):492-506.
87. Kim CH. Immune regulation by microbiome metabolites. Immunology. 2018;154(2):220-9.
88. Gill PA, Inniss S, Kumagai T, Rahman FZ, Smith AM. The role of diet and gut microbiota in regulating gastrointestinal and inflammatory disease. Frontiers in Immunology. 2022;13:866059.
89. Choi BSY, Daoust L, Pilon G, Marette A, Tremblay A. Potential therapeutic applications of the gut microbiome in obesity: from brain function to body detoxification. International Journal of Obesity. 2020;44(9):1818-31.
90. Cummings JH, Macfarlane GT. Role of intestinal bacteria in nutrient metabolism. Clinical Nutrition. 1997;16(1):3-11.
91. Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, Tuohy K. Gut microbiota functions: metabolism of nutrients and other food components. Eur J Nutr. 2018;57(1):1-24.
92. Swer NM, Venkidesh BS, Murali TS, Mumbrekar KD. Gut microbiota-derived metabolites and their importance in neurological disorders. Molecular Biology Reports. 2023;50(2):1663-75.
93. Banfi D, Moro E, Bosi A, Bistoletti M, Cerantola S, Crema F, et al. Impact of Microbial Metabolites on Microbiota-Gut-Brain Axis in Inflammatory Bowel Disease. Int J Mol Sci. 2021;22(4).
94. Fassarella M, Blaak EE, Penders J, Nauta A, Smidt H, Zoetendal EG. Gut microbiome stability and resilience: elucidating the response to perturbations in order to modulate gut health. Gut. 2020.
95. von Frieling J, Fink C, Hamm J, Klischies K, Forster M, Bosch TCG, et al. Grow With the Challenge - Microbial Effects on Epithelial Proliferation, Carcinogenesis, and Cancer Therapy. Front Microbiol. 2018;9:2020.
96. Wikoff WR, Anfora AT, Liu J, Schultz PG, Lesley SA, Peters EC, Siuzdak G. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. Proceedings of the National Academy of Sciences. 2009;106(10):3698.
97. Dodd D, Spitzer MH, Van Treuren W, Merrill BD, Hryckowian AJ, Higginbottom SK, et al. A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. Nature. 2017;551(7682):648-52.
98. Gao J, Xu K, Liu H, Liu G, Bai M, Peng C, et al. Impact of the Gut Microbiota on Intestinal Immunity Mediated by Tryptophan Metabolism. Frontiers in Cellular and Infection Microbiology. 2018;8(13).
99. Guillemin GJ. Quinolinic acid, the inescapable neurotoxin. Febs j. 2012;279(8):1356-65.
100. Fernstrom JD. Role of precursor availability in control of monoamine biosynthesis in brain. Physiol Rev. 1983;63(2):484-546.
101. Piñero-Fernandez S, Chimerel C, Keyser UF, Summers DK. Indole transport across Escherichia coli membranes. J Bacteriol. 2011;193(8):1793-8.

102. Darkoh C, Chappell C, Gonzales C, Okhuysen P. A rapid and specific method for the detection of indole in complex biological samples. Appl Environ Microbiol. 2015;81(23):8093-7.
103. Huc T, Konop M, Onyszkiewicz M, Podsadni P, Szczepańska A, Turło J, Ufnal M. Colonic indole, gut bacteria metabolite of tryptophan, increases portal blood pressure in rats. Am J Physiol Regul Integr Comp Physiol. 2018;315(4):R646-r55.
104. Whitt DD, Demoss RD. Effect of microflora on the free amino acid distribution in various regions of the mouse gastrointestinal tract. Appl Microbiol. 1975;30(4):609-15.
105. Zarkan A, Liu J, Matuszewska M, Gaimster H, Summers DK. Local and Universal Action: The Paradoxes of Indole Signalling in Bacteria. Trends Microbiol. 2020;28(7):566-77.
106. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol. 2010;8(3):207-17.
