Investigation of the Regulation of the Lysine Decarboxylase LdcI Activity by the Alarmone ppGpp and MoxR Family AAA+ ATPase RavA

by

Boyu Zhao

A thesis submitted in conformity with the requirements for the degree of Master of Science

Department of Biochemistry
University of Toronto

© Copyright by Boyu Zhao (2012)
Investigation of the Regulation of the Lysine Decarboxylase LdcI Activity by the Alarmone ppGpp and MoxR Family AAA+ ATPase RavA

Boyu Zhao
Master of Science
Department of Biochemistry
University of Toronto
2012

Abstract

The lysine-dependent acid stress response system in *Escherichia coli* protects the cells under moderately acidic conditions. It consists of LdcI, the inducible lysine decarboxylase and CadB, the lysine-cadaverine antiporter. LdcI interacts with both ppGpp, the signalling molecule in the stringent response, and RavA, a MoxR-family AAA+ protein induced in the stationary phase. Experiments *in vitro* have shown that ppGpp inhibits the activity of LdcI and that the interaction between LdcI and RavA antagonizes the inhibition. In this work, it was demonstrated, by using a media shift assay, that the antagonistic regulation of RavA and ppGpp of LdcI activity also takes place *in vivo*, thereby linking acid resistance to the stringent response. As part of this study, components of the lysine decarboxylase pathway and the *ravA-viaA* operon were endogenously tagged with fluorescent proteins. These strains are useful tools to study the localization behaviour of these proteins under different stress conditions.
Acknowledgments

My graduate study has come to a close with the completion of this thesis. For the past three years in graduate school, there are many people I would like to thank, both for their diligence and intelligence which have inspired me to constantly improve myself, and for their patience and kindness which have supported me through my struggle.

First, I am happy to have been working in Dr. Walid Houry’s lab. I had the opportunity to be exposed to a variety of equipments and techniques and was always given full support and encouragements from Dr. Houry for doing any experiments and trying out ideas.

Secondly, I would like to thank Dr. Usheer Kanjee who has been a wonderful mentor. He tolerated my pestering amount of questions at the beginning and has always been very patient. This has been invaluable to me as it is encouraging and makes me not afraid of expressing my opinions in the subsequent years of study. His quasi-perfectionistic attitude also raised very good habits in me for keeping my working area clean and sterilized. I would also like to thank all members of the lab, Dr. Majida El Bakkouri, Jen Huen, Dr. Yoshito Kakiha, Elisa Leung, Kaiyin Liu, Nardin Nano, Deji Ologbenla, Keith Wong, Angela Yu and Liang Zhao for building a very friendly environment for me to study where I feel comfortable for asking for help.

I would also like to thank everyone in the Biochemistry Graduate Student Union (BGSU), volleyball team and softball team. I learnt a lot through organizing, participating and hosting a variety of departmental events such as the biochemistry open house, student lunch and Wednesday student seminars. They are wonderful opportunities. Of course, I am always excited to learn new sports. They build up my confidence and make me feel more energetic and positive amongst the sometime emotionally tumultuous periods in research, as well as in personal life.

Lastly, I would like to thank my parents for their unconditional love and support.

Overall, I can say that I am happy with my improvements, my achievements and my learning experiences in graduate school. It has been great time learning from everyone and I will keep the skills and experiences and try to apply them well for my future.
# Table of Contents

## Contents

Acknowledgments .................................................................................................................. iii

Table of Contents ................................................................................................................ iv

List of Tables .......................................................................................................................... vii

List of Figures ....................................................................................................................... viii

Abbreviations ....................................................................................................................... x

1 Introduction .......................................................................................................................... 1

   1.1 Acid Stress Response in *Escherichia coli* ................................................................. 2

      1.1.1 Enzyme-based Acid Stress Response Systems in *E. coli* ............................... 3

         1.1.1.1 AR2: the Glutamate Decarboxylase System .............................................. 5

         1.1.1.1.1 Structures of the Glutamate Decarboxylases ....................................... 5

         1.1.1.2 Regulation of the Glutamate Decarboxylase System ............................... 11

         1.1.1.2 AR3: the Arginine Decarboxylase System ................................................. 17

         1.1.1.3 AR4: the Lysine Decarboxylase System .................................................... 19

         1.1.1.3.1 LdcI (CadA) ......................................................................................... 20

         1.1.1.3.2 CadB ................................................................................................. 21

         1.1.1.3.3 CadC ................................................................................................. 23

         1.1.1.4 AR5: the Ornithine Decarboxylase System ................................................. 24

      1.1.2 The Role of Chloride and Potassium Ions in Acid Resistance .......................... 24

      1.1.3 Chaperone-based Acid Stress Response .............................................................. 25

      1.1.4 The Acid Fitness Island in *E. coli* .................................................................... 28

      1.1.5 Acid Stress Response in Other Enterobacteria ............................................... 30
1.2 The Stringent Response and ppGpp ................................................................. 35
  1.2.1 Synthesis of ppGpp by RSH Proteins ......................................................... 35
  1.2.2 Cellular Adaptations Mediated by ppGpp in the Stringent Response .............. 40
1.3 MoxR Family AAA+ ATPase ........................................................................ 46
  1.3.1 Introduction to AAA+ ATPases ................................................................. 46
  1.3.2 Introduction of RavA and ViaA ................................................................. 49
    1.3.2.1 RavA ................................................................................................. 49
    1.3.2.2 ViaA ................................................................................................. 50
1.4 Thesis Rationale ............................................................................................ 52
2 Antagonistic Regulation of LdcI by ppGpp and RavA ........................................ 53
  2.1 Introduction .................................................................................................... 54
  2.2 Materials and Methods ............................................................................... 61
    2.2.1 Generation of the *AcadBA* Strains ......................................................... 61
      2.2.1.1 Preparation of the Kan\textsuperscript{R} Knockout Cassette from pKD4 ......... 61
      2.2.1.2 DY330 Transformation ..................................................................... 65
      2.2.1.3 P1 Transduction ............................................................................. 67
      2.2.1.4 Elimination of Kan\textsuperscript{R} Cassette ............................................. 67
    2.2.2 Generation of Complementation Strains ................................................. 68
    2.2.3 Media Shift Assay .................................................................................... 68
      2.2.3.1 Media Shift Assay to Test the Inhibition of LdcI by ppGpp ................. 68
      2.2.3.2 Media Shift Assay to Test the in vivo Effect of RavA ......................... 69
  2.3 Results ........................................................................................................... 69
    2.3.1 Inhibition of LdcI by ppGpp in vivo ......................................................... 69
    2.3.2 Antagonistic Regulation of RavA and ppGpp on the Activity of LdcI ........... 75
  2.4 Discussion ..................................................................................................... 77
3 Cellular Localization of LdcI, CadB, CadC, RavA and ViaA ................................................. 80

3.1 Introduction ......................................................................................................................... 81

3.2 Materials and Methods ................................................................................................. 82

3.2.1 Overall Strategy for Endogenously Tagging ldcI, cadB, cadC, ravA and viaA ... 82

3.2.2 Cloning sfGFP, mRFP and YFP into the pKD3 Plasmid ............................................. 84

3.2.3 Preparation of the DNA Cassettes for Tagging ldcI, cadB, cadC, ravA and viaA .......... 84

3.2.4 DY330 Transformation ................................................................................................. 86

3.2.5 P1 Transduction ............................................................................................................. 86

3.2.6 Elimination of CamR Cassette ..................................................................................... 87

3.2.7 Western Blot ................................................................................................................... 87

3.2.8 Fluorescence Microscopy ............................................................................................. 88

3.2.9 Media Shift Assay ......................................................................................................... 88

3.3 Results ............................................................................................................................... 89

3.3.1 LdcI (CadA) .................................................................................................................. 89

3.3.2 CadB ............................................................................................................................. 94

3.3.3 CadC ............................................................................................................................. 98

3.3.4 RavA ............................................................................................................................ 101

3.3.5 ViaA ............................................................................................................................ 104

3.4 Discussion ......................................................................................................................... 107

4 Future Directions ............................................................................................................... 110

References .............................................................................................................................. 112
List of Tables