107. Milani C, Ticinesi A, Gerritsen J, Nouvenne A, Lugli GA, Mancabelli L, et al. Gut microbiota composition and Clostridium difficile infection in hospitalized elderly individuals: a metagenomic study. Scientific Reports. 2016;6(1):25945.
108. Lee J-H, Lee J. Indole as an intercellular signal in microbial communities. FEMS Microbiology Reviews. 2010;34(4):426-44.
109. Shimada Y, Kinoshita M, Harada K, Mizutani M, Masahata K, Kayama H, Takeda K. Commensal Bacteria-Dependent Indole Production Enhances Epithelial Barrier Function in the Colon. PLOS ONE. 2013;8(11):e80604.
110. Chen D, Yang Z, Chen X, Huang Y, Yin B, Guo F, et al. The effect of *Lactobacillus rhamnosus* hsryfm 1301 on the intestinal microbiota of a hyperlipidemic rat model. BMC Complement Altern Med. 2014;14:386.
111. Chimerel C, Emery E, Summers DK, Keyser U, Gribble FM, Reimann F. Bacterial metabolite indole modulates incretin secretion from intestinal enteroendocrine L cells. Cell Rep. 2014;9(4):1202-8.
112. Powell DN, Swimm A, Sonowal R, Bretin A, Gewirtz AT, Jones RM, Kalman D. Indoles from the commensal microbiota act via the AHR and IL-10 to tune the cellular composition of the colonic epithelium during aging. Proceedings of the National Academy of Sciences. 2020;117(35):21519-26.
113. Ledala N, Malik M, Rezaul K, Paveglio S, Provatas A, Kiel A, et al. Bacterial Indole as a Multifunctional Regulator of *Klebsiella oxytoca* Complex Enterotoxicity. mBio. 2022;13(1):e03752-21.
114. Lee J, Bansal T, Jayaraman A, Bentley WE, Wood TK. Enterohemorrhagic *Escherichia coli* biofilms are inhibited by 7-hydroxyindole and stimulated by isatin. Appl Environ Microbiol. 2007;73(13):4100-9.
115. Rattanaphan P, Mittraparp-Arthorn P, Srinoun K, Vuddhakul V, Tansila N. Indole signaling decreases biofilm formation and related virulence of *Listeria* monocytogenes. FEMS Microbiol Lett. 2020;367(14).
116. Kohli N, Crisp Z, Riordan R, Li M, Alaniz RC, Jayaraman A. The microbiota metabolite indole inhibits *Salmonella virulence*: Involvement of the PhoPQ two-component system. PLoS One. 2018;13(1):e0190613.
117. Lee J, Attila C, Cirillo SL, Cirillo JD, Wood TK. Indole and 7-hydroxyindole diminish *Pseudomonas aeruginosa* virulence. Microb Biotechnol. 2009;2(1):75-90.

118. Oh S, Go GW, Mylonakis E, Kim Y. The bacterial signalling molecule indole attenuates the virulence of the fungal pathogen *Candida albicans*. J Appl Microbiol. 2012;113(3):622-8.
119. Hirakawa H, Kodama T, Takumi-Kobayashi A, Honda T, Yamaguchi A. Secreted indole serves as a signal for expression of type III secretion system translocators in enterohaemorrhagic *Escherichia coli* O157:H7. Microbiology. 2009;155(Pt 2):541-50.
120. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev. 1998;11(1):142-201.
121. Mir HD, Milman A, Monnoye M, Douard V, Philippe C, Aubert A, et al. The gut microbiota metabolite indole increases emotional responses and adrenal medulla activity in chronically stressed male mice. Psychoneuroendocrinology. 2020;119:104750.
122. Jaglin M, Rhimi M, Philippe C, Pons N, Bruneau A, Goustard B, et al. Indole, a Signaling Molecule Produced by the Gut Microbiota, Negatively Impacts Emotional Behaviors in Rats. Frontiers in Neuroscience. 2018;12(216).
123. Falconi CA, Junho C, Fogaça-Ruiz F, Vernier ICS, da Cunha RS, Stinghen AEM, Carneiro-Ramos MS. Uremic Toxins: An Alarming Danger Concerning the Cardiovascular System. Front Physiol. 2021;12:686249.
124. Roychowdhury P, Basak BS. The crystal strcuture of indole. Acta Crystallographica Section B. 1975(B31):1559-63.
125. Erb M, Veyrat N, Robert CAM, Xu H, Frey M, Ton J, Turlings TCJ. Indole is an essential herbivore-induced volatile priming signal in maize. Nature Communications. 2015;6(1):6273.
126. Knudsen JT, Eriksson R, Gershenzon J, Ståhl B. Diversity and distribution of floral scent. The Botanical Review. 2006;72(1):1.
127. Sun P, Huang Y, Yang X, Liao A, Wu J. The role of indole derivative in the growth of plants: A review. Front Plant Sci. 2022;13:1120613.
128. Dötterl S, Füssel U, Jürgens A, Aas G. 1,4-Dimethoxybenzene, a Floral Scent Compound in Willows that Attracts an Oligolectic Bee. Journal of Chemical Ecology. 2005;31(12):2993-8.
129. Zito P, Dötterl S, Sajeva M. Floral volatiles in a sapromyiophilous plant and their importance in attracting house fly pollinators. J Chem Ecol. 2015;41(4):340-9.
130. Lee J-H, Kim Y-G, Kim M, Kim E, Choi H, Kim Y, Lee J. Indole-associated predator–prey interactions between the nematode Caenorhabditis elegans and bacteria. Environmental Microbiology. 2017;19(5):1776-90.
131. Smith T. A modification of the method for determining the production of indole by bacteria. J Exp Med. 1897;2(5):543-7.
132. Wang D, Ding X, Rather PN. Indole can act as an extracellular signal in *Escherichia coli*. J Bacteriol. 2001;183(14):4210-6.
133. Isenberg HD, Sundheim LH. Indole reactions in bacteria. J Bacteriol. 1958;75(6):682-90.
134. Wood WA, Gunsalus IC, Umbreit WW. Function of the pyrixodal phosphate: resolution and purification of the tryptophanase enzyme of *Escherichia coli.* Journal of Biological Chemistry. 1947;170(1):313-21.
135. Stewart V, Yanofsky C. Role of leader peptide synthesis in tryptophanase operon expression in *Escherichia coli* K-12. Journal of Bacteriology. 1986;167(1):383-6.

136. Stewart V, Yanofsky C. Evidence for transcription antitermination control of tryptophanase operon expression in *Escherichia coli* K-12. Journal of Bacteriology. 1985;164(2):731-40.
137. Deeley MC, Yanofsky C. Transcription initiation at the tryptophanase promoter of *Escherichia coli* K-12. J Bacteriol. 1982;151(2):942-51.
138. Gong F, Yanofsky C. Analysis of tryptophanase operon expression in vitro: accumulation of TnaC-peptidyl-tRNA in a release factor 2-depleted S-30 extract prevents Rho factor action, simulating induction. J Biol Chem. 2002;277(19):17095-100.
139. Gong F, Ito K, Nakamura Y, Yanofsky C. The mechanism of tryptophan induction of tryptophanase operon expression: tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNA(Pro). Proc Natl Acad Sci U S A. 2001;98(16):8997-9001.
140. Bean RC, Shepherd WC, Chan H. Permeability of lipid bilayer membranes to organic solutes. J Gen Physiol. 1968;52(3):495-508.
141. Yanofsky C, Horn V, Gollnick P. Physiological studies of tryptophan transport and tryptophanase operon induction in *Escherichia coli*. J Bacteriol. 1991;173(19):6009-17.
142. Vega NM, Allison KR, Khalil AS, Collins JJ. Signaling-mediated bacterial persister formation. Nat Chem Biol. 2012;8(5):431-3.