Table 1: Comparison of decarboxylase systems in different enteric bacteria. ........................................ 34
Table 2: List of plasmids used in the thesis work. ...................................................................................... 62
Table 3: List of primers used in the thesis work. ....................................................................................... 63
Table 4: List of *E. coli* strains used in the thesis work. ........................................................................... 66
Table 5: Primers and plasmids used for tagging *ldcI, cadB, cadC, ravA* and *viaA*. .............................. 85
List of Figures

Figure 1: Schematic diagram of the Acid Resistance (AR) systems in *E. coli* ................................. 4

Figure 2: Domain arrangement of the acid stress induced amino acid decarboxylases and chaperones ........................................................................................................................................... 7

Figure 3: X-ray structures of the acid stress induced amino acid decarboxylases, antiporters, and chaperones ..................................................................................................................................................... 10

Figure 4: Schematic depiction of the gene organization and regulation of the different decarboxylase systems .............................................................................................................................................. 16

Figure 5: Structures of the components of the lysine decarboxylase system. ........................................ 22

Figure 6: The acid fitness island in *E. coli* ............................................................................................ 29

Figure 7: Domain arrangements and structure of RSH family proteins. ............................................. 37

Figure 8: Schematic diagram of the stringent response in *E. coli*. ................................................... 41

Figure 9: The classical AAA+ ATPases domain structure. ................................................................. 48

Figure 10: Structure and domain arrangements of RavA and ViaA ...................................................... 51

Figure 11: Inhibition of LdcI by ppGpp *in vitro* .................................................................................. 56

Figure 12: RavA antagonizes the inhibition of LdcI by ppGpp *in vitro*. ......................................... 60

Figure 13: Inhibition of LdcI by ppGpp *in vivo*. ................................................................................ 72

Figure 14: Media Shift Assay on ΔrelAΔspoTΔcadBA strains. ............................................................ 74

Figure 15: RavA antagonizes the inhibitory effect of ppGpp on LdcI activity *in vivo*. ....................... 76

Figure 16: Physiology significance of the inhibition of LdcI by ppGpp. ............................................. 79
Figure 17: Schematic depiction of the cloning of sfGFP, mRFP and YFP into the pKD3 plasmid.
................................................................. 83

Figure 18: Endogenous tagging of LdcI with sfGFP. ................................................................. 91

Figure 19: Verification of MG1655 ldcI-(6Gly)-sfGFP strains.................................................. 93

Figure 20: Endogenous tagging of CadB with mRFP. ................................................................. 95

Figure 21: Verification of MG1655 cadB-(6Gly)-mRFP strains.................................................. 97

Figure 22: Endogenous tagging of CadC with sfGFP. ................................................................. 99

Figure 23: Verification of MG1655 cadC-(6Gly)-sfGFP strains.................................................. 100

Figure 24: Endogenous tagging of RavA with mRFP ................................................................. 102

Figure 25: Verification of MG1655 ravA-(6Gly)-mRFP strains.................................................. 103

Figure 26: Endogenous tagging of ViaA with YFP................................................................. 105

Figure 27: Verification of MG1655 viaA-(6Gly)-YFP strains.................................................. 106

Figure 28: Individual aggregates are located to the old pole through cycles of cell divisions. .. 109
Abbreviations

AAA+ - ATPase associated with diverse cellular activities

ΔΨ - transmembrane electrical potential

AR1 - acid resistance system 1

AR2 - acid resistance system 2

AR3 - acid resistance system 3

AR4 - acid resistance system 4

AR5 - acid resistance system 5

ATP - adenosine triphosphate

CRP - cAMP receptor protein

DNA - deoxyribonucleic acid

Dps - DNA-binding protein from starved cells

EM - electron microscopy

GABA - γ-amino butyric acid

GDP - guanosine diphosphate

GTP - guanosine triphosphate

H⁺ - proton

LARA - LdcI associating domain of RavA

LB - lysogeny broth

LPS - lipopolysaccharide
LRP-leucine response protein

MIDAS-metal-ion-dependent adhesion site

PLP-pyridoxal-5'-phosphate

PLP-SD-PLP-binding subdomain

polyP-polyphosphate

ppGpp-guanosine tetraphosphate or guanosine 3', 5'-bis (diphosphate)

pppGpp-guanosine pentaphosphate or guanosine 3'-diphosphate, 5'- triphosphate

RavA-regulatory ATPase variant A

RMF-ribosome modulating factor

RNA-ribonucleic acid

RNAP-RNA polymerase

rRNA-ribosomal RNA

RSH-RelA/SpoT homologs

SD4-sub-domain 4

SDS-PAGE-sodium dodecyl sulfate polyacrylamide gel electrophoresis

TGS-threonine aminoacyl-rTNR synthetase ThrRS, GTPases and SpoT small domain

TNBS-trinitrobenzenesulfonic acid

ViaA-VWA interacting with AAA+ ATPase

VWA-Von Willebrand factor type A
1 Introduction


Note: Sections 1.1.2 to 1.1.5 cover materials beyond the scope of this thesis. It is included because it was an integral part of the review. They provide a more comprehensive understanding of the acid stress system in *E. coli* and other enteric bacteria.
The acid stress response system provides the ability and resilience to enteric bacteria to survive at low pH, from extremely acidic conditions (pH 2–3) to moderately acidic conditions (pH 4–5) [1]. It also contributes to the virulence of many pathogenic bacteria such as the *Vibrio* species, *Shigella* species, and *Salmonella* species by allowing them to survive and replicate in the acidic environments in gastric and urinary systems to cause infections. To date, most of the details of the acid stress response have been obtained from studies performed with wildtype *Escherichia coli*, one of the most common model organisms in the study of bacteria in laboratories.

The lysine-dependent acid stress response system is one of the major systems involved in the bacterial acid stress response. It protects bacteria under moderate acidic conditions. The lysine decarboxylase LdcI (or CadA) is a major component of this pathway. Based on previous work from our group, LdcI has been found to interact, *in vitro*, with both ppGpp, the signaling molecule in the stringent response in *E. coli*, and with RavA, a MoxR-family AAA+ ATPase [2, 3]. The introduction chapter provides detailed background information on acid stress response, the stringent response and the AAA+ ATPase RavA. It serves as a context to the main focus of my thesis work, which is to understand the regulation and localization of LdcI under various stress conditions *in vivo*. In this chapter, the acid stress response is discussed first, with detailed descriptions of the structures, functional mechanisms and regulations of its major components. Secondly, the stringent response is introduced with respect to the upstream signals that trigger the stringent response, the synthesis of ppGpp and the downstream effects of ppGpp signalling. Lastly, an overview on the AAA+ domain structure and RavA is presented.

1.1 Acid Stress Response in *Escherichia coli*

The acid stress response consists of both enzyme-based and chaperone-based systems, both of which are under complex regulatory networks that encompass chaperones, small non-coding RNAs, and protein factors involved in a variety of cellular functions such as DNA topology modulation, superoxide stress, ion concentration responses, and cellular transport [1]. The complex machinery of acid stress response can thus be seen as a sign of the absolute necessity for this response to initiate accurately and function properly for cell survival.

In this part of the introduction, the decarboxylase-based and chaperone-based pathways in *Escherichia coli* are described in detail first. Subsequently, as the glutamate, arginine, and lysine decarboxylase systems are shared in other enteric bacteria species, the acid stress systems are
compared and contrasted among five enteric bacterial species: *Escherichia coli*, *Vibrio cholera*, *Vibrio vulnifus*, *Shigella flexneri*, and *Salmonella typhimurium* [1].

### 1.1.1 Enzyme-based Acid Stress Response Systems in *E. coli*

Five acid resistance (AR) pathways, named AR1 to 5, have been identified in *E. coli* (Figure 1) [4, 5]. The AR1 pathway, though poorly understood, is activated when cells are placed in minimal media at pH 2.5 without external supply of any amino acid after the cells are grown at pH 5.5 to stationary phase in Lysogeny Broth (LB) buffered media [5, 6]. This pathway requires the alternative sigma factor $\sigma^S$ and cAMP receptor protein (CRP). Because CRP is involved, the system is repressed by glucose [5]. The AR2-5 pathways are all decarboxylase-based pathways. Each of them consists of a decarboxylase enzyme that is induced by both low pH and the presence of a specific amino acid, and an antiporter. The enzymes involved in AR 2, 3, 4 and 5 are the glutamate, arginine, lysine and ornithine decarboxylases, respectively. The AR2 and AR3 pathways protect bacteria in extreme acidic environment (e.g. pH 2.5), while the AR4 and AR5 pathways protect them in moderately acidic environment (e.g. pH 4.5) [4, 5].
**Figure 1: Schematic diagram of the Acid Resistance (AR) systems in *E. coli*.