143. Heatwole VM, Somerville RL. Cloning, nucleotide sequence, and characterization of mtr, the structural gene for a tryptophan-specific permease of *Escherichia coli* K-12. J Bacteriol. 1991;173(1):108-15.
144. Han TH, Lee J-H, Cho MH, Wood TK, Lee J. Environmental factors affecting indole production in Escherichia coli. Research in Microbiology. 2011;162(2):108-16.
145. Hirakawa H, Inazumi Y, Masaki T, Hirata T, Yamaguchi A. Indole induces the expression of multidrug exporter genes in *Escherichia coli*. Mol Microbiol. 2005;55(4):1113-26.
146. Chimerel C, Field CM, Piñero-Fernandez S, Keyser UF, Summers DK. Indole prevents *Escherichia coli* cell division by modulating membrane potential. Biochim Biophys Acta. 2012;1818(7):1590-4.
147. Gaimster H, Cama J, Hernandez-Ainsa S, Keyser UF, Summers DK. The indole pulse: a new perspective on indole signalling in *Escherichia coli*. PLoS One. 2014;9(4):e93168.
148. Domka J, Lee J, Wood TK. YliH (BssR) and YceP (BssS) regulate *Escherichia coli* K-12 biofilm formation by influencing cell signaling. Appl Environ Microbiol. 2006;72(4):2449-59.
149. Martino PD, Fursy R, Bret L, Sundararaju B, Phillips RS. Indole can act as an extracellular signal to regulate biofilm formation of *Escherichia coli* and other indole-producing bacteria. Can J Microbiol. 2003;49(7):443-9.
150. Vega NM, Allison KR, Samuels AN, Klempner MS, Collins JJ. *Salmonella typhimurium* intercepts *Escherichia coli* signaling to enhance antibiotic tolerance. Proceedings of the National Academy of Sciences. 2013;110(35):14420.
151. Smith JL. The role of gastric acid in preventing foodborne disease and how bacteria overcome acid conditions. J Food Prot. 2003;66(7):1292-303.
152. Foster JW. Escherichia coli acid resistance: tales of an amateur acidophile. Nature Reviews Microbiology. 2004;2(11):898-907.
153. Hirakawa H, Hayashi-Nishino M, Yamaguchi A, Nishino K. Indole enhances acid resistance in *Escherichia coli.* Microb Pathog. 2010;49(3):90-4.

154. Boon N, Kaur M, Aziz A, Bradnick M, Shibayama K, Eguchi Y, Lund PA. The Signaling Molecule Indole Inhibits Induction of the AR2 Acid Resistance System in *Escherichia coli*. Front Microbiol. 2020;11:474.
155. Baev MV, Baev D, Radek AJ, Campbell JW. Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of sugars, alcohols, and organic acids with transcriptional microarrays. Applied Microbiology and Biotechnology. 2006;71(3):310-6.
156. Juneja VK, Snyder OP, Jr., Marmer BS. Thermal destruction of *Escherichia coli* O157:H7 in beef and chicken: determination of D- and z-values. Int J Food Microbiol. 1997;35(3):231-7.
157. Lee HH, Molla MN, Cantor CR, Collins JJ. Bacterial charity work leads to population-wide resistance. Nature. 2010;467(7311):82-5.
158. Liu J, Summers D. Indole at low concentration helps exponentially growing Escherichia coli survive at high temperature. PLoS One. 2017;12(12):e0188853.
159. Masuda Y, Sakamoto E, Honjoh K-i, Miyamoto T. Role of Toxin-Antitoxin-Regulated Persister Population and Indole in Bacterial Heat Tolerance. Applied and Environmental Microbiology. 2020;86(16):e00935-20.
160. Balaban NQ, Helaine S, Lewis K, Ackermann M, Aldridge B, Andersson DI, et al. Definitions and guidelines for research on antibiotic persistence. Nature Reviews Microbiology. 2019;17(7):441-8.
161. Kuczyńska-Wiśnik D, Stojowska K, Matuszewska E, Leszczyńska D, Algara MM, Augustynowicz M, Laskowska E. Lack of intracellular trehalose affects formation of *Escherichia coli* persister cells. Microbiology. 2015;161(Pt 4):786-96.
162. Hu Y, Kwan BW, Osbourne DO, Benedik MJ, Wood TK. Toxin YafQ increases persister cell formation by reducing indole signalling. Environmental Microbiology. 2015;17(4):1275-85.
163. Kwan BW, Valenta JA, Benedik MJ, Wood TK. Arrested protein synthesis increases persister-like cell formation. Antimicrob Agents Chemother. 2013;57(3):1468-73.
164. Goode O, Smith A, Zarkan A, Cama J, Invergo BM, Belgami D, et al. Persister Escherichia coli Cells Have a Lower Intracellular pH than Susceptible Cells but Maintain Their pH in Response to Antibiotic Treatment. mBio. 2021;12(4):10.1128/mbio.00909-21.
165. Bamford RA, Smith A, Metz J, Glover G, Titball RW, Pagliara S. Investigating the physiology of viable but non-culturable bacteria by microfluidics and time-lapse microscopy. BMC Biology. 2017;15(1):121.
166. Wang P, Robert L, Pelletier J, Dang WL, Taddei F, Wright A, Jun S. Robust growth of Escherichia coli. Curr Biol. 2010;20(12):1099-103.
167. Römling U, Balsalobre C. Biofilm infections, their resilience to therapy and innovative treatment strategies. J Intern Med. 2012;272(6):541-61.
168. Lee JH, Wood TK, Lee J. Roles of indole as an interspecies and interkingdom signaling molecule. Trends Microbiol. 2015;23(11):707-18.
169. Hu M, Zhang C, Mu Y, Shen Q, Feng Y. Indole affects biofilm formation in bacteria. Indian J Microbiol. 2010;50(4):362-8.
170. Ito A, May T, Kawata K, Okabe S. Significance of rpoS during maturation of *Escherichia coli* biofilms. Biotechnol Bioeng. 2008;99(6):1462-71.
171. Martino PD, Merieau A, Phillips R, Orange N, Hulen C. Isolation of an *Escherichia coli* strain mutant unable to form biofilm on polystyrene and to adhere to human

pneumocyte cells: involvement of tryptophanase. Canadian Journal of Microbiology. 2002;48(2):132-7.
172. Gaimster H, Summers D. Regulation of Indole Signalling during the Transition of *E. coli* from Exponential to Stationary Phase. PLoS One. 2015;10(9):e0136691.
173. Chant EL, Summers DK. Indole signalling contributes to the stable maintenance of *Escherichia coli* multicopy plasmids. Mol Microbiol. 2007;63(1):35-43.
174. Lewis K. Persister cells, dormancy and infectious disease. Nature Reviews Microbiology. 2007;5(1):48-56.
175. Zarkan A, Matuszewska M, Trigg SB, Zhang M, Belgami D, Croft C, et al. Inhibition of indole production increases the activity of quinolone antibiotics against E. coli persisters. Scientific Reports. 2020;10(1):11742.
176. Gellert M, Mizuuchi K, O'Dea MH, Nash HA. DNA gyrase: an enzyme that introduces superhelical turns into DNA. Proc Natl Acad Sci U S A. 1976;73(11):3872-6.
177. Snyder M, Drlica K. DNA gyrase on the bacterial chromosome: DNA cleavage induced by oxolinic acid. J Mol Biol. 1979;131(2):287-302.
178. Field CM, Summers DK. Indole inhibition of ColE1 replication contributes to stable plasmid maintenance. Plasmid. 2012;67(2):88-94.
179. Sugino A, Peebles CL, Kreuzer KN, Cozzarelli NR. Mechanism of action of nalidixic acid: purification of Escherichia coli nalA gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. Proc Natl Acad Sci U S A. 1977;74(11):4767-71.