The amino acid and decarboxylation products are shown in chemical notation and the proteins responsible for the reactions are shown under the reaction arrows. All the decarboxylation reactions consume a proton and release carbon dioxide. AR1 is not shown because the system is not well characterized. The HdeA/B and Hsp31 chaperones are represented by ovals. This figure is adapted from [5, 7, 8].
1.1.1.1 AR2: the Glutamate Decarboxylase System

This system has been extensively studied and constitutes one of the major acid stress response systems in *E. coli* under extreme acidic environment. The system consists of paralogous GadA and GadB decarboxylases and an inner-membrane antiporter GadC [5]. GadA and GadB are pyridoxal 5′-phosphate (PLP)-dependent enzymes that convert glutamate to gamma-amino butyric acid (GABA) and carbon dioxide (CO$_2$) in a reaction that consumes a cytoplasmic proton (Figure 1) [5, 9]. GABA is transported out of the cell by the inner membrane antiporter GadC in exchange for more glutamate from the external medium [9].

1.1.1.1.1 Structures of the Glutamate Decarboxylases

The structures of both GadA and GadB have been solved by X-ray crystallography [10, 11]. The two isozymes are different in primary sequence at only five residues and, hence, are very similar in structure. They form 330 kDa hexamers assembled from trimerization of GadA (B) dimers with pH optimum of 4 [10]. The GadA (B) monomer can be divided into three domains: the N-terminal domain, the large domain (PLP binding domain), and the C-terminal small domain (Figure 3). The N-terminal domain is critical for the function of GadB because it is responsible for the preferential association of GadB with the inner membrane when pH decreases [10]. Deletion of the first 14 residues of GadB diminishes its ability to migrate to the inner membrane.

By comparing the structures of GadB at neutral (pH 7.6) and acidic pHs (pH 4.6), the N-terminal domain is found to undergo a conformational change from a disordered state containing little secondary structure at neutral pH to an α-helix structure at acidic pH (Figure 3) [10]. This α-helix is oriented perpendicularly to the subunit surface. As a result, the active (low pH) form of GadB hexamer has a three-helical bundle on each of its two opposing surfaces (Figure 3). The bundles have a hydrophobic core and are charged on the outside. There are three Asp residues and one Glu residue in the first 15 residues of GadB. At least two of them are protonated upon acidification, resulting in the conformational changes in GadB N-terminus [10].

More interestingly, halide ions such as Cl$^-$ are found, by X-ray crystallography, to be able to bind to the bottom of the the two triple-helix bundles formed by the N-terminal domain [12]. The binding fixes the turn formed by residues 16-19 and, as a result, stabilizes the triple α-helix bundle required for GadB hexamer interaction with the inner membrane [12]. In addition, presence of Cl$^-$ ions increase the pH at which GadB becomes activated to near pH 6. This allows
GadB to start functioning before the intracellular pH becomes extremely acidic. These findings demonstrate the significance of importing Cl⁻ ions in addition to its role in protecting the membrane potential during extreme acid stress (see Section 1.1.5).
Figure 2: Domain arrangement of the acid stress induced amino acid decarboxylases and chaperones.

The domain boundaries for *E. coli* GadA/B, AdiA, LdcI, HdeA, HdeB and Hsp31 are based on the solved X-ray structures. For *E. coli* SpeF, the domain boundary is defined based on sequence alignment with *Lactobacillus* 30a SpeF whose X-ray structure has been solved [13].
The X-ray structure of GadA hexamer was so far only solved at pH 4.6 which also forms the helix structure at its N-terminus. Therefore, although similar biochemical experiments to GadB have not yet been performed with GadA, it is highly likely that N-terminal domain in GadA also undergoes a conformational change from a disordered state at neutral pH to an α-helix structure at acidic pH [11].

The active site of GadB (as well as of GadA) locates in the PLP-binding large domain. Lys276, in GadB (GadA), forms a Schiff base linkage with the C4 atom of the pyridine ring of PLP. Unlike other PLP-dependent enzymes, GadB uses Gln163 instead of an aromatic residue for a stacking interaction with the pyridine ring of PLP [10]. When the glutamate substrate interacts with the active site of GadB, it is held in place by a hydrogen bond between its γ-carboxylate group and the Phe63 main chain, the Thr62 side chain, and the carboxylate side chain of Asp86 of the neighboring subunit [10]. This binding explains the maximum activity of GadB at low pH, since either the Glu substrate or Asp86 of GadB must be protonated for this interaction to occur. In addition, an Arg422 residue that binds the α-carboxylate of the substrate in many PLP-dependent enzymes is kept away from interacting with the Glu substrate in GadB, so that it does not interfere with the decarboxylation process [10].

When the pH increases to neutral, it is speculated that GadB undergoes conformational changes to its inactive form in the following steps. At first, each of the N-terminal triple helical bundles unfolds independently. When both are unfolded, an aldimine structure (imine derived from an aldehyde) forms between the imadizole ring of His465 at the C-terminal end of GadA/B and Lys276-PLP imine to close the active site [12]. This covalent adduct is the 340nm-absorbing chromophore that is the signature of the inactive form of GadB [12]. Hence, at neutral pH, each active site funnel is blocked by the C-terminus of the same subunit and a β-hairpin from the neighboring subunit. The structure of GadC from *E. coli* O157:H7 has been solved recently [14]. The GABA-glutamine antiporter has 12 transmembrane helices. At below pH 6.5, the GadC adopts an inward-open conformation that exposes negative-charged binding pocket to bind substrates in the cytoplasm. Once the pH increases to above 6.5, a folded domain named C-plug which is formed by the last 50 residues at the C-terminal domain, closes the pocket to form the inactive conformation of GadC. The C-plug domain contains basic residues and is required for the pH-dependent activity of GadC. The GadC structure was solved at pH 8 and shows the closed and inactive conformation of GadC (Figure 3).
Decarboxylases

GadA (Top View)
pH 4.6

GadB (Side View)
pH 4.6, pH 7.6

Membrane Transporters

AdiC (Side View)
GadC (Monomer)

1H⁺/2Cl⁻ channel (Side View)

Chaperones

AdiA (Top View)

Ldcl (Top View)

SpeF (Top View)

HdeA (Side View)
HdeB (Side View)

Hsp31 (Side View)
Figure 3: X-ray structures of the acid stress induced amino acid decarboxylases, antiporters, and chaperones.

The domains of the decarboxylases and chaperones are colour-coded according to Figure 2. The structures of *E. coli* GadA [PDB ID:1XEY [11]], *E. coli* GadB [1PMO – inactive form at pH 7.6; 1PMM – active form at pH 4.6 [10]], *E. coli* AdiA [2VYC [15]], *E. coli* LdcI [3Q16 [2]] and *Lactobacillus* 30a OrnDC [1ORD [13]] are displayed in the same column for comparison. The X-ray structures of the *E. coli* AdiC antiporter [3H5M [16]], *E. coli* ClC Cl channel [1KPK [17]] and *E. coli* O157: H7 GadC antiporter [4DJL [14]] are also shown. *E. coli* HdeA [1DJ8 [7]], HdeB [2XUV,[18]] and *E. coli* Hsp31 [1ONS [8]] are shown in their physiological dimeric form. This figure was prepared using *PyMOL* (DeLano 2002).
1.1.1.1.2 Regulation of the Glutamate Decarboxylase System

As the most effective acid stress response pathway under extreme acid stress conditions, the AR2 system is intricately regulated. To date, there are over 20 proteins and three small non-coding RNAs that are identified to regulate the Gad system (Figure 4A). The proteins and factors to be discussed in details below include: CRP, Dps, EvgA/S, GadE, GadX, GadW, H-NS, Lon, PhoP/Q, RNaseE, $\sigma^{70}$, $\sigma^{8}$, SspA, TrmE, TopA, TorS/R, and YdeO. The three small non-coding RNAs are DsrA, GadY, and GcvB.

In the *E. coli* genome, the *gadA* and *gadB* genes are located 2100 kbp apart (Figure 4A). The *gadC* gene is located downstream of *gadB* and they form an operon [11]. The central transcriptional activator of the *gadA* and *gadBC* genes is GadE [5, 19]. GadE induces the expression of the Gad system by binding to a 20 bp sequence, termed the *gad* box, which is located around 60 bp upstream of the transcription start sites of *gadA* and *gadBC* [20, 21].

Apart from the Gad system, GadE also serves as a global transcriptional activator for many genes. When overexpressed, GadE is found to induce genes involved in stress responses such as high osmolarity and acid stress (e.g. *osmC*, *hdeA*, and *ycgG*), in biosynthesis of glutamate (e.g. *gltD* and *gltH*), and also in LPS (lipopolysaccharide) biosynthesis (e.g. *rcsA* and *rfaG*) [22].

Other protein factors regulate, directly or indirectly, the expression of GadE depending on the growth phase of the cells and on the media (Figure 4A). There is a 750 bp regulatory region upstream of the transcription start site of the *gadE* gene [23]. This region contains three promoters (P1, P2, and P3) allowing different regulators to bind and produce three *gadE* transcripts: T1, T2, and T3. T1 starts at -124, T2 at -324 and T3 at -566 position from the *gadE* start codon +1. The P1, P2 and P3 promoters are located about 200 bp upstream of the start of each transcript. P1 is where GadE acts to auto induce itself in minimal medium containing glucose [23, 24]. This auto activation also requires $\sigma^{8}$, the alternative sigma factor responsible for the transcription of many stress response genes in stationary phase, and another unidentified factor. P2 and P3 are activated by GadX (YhiX) and GadW (YhiW) during stationary phase growth and by the EvgA/S and YdeO pathway during exponential phase growth in minimal medium at acidic pH [24].

The GadX and GadW are two AraC-like regulators; AraC is a well-studied transcription
activator of the arabinose operon that encodes proteins involved in metabolizing L-arabinose. Along with $\sigma^8$, GadX and GadW are important for inducing the Gad system in stationary phase cells grown in either minimal or rich media [21]. The transcription of $\text{gad}X$ is induced by $\sigma^8$ (Figure 4A). CRP and H-NS proteins are two repressors of $\text{gad}X$ transcription. CRP represses the production of $\sigma^8$ during normal cell growth and, thus, indirectly represses $\text{gad}X$ transcription [20]. H-NS is a major component of bacterial chromatin. It preferentially binds AT-rich DNA sequences often found in $E. coli$ promoters, to repress expression of downstream genes. H-NS is a repressor of $\text{gad}A$, $\text{gad}E$, and $\text{gad}X$, but not of $\text{gad}BC$ [25]. Several in vitro experiments showed that GadW and GadX can bind directly to the promoter regions of $\text{gad}A$ and $\text{gad}BC$ and induce their expression through different mechanisms [21, 25, 26]. However, it seems that in vivo GadX and GadW activate the Gad system indirectly by activating the $\text{gad}E$ transcription (Figure 4A), because overexpression of GadX cannot induce $\text{gad}A$ or $\text{gad}BC$ in a $\text{gad}E$ mutant background [19, 24, 27, 28]. Moreover, overexpression of GadE seems to diminish the requirement for GadX and GadW in $E. coli$ acid resistance but not vice versa [19]. Therefore, the in vitro results do not exactly agree with the in vivo results. So far, the exact mechanism of the GadX/W/E-mediated activation has not been elucidated. It is probable that subsequent to binding of GadX/W to the $\text{gad}A/BC$ promoters, cooperation with GadE is still required for full activation of the Gad system [28]. Moreover, the mechanism of GadX/W/E activation maybe different depending on the environmental conditions under which the Gad system is induced [28].

EvgA/S and YdeO specifically enhance the transcription of $\text{gad}E$ during exponential growth in minimal medium with glucose (Figure 4A). EvgS is a membrane-bound sensor kinase, while EvgA is the response regulator in the two component system and YdeO is an AraC-like transcription regulator. EvgA and YdeO have additive effects on the activation of $\text{gad}E$ transcription. EvgA can also indirectly activate $\text{gad}E$ by activating YdeO through phosphorylation [24]. Another two component system called TorS/R has also been found to repress $\text{gad}E$ transcription (Figure 4A) [29]. TorS/R induces genes involved in using alternative electron donor in the absence of oxygen.

The Lon protease is found to constitutively degrade the GadE protein, even under acid induction [30]. Lon acts as a major cellular protein quality control in $E. coli$. Its effect on GadE can rapidly terminate the acid stress response when pH goes back to neutral, and it also prevents over-expression of acid resistance genes in stationary phase cells [30]. Timely termination of the
Gad system is important for restoring the intracellular pH with the help of the CIC proton-chloride $1\text{H}^+/2\text{Cl}^-$ antiporter (see below).

TrmE (MnmE) is a GTPase involved in tRNA modification. Its GTPase domain is critical for regulating $\text{gadE}$ transcription and also the transcription and translation of $\text{gadA}$ and $\text{gadBC}$ in stationary phase cells grown in rich medium with glucose [27]. $\sigma^{70}$ is thought to initiate $\text{gadA}$ and $\text{gadBC}$ expression in minimal media in an $\text{hns}\Delta\text{rpoS}\Delta$ mutant [31].

SspA, Topoisomerase I, and Dps are also involved in the regulation of the Gad system (Figure 4A). SspA, or stringent starvation protein A, can reduce the H-NS levels post-transcriptionally in stationary phase cells. Because H-NS inhibits the transcription of $\text{rpoS}$, SspA is required to activate the transcription of $\text{rpoS}$ and, hence, genes that are further regulated downstream such as $\text{gadX}$ [32]. A Topoisomerase I deletion strain ($\Delta\text{topA}$) has reduced levels of GadA/B and is acid sensitive. Because Topoisomerase I relieves hypernegative supercoiling in transcription elongation, it is thought that it is required to counteract H-NS repression of $\text{gadA/B}$ genes [33]. Dps stands for DNA-binding protein in starved cells. It maintains the integrity of the bacterial chromosome under low pH by binding to DNA directly, thereby participating in acid stress protection [34].

The expression of the Gad system is also influenced by the concentration of metal ions, such as $\text{Mg}^{2+}$ and $\text{Na}^+$ (Figure 4A). A decrease in external $\text{Mg}^{2+}$ levels is sensed by the $\text{Mg}^{2+}$ membrane sensor PhoQ of the PhoP/Q two-component system. PhoQ then phosphorylates and activates PhoP, which promotes transcription of $\text{gadW}$. Deletion of $\text{phoP}$ results in an acid sensitive $\text{E. coli}$ strain [35]. $\text{Na}^+$ ions modulate the activity of GadX and GadW post-transcriptionally. GadX activates the transcription of $\text{gadE}$ only when the concentration of $\text{Na}^+$ ions is high (e.g. 100 mM) [36].

Indole is secreted by $\text{E. coli}$ as a stationary-phase signal in enriched media and can reach 600 $\mu$M. It can serve as an extracellular signal that controls multidrug exporters and to enhance multicopy plasmid maintenance, cell division, and biofilm formation. Extracellular indole has been shown to activate the glutamine decarboxylase pathway. At pH 3.5, the expression of the GadA/B/C and GadE was found to be activated by approximately 10 fold in the presence of 1 mM indole [37].
Finally, there are three small non-coding RNAs that regulate the Gad system: DsrA, GadY, and GcvB (Figure 4A). DsrA stabilizes the rpoS mRNA to enhance its translation and increases hns mRNA turnover to block translation of H-NS. It does so by sequence-specific RNA-RNA interactions. Deletion of dsrA compromises the E. coli acid resistance ability [38]. The GadY small RNA is encoded in an intergenic region between gadX and gadW (Figure 4), and its transcription is dependent on σ^s [39]. GadY transcripts exist in three different forms: a full-length 105 bp transcript, and two processed transcripts with 90 and 59 nucleotides, respectively. GadY indirectly induces gadE expression by base-pairing with the 3'UTR of the gadX mRNA to increase its stability and translation [39]. The expression of GadY is reduced when the rne gene, which codes for RNaseE, is disrupted. RNaseE is thought to be involved in the processing of a putative unstable GadY precursor into the mature forms of the GadY transcripts. This mutant also has decreased levels of gadA and gadB expression and lower survival rate at pH 2.5. Thus, this essential endoribonuclease is also needed for the proper function of the Gad system [40]. The third small RNA, GcvB, positively regulates acid resistance by increasing the expression level of rpoS. The mechanism of how GcvB does so is unknown, as GcvB does not directly affect rpoS transcription, nor does it interact with H-NS, GadW, or CRP [41].