180. Zarkan A, Cano-Muniz S, Zhu J, Al Nahas K, Cama J, Keyser UF, Summers DK. Indole Pulse Signalling Regulates the Cytoplasmic pH of *E. coli* in a Memory-Like Manner. Sci Rep. 2019;9(1):3868.
181. Lan G, Tu Y. Information processing in bacteria: memory, computation, and statistical physics: a key issues review. Rep Prog Phys. 2016;79(5):052601.
182. Jõers A, Tenson T. Growth resumption from stationary phase reveals memory in Escherichia coli cultures. Scientific Reports. 2016;6(1):24055.
183. Snell EE. Tryptophanase: structure, catalytic activities, and mechanism of action. Adv Enzymol Relat Areas Mol Biol. 1975;42:287-333.
184. Robert G, Yadu T. Thermodynamics of enzyme catalyzed reactions: 2. Transferases. J. Phys. & Chem. Ref. Data (JPCRD), National Institute of Standards and Technology, Gaithersburg, MD; 1994.
185. Ku SY, Yip P, Howell PL. Structure of *Escherichia coli* tryptophanase. Acta Crystallogr D Biol Crystallogr. 2006;62(Pt 7):814-23.
186. TORAYA T, NIHIRA T, FUKUI S. Essential Role of Monovalent Cations in the Firm Binding of Pyridoxal 5′-Phosphate to Tryptophanase and β-Tyrosinase. European Journal of Biochemistry. 1976;69(2):411-9.
187. Yanofsky C. RNA-based regulation of genes of tryptophan synthesis and degradation, in bacteria. Rna. 2007;13(8):1141-54.
188. Yanofsky C. Attenuation in the control of expression of bacterial operons. Nature. 1981;289(5800):751-8.
189. Edwards RM, Yudkin MD. Location of the gene for the low-affinity tryptophan-specific permease of *Escherichia coli.* Biochemical Journal. 1982;204(2):617-9.

190. Gish K, Yanofsky C. Inhibition of expression of the tryptophanase operon in *Escherichia coli* by extrachromosomal copies of the tna leader region. J Bacteriol. 1993;175(11):3380-7.
191. Stewart V, Yanofsky C. Evidence for transcription antitermination control of tryptophanase operon expression in *Escherichia coli* K-12. J Bacteriol. 1985;164(2):731-40.
192. Botsford JL, DeMoss RD. Catabolite repression of tryptophanase in *Escherichia coli*. J Bacteriol. 1971;105(1):303-12.
193. Ward DF, Yudkin MD. Mutations in *Escherichia coli* that relieve catabolite repression of tryptophanase synthesis. Tryptophanase promoter-like mutations. J Gen Microbiol. 1976;92(1):133-7.
194. Bansal T, Englert D, Lee J, Hegde M, Wood TK, Jayaraman A. Differential effects of epinephrine, norepinephrine, and indole on *Escherichia coli* O157:H7 chemotaxis, colonization, and gene expression. Infect Immun. 2007;75(9):4597-607.
195. Cruz-Vera LR, Gong M, Yanofsky C. Changes produced by bound tryptophan in the ribosome peptidyl transferase center in response to TnaC, a nascent leader peptide. Proc Natl Acad Sci U S A. 2006;103(10):3598-603.
196. Stewart V, Landick R, Yanofsky C. Rho-dependent transcription termination in the tryptophanase operon leader region of *Escherichia coli* K-12. Journal of Bacteriology. 1986;166(1):217.
197. Konan KV, Yanofsky C. Rho-dependent transcription termination in the tna operon of *Escherichia coli:* roles of the boxA sequence and the rut site. J Bacteriol. 2000;182(14):3981-8.
198. Cruz-Vera LR, New A, Squires C, Yanofsky C. Ribosomal features essential for tna operon induction: tryptophan binding at the peptidyl transferase center. J Bacteriol. 2007;189(8):3140-6.