Based on the above discussion, it can be seen that the Gad system (AR2) is under the control of an extraordinarily complex regulatory network that integrates multiple environmental cues and physiological adaptations (Figure 4A).
Figure 4: Schematic depiction of the gene organization and regulation of the different decarboxylase systems.

A) Regulation of the glutamate decarboxylase system. Arrows indicate activation, while T-shape lines indicate repression. Proteins highlighted in orange indicate that they affect the gene transcription of the connected protein, while those highlighted in blue indicate other mechanisms of regulation such as through protein-protein, protein-DNA, or RNA-RNA interaction. Further details are given in the text.

B) Gene organization of the arginine decarboxylase system. C) Gene organization of the lysine decarboxylase system.
1.1.1.2 AR3: the Arginine Decarboxylase System

Like AR2, this system also enables \textit{E. coli} to survive in extreme acid environment (e.g. pH 2.5). The AR3 pathway consists of the arginine decarboxylase AdiA and an antiporter AdiC [5]. By a mechanism similar to AR2, the AR3 pathway utilizes arginine to increase intracellular pH (Figure 1). AdiA converts arginine into agmatine (Agm) and CO\textsubscript{2}. The agmatine is then transported out of the cell by AdiC in exchange for more arginine [42]. Both AR2 and AR3 pathways are able to increase the intracellular pH of the bacteria to approximately 4.5 in the presence of the corresponding amino acid [5].

The structures of both AdiA and AdiC have been solved by X-ray crystallography [15, 16] (Figure 3). The AdiA monomer has five domains which are, from N- to C-terminus: the wing domain, the linker domain, the PLP-binding domain, the aspartate aminotransferase (AspAT)-like small domain, and the C-terminal domain (Figure 3) [15]. Above pH 6.5, AdiA exists as an inactive dimer. When pH drops below pH 6.0, five dimers assemble to become an 800 kDa active AdiA decamer. Interaction between the N-terminal wing domains in neighboring AdiA dimers is responsible for this association. The wing domain contains many acidic residues exposed at the dimer surface. At neutral pH, the wing domains carry a net negative charge and repulse each other, thus, preventing dimer association. At acidic pH, the surface charges are partially neutralized to facilitate the formation of decamers. Thus, the AdiA decamer is not only the active form of AdiA, but also serves as a proton buffer as pH decreases [15].

The active site of AdiA is buried at the dimer interface. The cofactor PLP is bound in a deep cleft formed by the C-terminal domain of one monomer and the PLP-binding domain of the other monomer. PLP is stabilized by multiple interactions in the PLP-binding domain: the PLP phosphate group is stabilized by the dipole of the N-terminus of the \(\alpha\)12 helix. Lys386 covalently binds to the C4 atom of the PLP pyridine ring which is stacked between the side chains of His255 and Ala349 [15, 16]. When the substrate arginine is not bound, the active site is open and PLP is visible from a cleft formed at the active site. The rim of the cleft at the C-terminal domain is lined with negatively charged residues, including three Glu residues that serve to interact with the positively charged substrate [15]. The active site then closes upon binding of substrate to bring a Glu from one monomer close to an Asp residue from the other monomer. These two residues then coordinate the guanidinium group of the arginine substrate. Thus, it has been
proposed that the AdiA structure explains the maximal activity of AdiA at pH 5.2. At this pH, both the Glu and Asp residues can retain their negative charge for substrate coordination, and the acidic surface of the AdiA dimer can be optimally neutralized to assemble into AdiA decamers [15].

The X-ray structures of the arginine/agmatine antiporter AdiC from wildtype and pathogenic O157:H7 *E. coli* strains have been solved (Figure 3) [16, 43, 44]. AdiC exists as a homodimer in the cytoplasmic membrane and each monomer has 12 transmembrane segments (TM). The AdiC structures in these two *E. coli* strains are identical [43]. The homodimeric interface is mainly formed by interactions between hydrophobic amino acids from TM11 of one monomer and nonpolar residues from TM12 of the other. At the default state, AdiC opens to the periplasm, exposing its central cavity where the active site lies (Figure 3). It is lined with Tyr, Asn, Ser, and Glu residues from multiple TMs that bind positively charged Arg or Agm. Binding of Arg and Agm generally involves different tyrosine residues, but Y93 is required for both.

Recently, the crystal structures of two substrate-bound forms of AdiC have been solved: one is an N22A *E. coli* AdiC mutant in an arginine-bound occluded conformation [44], the other is the N101A *E. coli* AdiC mutant in an arginine-bound periplasm-facing open conformation [45]. Based on these structures, the transport mechanism by AdiC can be proposed as follows: Arg from the extracellular milieu first binds at the central cavity of AdiC in a pocket surround by five transmembrane helices, TM1, 3, 6, 8 and 10. The W293 from TM8 and N101 from TM3 interact with Arg, which then induces conformational changes in the binding site to lead to the Arg-occluded conformation [45]. In the occluded state, residues Ile23 in TM1, Trp202 and Ile205 in TM6, Trp293 in TM8, Ser357 in TM10 and Asn101, Ala96, Cys97 in TM3 constitute the binding pocket. The aliphatic portion of Arg interacts with the side chains of three hydrophobic amino acids, Met104 in TM3, Trp 202 and Ile205 on TM6. The binding induces major conformational changes in TM2, 6 and 10, with structural shift in TM6 being the most pronounced [44, 46].

AdiC then undergoes a conformational change to become open to the cytoplasm, where Arg is displaced by Agm. This series of conformational changes has been proposed to be mediated by polar amino acid clusters that lay proximal, middle and distal with respect to the periplasm [44]. AdiC then closes over to occlude Agm from the cytoplasm until AdiC switches back to the
default conformation to release Agm to the periplasm. The Glu208 residue found deep in the substrate binding pocket is proposed to be the pH sensor by molecular dynamics simulations [47]. With a pKa of 4.25, Glu208 is predominantly protonated at low pH, such as in the stomach (pH 2). At this pH, it can bind the head group of its Arg substrate, which would have no net charge at pH 2 with the positively charged α-amino group offsetting the negatively charged α-carboxyl group. Once facing the cytoplasm (pH 4-5), Glu208 is deprotonated and can bind the positively charged head group of Agm. The Agm is then released by protonation of Glu208 in the periplasm [47].

The regulation of the AR3 pathway is not as extensively studied as AR2. So far, only a CysB protein is shown to act as an activator of the adiA/C genes in anaerobic conditions in complex medium at low pH [5, 48]. The AraC-like regulator AdiY, located between adiA and adiC in the genome (Figure 4B), can also enhance the expression of the two genes when overexpressed. However, AdiY is not essential for the function of the Adi system in E. coli since the transcription of the AdiA/C genes is not affected by mutations in AdiY [5].

1.1.1.3 AR4: the Lysine Decarboxylase System

In contrast to AR2 and AR3, the AR4 pathway provides protection for E. coli at moderate acidic conditions (pH 5), anaerobic environment and phosphate starvation [49, 50]. The AR4 system consists of the lysine decarboxylase LdcI (CadA) and an inner-membrane lysine-cadaverine antiporter, CadB [51]. The two genes are organized into a cadBA operon (Figure 4C) [52]. LdcI functions optimally at pH 5.7. It converts lysine into cadaverine and CO2 (Figure 1) [51].

The cadBA operon is induced by low pH, lysine, and anaerobiosis [53]. Kinetic studies showed that induction of cadBA operon is slower when cells are exposed to lysine and low pH simultaneously than when they are previously adapted to either [53]. H-NS is a repressor of the cadBA operon under normal conditions, possibly by affecting DNA topology [54]. The expression of cadBA is also repressed by the accumulation of cadaverine, which binds to the C-terminal periplasmic domain of CadC [55, 56]. Docking experiments suggest that a cavity at the interface of the two subdomains is where cadaverine is bound by the side chains of Glu447, Asp225, Thr229, Tyr374 and Gln421[57]. Recently, it has been found that in the absence of extracellular lysine, endogenous lysine can also induce low level expression of the cadBA system [58].
OmpC (36 kDa) and OmpF (35 kDa) are required in the arginine and lysine decarboxylase pathway in anaerobic conditions [59]. They are classical outer membrane proteins (OMP) that function as porins to allow small hydrophilic molecules to pass through under harsh conditions in gram-negative bacteria for cell survival. Endogenous cadaverine produced from LdcI activity has been found to inhibit the function of OmpC and OmpF without affecting their expression [60]. Double deletion of OmpC and OmpF in *E. coli* significantly decreases the expression of the *adiA* and *cadB* genes. The deletion mutant has an acid-sensitive phenotype that can be restored by supplementing glutamine, but not arginine or lysine [59, 61].

### 1.1.1.3.1 LdcI (CadA)

LdcI is a Fold Type I family member of PLP-dependent enzyme [2]. The domain organization of LdcI monomer is the same as that of AdiA (Figure 2, Figure 3 and Figure 5A) [2, 13, 62]. It consists of an N-terminal Wing domain (residues 1-129), a core domain (residues 130-563) and a C-terminal domain (CTD) (residues 564-715). The core domain is made up of a linker region (residues 130-183) and two subdomains: a PLP binding subdomain (PLP-SD) (residues 184-417) and subdomain 4 (SD4) (residues 418-563) (Figure 5A). In its active form, LdcI exists as decamers composed of five dimers (Figure 5B). The Wing and SD4 domains of each monomer mediate the interaction between neighbouring dimers to form the LdcI decamer [2]. The dimerization of LdcI monomers completes the active site, which is buried and can be accessed through a narrow cleft formed at the dimer interface. In the active site, the PLP cofactor is covalently bound to the ε-amino group of K367 via a Schiff base to form an internal aldimine [2].

Both *in vitro* and *in vivo*, LdcI has been shown to be inhibited by ppGpp, the small signaling molecule in stringent response (see Section 1.2). The binding pocket for ppGpp is located at the interface between neighboring monomers within each ring of the LdcI decamer. There are ten ppGpp binding sites in total, five high affinity sites (Kₐ ~13 nM) and five low affinity sites (Kₐ ~685 nM). The ppGpp binding pocket is consisted of basic residues from neighbouring monomers that interact with the negatively charged phosphate groups of ppGpp [2]. The interaction directly links the stringent response to the acid stress response in *E. coli*.

LdcI is also found to interact with RavA, a MoxR-family chaperone-like AAA+ ATPase (see Section 1.3) [3]. By electron microscopy, it was found that five RavA hexamers bridge two LdcI
decamers to form a cage-like structure [63]. Interaction of RavA with LdcI enhances RavA ATPase activity but not LdcI activity [63]. However, this interaction reduces the inhibition of LdcI by ppGpp both in vitro and in vivo [3].

1.1.1.3.2 CadB

CadB is a 445 amino acid antiporter that exchanges cadaverine for lysine. The $K_m$ value of cadaverine for antiporter activity of CadB is 300 $\mu$M [51]. Although the crystal structure of CadB has not been solved so far, its structure has been modeled based on the structure of AdiC (Figure 5C). It contains 12 transmembrane helices with both its N- and C-termini in the cytoplasm. The cavity of CadB consists mainly of TM2 2, 3,6,7,8 and 10[64]. By mutagenesis studies, residues Trp43, Tyr58, Tyr107, Tyr366 and Tyr368 have been identified to involve in cadaverine uptake at neutral pH. Arg299 is involved in cadaverine-lysine antiport activities. Two tyrosine residues in TM10 of CadB are involved in the uptake activities [64]. Other important residues involved in the uptake or excretion of cadaverine include Glu (76, 204, 377, and 408), Asp (185 and 303), Tyr (57, 73, 89, 90, 174, 368, 235, 246, 310, and 423), and Trp 41. Cys370 is important for both uptake and excretion of cadaverine [65].

The proposed mechanism for CadB mediated lysine transport is based on the structure of AdiC. First, lysine binds to the central cavity and it is exchanged with cadaverine. The bound cadaverine then moves into the periplasm through interaction with aromatic residues Tyr89 and Tyr90 on TM3 and Tyr235 on TM7.
Figure 5: Structures of the components of the lysine decarboxylase system.

A) The crystal structure of LdcI (CadA) dimer is shown with domains colored for one monomer [PDB ID: 3Ql6, [2]]. B) LdcI exists as a decamer in its active form [2]. The X-ray crystal structure of LdcI decamer (red) is fitted to the EM image of LdcI (gray). C) The schematic structure of CadB. It contains 12 transmembrane helix and both its N- and C-termini are in the cytoplasm [66]. D) The crystal structure of the periplasmic domain of CadC is shown with the N- and C-termini indicated [3LY7,[67]]. Cys208 and Cys272 residues that form an intramolecular bond are shown as sticks. The structure of CadC was prepared using PyMOL (DeLano 2002).

Note: Figure 5 A) and B) is reprinted ©2011 European Molecular Biology Organization [2].

Figure 5 C) is reprinted with kind permission from Springer Science and Business Media.
1.1.1.3.3 CadC

CadC is encoded upstream of the cadBA operon and acts as its transcriptional activator. CadC is a 58 kDa integral membrane protein consisted of three domains: the cytoplasmic DNA binding domain (residues 1-159), a transmembrane helix domain (residues 160-187) and a C-terminal periplasmic domain (residues 188-512) [56]. The crystal structure of the periplasmic domain of CadC has been solved at 2.3 Å (Figure 5D) [68]. It is composed of two subdomains: one with a mixed parallel/anti-parallel β sheet packed against two α-helices from a three-helix bundle and the other with an 11-α-helix bundle [57].

The cadBA operon is activated at low external pH (pH 5) and the presence of lysine. The presence of lysine is sensed directly by LysP, a lysine permease that has 12 transmembrane segments [56]. In the absence of lysine, LysP inhibits CadC via an interaction between the transmembrane segments of LysP and a cluster of six aromatic amino acids, in particular Phe165, in the transmembrane domain of CadC [56]. In addition, Arg265 is the only residue in the periplasmic domain that is involved in the lysine-dependent activity of CadC [55]. When lysine is abundantly available, LysP activity is repressed and CadC is released to activate the transcription of cadBA via its N-terminal cytoplasmic domain [53, 56].

By both computational modeling and size exclusion chromatography experiments, CadC has been proposed to form dimers in order to activate the cadBA operon as pH decreases [57]. Moreover, two CadC binding sites have been identified in the cadBA promoter region. Changes in pH are sensed by a negative patch, which includes Asp198, Asp200, Glu461 and Asp471, in the periplasmic domain of CadC. Mutations of these residues result in either a pH insensitive or pH-nonresponsive phenotype. These residues are also located at the dimer interface. As pH decreases, the acidic residues become more protonated, thus reducing the repulsion between the two CadC monomers to promote the formation of dimers [69].

Additionally, an intramolecular disulfide bond formed between Cys208 and Cys272 in the periplasmic domain has been found to be important for the conformational change in CadC in response to change in pH (Figure 5D)[70]. Cys208 is located within a flexible loop in the N-terminal subdomain of the C-terminal periplasmic domain. When either cysteine is mutated, cadBA expression is partially induced at pH 7.6, indicating that CadC is at a semi-active state [70]. The exact mechanism for the formation and opening of the disulfide bond remains
unknown. However, it was speculated that oxidation of the two cysteines occurs *de novo* as CadC rapidly turn over in the presence of oxygen at pH 7.6. At low pH (pH 5.8), either the disulfide bond formation is prevented or is broken by an unknown mechanism [70].

Building on the experimental results described, a mechanism for the induction of the *cadBA* operon by CadC has been proposed [70]. First, the disulfide bond is formed at pH 7.6 and reduced at pH 5.8. At neutral pH, CadC is inactive due to both the inhibition by LysP and the formation of the disulfide bond. When lysine is present at neutral pH, even though LysP inhibition is relieved, CadC is still inactive due to the presence of the disulfide bond. At pH 5.8 and in the absence of lysine, the disulfide bond is reduced in CadC which is still inactive due to the inhibition by LysP. Therefore, only at low pH and in the presence of lysine, CadC becomes activated via dimerization-induced conformational changes [70].