199. Nielsen BL, Willis VC, Lin CY. Western blot analysis to illustrate relative control levels of the lac and ara promoters in *Escherichia coli.* Biochem Mol Biol Educ. 2007;35(2):133-7.
200. Lacour S, Landini P. SigmaS-dependent gene expression at the onset of stationary phase in *Escherichia coli:* function of sigmaS-dependent genes and identification of their promoter sequences. J Bacteriol. 2004;186(21):7186-95.
201. Barbosa TM, Levy SB. Differential expression of over 60 chromosomal genes in *Escherichia coli* by constitutive expression of MarA. J Bacteriol. 2000;182(12):3467-74.
202. Patient ME, Summers DK. ColE1 multimer formation triggers inhibition of *Escherichia coli* cell division. Molecular Microbiology. 1993;9(5):1089-95.
203. Summers DK, Beton CW, Withers HL. Multicopy plasmid instability: the dimer catastrophe hypothesis. Mol Microbiol. 1993;8(6):1031-8.
204. Summers DK, Sherratt DJ. Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. Cell. 1984;36(4):1097-103.
205. Gaimster H, Summers D. Plasmids in the driving seat: The regulatory RNA Rcd gives plasmid ColE1 control over division and growth of its *E. coli* host. Plasmid. 2015;78:59-64.

206. Li G, Young KD. A new suite of tnaA mutants suggests that *Escherichia coli* tryptophanase is regulated by intracellular sequestration and by occlusion of its active site. BMC Microbiol. 2015;15(1):14.
207. Li G, Young KD. A cAMP-independent carbohydrate-driven mechanism inhibits tnaA expression and TnaA enzyme activity in *Escherichia coli*. Microbiology. 2014;160(9):2079-88.
208. Mohiuddin SG, Kavousi P, Orman MA. Flow-cytometry analysis reveals persister resuscitation characteristics. BMC Microbiology. 2020;20(1):202.
209. Dusny C, Grünberger A. Microfluidic single-cell analysis in biotechnology: from monitoring towards understanding. Current Opinion in Biotechnology. 2020;63:26-33.
210. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97(12):6640-5.
211. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006.0008.
212. Kim J, Webb AM, Kershner JP, Blaskowski S, Copley SD. A versatile and highly efficient method for scarless genome editing in *Escherichia coli* and *Salmonella enterica*. BMC biotechnology. 2014;14:84-.
213. Volkmer B, Heinemann M. Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. PLoS One. 2011;6(7):e23126.
214. Bakshi S, Leoncini E, Baker C, Cañas-Duarte SJ, Okumus B, Paulsson J. Tracking bacterial lineages in complex and dynamic environments with applications for growth control and persistence. Nature Microbiology. 2021;6(6):783-91.
215. Hardo G, Li R, Bakshi S. Quantitative Microbiology with Microscopy: Effects of Projection and Diffraction. bioRxiv. 2023:2023.05.15.540883.
216. Hardo G, Noka M, Bakshi S. Synthetic Micrographs of Bacteria (SyMBac) allows accurate segmentation of bacterial cells using deep neural networks. BMC Biology. 2022;20(1):263.
217. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, et al. Sex and virulence in *Escherichia coli:* an evolutionary perspective. Mol Microbiol. 2006;60(5):1136-51.
218. Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol. 2000;7(1-2):203-14.
219. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research. 2021;49(W1):W293-W6.
220. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256-w9.
221. Bernabeu M, Sánchez-Herrero JF, Huedo P, Prieto A, Hüttener M, Rozas J, Juárez A. Gene duplications in the *E. coli* genome: common themes among pathotypes. BMC Genomics. 2019;20(1):313.
222. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688-90.
223. Hillis DM, Huelsenbeck JP, Cunningham CW. Application and accuracy of molecular phylogenies. Science. 1994;264(5159):671-7.