### 1.1.1.4 AR5: the Ornithine Decarboxylase System

The ornithine decarboxylase system is poorly studied compared to the other AR system. It is composed of the inducible ornithine decarboxylase SpeF, and the ornithine/putrescine antiporter PotE (Figure 1) [71, 72]. They are organized on the same operon and are induced when cells are grown at pH 5.0.

SpeF is a PLP-dependent decarboxylase that catalyzes the decarboxylation of L-ornithine to putrescine. Its structure is paralogous to LdcI, AdiA and SpeC, the constitutive ornithine decarboxylase. SpeF has three domains: an N-terminal Wing domain, a core domain with PLP-binding site and a C-terminal domain (Figure 2). It oligomerizes to form a homodimer and has a pH optimum of 6.9 [4, 73].

PotE is a 439-amino-acid transmembrane protein that is structurally related to both AdiA and CadB [4]. It contains 12 transmembrane helices and exchanges putrescine for more ornithine under acid stress [4].

### 1.1.2 The Role of Chloride and Potassium Ions in Acid Resistance

Upon encountering acidic conditions, it is important for *E. coli* to regulate its transmembrane potential which is disrupted in extreme acid (e.g. at pH 2) [5]. The strategy that *E. coli* has
developed is to reverse the electrical membrane potential from approximately -90 mV to +30 mV (when glutamate is present) or +80 mV (when arginine is present) [5, 74]. This flip in polarity of the membrane potential is thought to be due to the combined effect of an increasing amount of protons in the cytoplasm as a result of the external pH decrease and due to the accumulation of decarboxylation products of the AR2 or AR3 pathways (Glu to GABA, Arg$^+$ to Agm$^{2+}$, respectively). When the acid stress is relieved, the membrane potential is restored with the help of the E. coli CIC chloride channel [5, 12] (Figure 1).

The CIC channel is a $1H^+/2Cl^-$ antiporter, and its structure has been solved by X-ray crystallography (Figure 3) [17]. It is a homodimeric membrane protein in which each monomer can transport one Cl$^-$ ion. The Cl$^-$ selectivity filter is formed by Cl$^-$ ion interacting with the dipoles of an $\alpha$-helix and with nitrogen and hydroxyl groups [17]. The CIC protein imports Cl$^-$ to bring negative charges into the cell and simultaneously removes excess protons, which facilitates the recovery of a negative inner membrane potential [5].

Potassium ions are also implicated in the regulation of cytoplasmic pH in E. coli. The K$^+$ influx systems including the Trk and Ktr K$^+/H^+$ symporters and K$^+$ channels play important roles in maintaining the electrochemical membrane potential and in regulating the cytoplasmic pH homeostasis [75]. The requirement of K$^+$ ions depends on the external pH. White et al. showed that cells depleted of the K$^+$ transport systems can survive as well as wild type at pH 8.0, but fail to survive at an external pH of 6.0 when the concentration of K$^+$ is below 40 mM [76].

### 1.1.3 Chaperone-based Acid Stress Response

Proteins involved in the above three enzyme-based pathways function mainly in the cytoplasm. However, E. coli also has acid resistance systems to protect periplasmic proteins, which are more vulnerable to acid denaturation and damage as the outer membrane porins are permeable to extracellular molecules. The small chaperones HdeA (9.7 kDa) and HdeB (9 kDa) are the primary players in the acid stress response in the periplasm (Figure 1) [7]. The $hdeA$ and $hdeB$ genes are expressed from the same operon (Figure 5). Although they only share 17% sequence identity, HdeA is structurally homologous to HdeB [7].

HdeA and HdeB are general chaperones that function in an ATP-independent manner. They recognize a variety of substrates and maintain them in a soluble state in acidic environment.
They can also form mixed aggregates with proteins that have failed to be solubilized (Figure 1) [77]. The presence of HdeA and HdeB was shown to decrease the size of protein aggregates in extreme acid (below pH 3); the chaperones also make these aggregates less hydrophobic. In vitro data shows that HdeA functions optimally at below pH 3, while HdeB functions optimally at pH 3 [78]. However, it seems that in vivo, both are required for optimal protection of periplasmic proteins at external pH 2 and 3. They were shown in vitro to be able to help proteins refold once pH becomes neutral, it is possible that, in vivo, other enzymes such as periplasmic chaperones, disulfide isomerases, and/or peptidyl prolyl isomerase are involved in the folding process [77].

The structure of HdeA has been solved at pH 4 by X-ray crystallography (Figure 3) [7]. HdeA is a compact single-domain protein with a hydrophobic core created by four α-helices. There is one disulfide bond formed between Cys18 and Cys66 that stabilizes the structure of the monomer. At neutral pH, HdeA exists in an inactive homodimeric form. The dimer is held together by both electrostatic interaction and hydrophobic interactions. Residues in the loop connecting helix B and C, Asp43, Glu46, Asp47 and Asp51 from one monomer electrostatically interact with lysine clusters in the other HdeA monomer [79]. At pH 2~3, these residues become protonated, and thereby destabilize the dimerization [79]. In addition, the dimerization is mediated by the interaction between hydrophobic residues, such as Val, Thr, and Ala, from both monomers. When the pH is abruptly shifted to pH below 3, HdeA dissociates into active monomers in a fraction of a second [80]. The monomer turns into a partially unfolded state that retains most of its secondary structure but not its tertiary structure [80]. The Cys18-Cys66 disulfide bond is essential for HdeA function by holding the protein together at this stage. The hydrophobic residues previously at the dimer interface are consequently exposed and are shown to adaptively interact with misfolded/unfolded substrates to form different conformations of substrate-HdeA complexes [80] The primary sequence of HdeA reveals an amphipathic property of this protein, in that the N- and C-termini of this protein are both basic and the middle region is hydrophobic [81]. The N- and C-termini may help increase the solubility of the HdeA-substrate complexes at extreme acidic pH [81].

Two HdeA substrates, DegP and SurA, have been identified [82, 83]. These two proteins function as chaperones in the periplasm as part of an essential quality control system for the biogenesis of the outer membrane proteins. HdeA likely protects them at low pH and releases them as pH increases. They may then refold to aid in HdeA-mediated acid recovery of other
client proteins in an ATP-independent manner [82].

Similar to HdeA, HdeB is a dimer at neutral pH and dissociates into monomers at pH 3.5 [78]. The crystal structure of HdeB solved at pH 4.5 shows that the HdeB and HdeA are very similar in their monomeric, but not dimeric, form (Figure 2 and Figure 3) [18]. The HdeB monomer has a very long loop that connects the two large helices (residues 64-72). In the HdeB dimer, the dimer interface contains a pH-sensitive interunit salt-bridge formed between Glu41 and Lys48. In addition, there is a cluster of aromatic residues formed by Trp55, Trp56 and Tyr64 from each monomer. This local structural environment of this cluster is maintained by a salt bridge between Asp76 and His59, which is disrupted upon decrease in pH [18].

The mechanism of function of HdeB is proposed to be similar to that of HdeA. For HdeB, the dissociation of the dimer is triggered by breaking the Glu41-Lys48 salt bridge [18]. The exposed surface of HdeB when it becomes disordered at pH 3 is less hydrophobic than that of HdeA [78]. The functions of HdeA and HdeB at above pH 3 and below pH 7 have not been investigated [77].

The hdeA/B genes are induced in stationary phase cells or at acidic pH. Regulators of hdeA/B include GadE, GadX, GadW, H-NS, LRP, and MarA. GadE activates the transcription of the hdeA/B genes chaperones under all conditions. GadX and GadW are both transcriptional repressors of hdeA/B. H-NS represses hdeA/B transcription at neutral pH [77]. The leucine-response protein (LRP) is a strong repressor for hdeA/B transcription in minimal media. MarA, an AraC/XylS transcriptional regulator, represses hdeA/B transcription by increasing H-NS repression and interfering with GadE activation in stationary phase in the presence of sodium salicylate [84].