224. Zhou Z, Alikhan NF, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res. 2018;28(9):1395-404.
225. Sezonov G, Joseleau-Petit D, D'Ari R. *Escherichia coli* physiology in Luria-Bertani broth. J Bacteriol. 2007;189(23):8746-9.
226. Isaacs H, Jr., Chao D, Yanofsky C, Saier MH, Jr. Mechanism of catabolite repression of tryptophanase synthesis in *Escherichia coli.* Microbiology (Reading). 1994;140 (Pt 8):2125-34.
227. Lelong C, Aguiluz K, Luche S, Kuhn L, Garin J, Rabilloud T, Geiselmann J. The Crl-RpoS regulon of *Escherichia coli.* Mol Cell Proteomics. 2007;6(4):648-59.
228. Li G, Young KD. Isolation and identification of new inner membrane-associated proteins that localize to cell poles in *Escherichia coli.* Mol Microbiol. 2012;84(2):276-95.
229. Summers DK. Plasmid Replication and its Control. The Biology of Plasmids1996. p. 31-64.
230. Orozco-Gómez DI, Sosa-Hernández JE, Gallardo-Navarro ÓA, Santana-Solano J, Santillán M. Bistable behaviour and medium-dependent post-translational regulation of the tryptophanase operon regulatory pathway in *Escherichia coli.* Scientific Reports. 2019;9(1):5451.
231. Wood TK, Knabel SJ, Kwan BW. Bacterial persister cell formation and dormancy. Appl Environ Microbiol. 2013;79(23):7116-21.
232. Baumler AJ, Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. Nature. 2016;535(7610):85-93.
233. Cameron EA, Sperandio V. Frenemies: Signaling and Nutritional Integration in Pathogen-Microbiota-Host Interactions. Cell Host Microbe. 2015;18(3):275-84.
234. Kumar A, Sperandio V. Indole Signaling at the Host-Microbiota-Pathogen Interface. mBio. 2019;10(3).
235. Levy M, Blacher E, Elinav E. Microbiome, metabolites and host immunity. Curr Opin Microbiol. 2017;35:8-15.
236. Li X, Zhang B, Hu Y, Zhao Y. New Insights Into Gut-Bacteria-Derived Indole and Its Derivatives in Intestinal and Liver Diseases. Front Pharmacol. 2021;12:769501.
237. Liu S, Feng J, Pu J, Xu X, Lu S, Yang J, et al. Genomic and molecular characterisation of *Escherichia marmotae* from wild rodents in Qinghai-Tibet plateau as a potential pathogen. Scientific Reports. 2019;9(1):10619.
238. Sarsero JP, Wookey PJ, Pittard AJ. Regulation of expression of the E*scherichia coli* K-12 mtr gene by TyrR protein and Trp repressor. J Bacteriol. 1991;173(13):4133-43.
239. Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. Science. 2007;315(5811):476-80.
240. Thomas CM, Nielsen KM. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol. 2005;3(9):711-21.
241. Conant GC, Wolfe KH. Turning a hobby into a job: How duplicated genes find new functions. Nature Reviews Genetics. 2008;9(12):938-50.
242. Dubnau D, Losick R. Bistability in bacteria. Molecular microbiology. 2006;61(3):564-72.
243. Veening JW, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. Annu Rev Microbiol. 2008;62:193-210.

244. Choudhary D, Lagage V, Foster KR, Uphoff S. Phenotypic heterogeneity in the bacterial oxidative stress response is driven by cell-cell interactions. Cell Reports. 2023;42(3).
245. Davis KM, Isberg RR. Defining heterogeneity within bacterial populations via single cell approaches. BioEssays. 2016;38(8):782-90.
246. Lewis K. Multidrug Tolerance of Biofilms and Persister Cells. In: Romeo T, editor. Bacterial Biofilms. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 107-31.
247. Zhou Y, Gottesman S. Regulation of proteolysis of the stationary-phase sigma factor RpoS. J Bacteriol. 1998;180(5):1154-8.