Recently, a cytoplasmic chaperone Hsp31 (hchA) was implicated in acid resistance in E. coli (Figure 1) [8, 85]. Hsp31 is a heat-inducible homodimeric protein (Figure 2 and Figure 3). The Hsp31 monomer consists of an α/β/α domain and a small α/β domain. Its dimerization is mediated mostly by Van der Waal forces between polar residues including Tyr82, Gln121, Ser124 and Asn128 on each monomer of the dimer [8]. Hsp31 functions as a holdase that stabilizes unfolded protein intermediates until stress is elevated. The transcription of hchA is induced by σ8 in stationary phase cells. Deletion of hchA results in markedly decreased activity of the AR2 and AR3 pathways. However, the underlying mechanism is not known [85].
1.1.4 The Acid Fitness Island in *E. coli*

Many of the regulatory genes for the Gad system (Figure 4A), including *gadA/E/W/X*, as well as the chaperone/chaperone-related genes *hdeA/B/D* are located as a cluster at position 3,651,984 – 3,665,603 bp on the chromosome (Figure 6) [22]. HdeD is an acid-resistance membrane protein with unknown function. This cluster is unique to *E. coli* and is termed the ‘acid fitness island’ by Hommais et al.[22].

Recently, more genes in this island were found to participate in acid resistance, including an outer membrane lipoprotein Slp, a transcription regulator YhiF, and the predicated membrane proteins YhiD and HdeD. Slp and YhiF are required to protect cells against excreted toxic metabolites including the accumulated anions of dissociated weak acids after growth at low pH such as lactate, succinate, and formate. HdeD and YhiD are required for acid stress response when cells are grown at high density (> 10⁸ CFU/mL) [86]. Moreover, transcription of these genes is activated by GadX and GadW, but it is unknown whether the activation is direct [87].
Figure 6: The acid fitness island in *E. coli*.

The acid fitness island is located at 3651984 – 3665603 bp on the *E. coli* genome. The *yhiUV* genes are also termed *mdtEF*. The figure is drawn to scale.
1.1.5 Acid Stress Response in Other Enterobacteria

Although *E. coli* is the most commonly used model organism for studying acid resistance, it is important to also study other pathogenic bacteria in order to understand the differences in their acid resistance systems. Four other commonly studied enteropathogenic bacteria are *Vibrio cholera*, *Vibrio vulnificus*, *Shigella flexneri*, and *Salmonella enterica serovar typhimurium* (Table 1).

The two *Vibrio* species are intestinal pathogens. *Vibrio cholera* has an acid tolerance response (ATR) and an essential component of the ATR is the lysine decarboxylase system [88]. The *V. cholera* CadA and CadB function similarly to the *E. coli* Cad system. The *cadA* and *cadB* genes are organized into an operon and upstream of the *cadBA* operon lies the *cadC* gene. The *V. cholera* CadC is similar in sequence and function to the *E. coli* CadC. It specifically activates the transcription of *cadBA* operon under acid induction (pH 4.5). However, unlike *E. coli*, the *V. cholera* cadA also possesses an independent promoter, so the *cadB* and *cadA* genes can also be transcribed monocistronically. The *cadBA* operon is transcribed constitutively at a low level independent of acid or CadC induction, although this basal level expression is insufficient for acid resistance [88]. In addition, a ClpB protein encoded by the *clpB1* gene in *V. cholera* was also implicated in the acid stress response [89]. ClpB protein is a member of the Hsp100-family chaperone ATPases that function to dissolve protein aggregates.

The *Vibrio vulnificus* species which causes food-borne gastroenteritis also encodes a Cad system [90]. The *V. vulnificus* CadB is 82% identical to the *E. coli* CadB, and was recently found to transport cadaverine into the cell at pH 8.5 [91]. As cadaverine has been found to directly scavenge superoxide radicals, the import of cadaverine in the basic condition of the anterior intestine likely protects cells against oxidative stress [91, 92]. The expression of the *cadBA* operon in *Vibrio vulnificus* is regulated by multiple factors. Because acid stress often induces superoxide stress, SoxR, a protein induced under superoxide stress, together with CadC induces *cadBA* at low pH. The two activators bind at different regions on the same *cadBA* promoter: SoxR at -10--38 bp and CadC at approximately -233 bp from the transcription start of *cadBA* [90]. AphB, a LysR family transcriptional regulator, indirectly induces the expression of *cadBA* by activating the transcription of *cadC*. The AphB homologue in *V. cholera*, however, functions in a virulence cascade instead of with CadC [93]. Lastly, LRP, the leucine responsive protein,
cooperates with CadC to bind to the cadBA promoter and induce the expression of cadBA [94]. So far, acid resistance pathways that permit survival in extreme acid (pH 2), such as the Gad and Arg system, have not been found in the Vibrio species. Thus, the two Vibrio species are more acid sensitive than E. coli and Shigella flexneri.

*S. flexneri* has the AR1 (the decarboxylase-independent pathway), AR2 and AR3 pathways [95]. However, there are several notable differences in AR2 of *S. flexneri* from that of *E. coli*. Expression of the *Shigella* Gad system requires acid induction and σ8 in stationary phase cells grown in minimal medium, while the *E. coli* Gad system is induced in stationary phase regardless of medium [95]. In stationary phase *Shigella* cells, the expression levels of gadA, gadBC, gadE, and hdeA genes decrease with increase in pH; gadE is not even transcribed at above pH 6.9. However, in *E. coli*, gadE transcripts can still be detected at pH 7.5 [95]. The *Shigella* Gad system is not found to be induced in exponential phase cells [96]. The σ8-independent induction of the Gad system of *S. flexneri* requires a much stronger environmental signal, such as anaerobiosis and growth on glucose, while in *E. coli* the induction of the Gad system can happen under semi-aerobic conditions. Furthermore, HdeA is essential for the proper function of the Gad system in *S. flexneri* [95].

In both *S. flexneri* and enteroinvasive *E. coli* (EIEC), the cad locus is inactivated [97]. Complementation of EIEC and *S. flexneri* strains with CadC has been found to negatively regulate the expression of the arginine decarboxylase system. This suggests that loss of functional lysine decarboxylase is counterbalanced by a higher expression of the *adi* system [97].

*Salmonella enterica serovar typhimurium* was previously thought not to be able to survive in extreme acid (below pH 3) because it lacks AR2 and AR3 systems. While the AR2 pathway is shown to be absent in *Salmonella* by Southern blotting against Salmonella DNA with gadC probe, recently, a functional arginine decarboxylase (AR3) system was discovered in this species [98-100]. The *Salmonella* AR3 system is only induced when cells are grown in anaerobic conditions (i.e. in the absence of oxygen), but, unlike in *E. coli*, the adiY gene in *Salmonella* is required for the proper activation of this system [99, 100].

*Salmonella* also has the AR4 lysine decarboxylase system [101]. The cadBA operon is activated by CadC. However, unlike in *E. coli* where cadC is constitutively expressed, the *S. enterica cadC* is induced by low pH and the presence of lysine. This could be explained by the
observation that the promoters of the cadC gene in the two species have little sequence similarity [101].

The Salmonella CadC has three domains, similar to that of E. coli: the N-terminal DNA-binding domain, the transmembrane domain, and the C-terminal periplasmic domain [102]. At low pH and in the presence of lysine, CadC in Salmonella is cleaved in the vicinity of residue 210 located in a segment linking the transmembrane and periplasmic domains [102]. This cleavage is proposed to then lead to the activation of the cadBA operon [102]. The CadC proteins of E. coli and Salmonella share 58.4% sequence identity, and they are predicted to have similar structures [101]. However, it is not yet clear whether E. coli CadC undergoes a similar processing event to activate cadBA as Salmonella CadC.

Salmonella also has two other major acid tolerance response (ATR) systems [103]. One of them is a log-phase ATR system. This system is induced when exponentially growing cells adapted at a moderate pH (4.5 – 5.8) undergo a transition to low pH (pH 3) [104]. Over 60 acid shock proteins (ASPs) are produced during this response. These proteins include σ5, Fur (the major iron regulator), Ada-DNA methyltransferase involved in DNA damage and repair, and the two-component PhoP/Q system. Protons prevent Mg²⁺-dependent inhibition of PhoQ by affecting its Mg²⁺ binding site. PhoQ can then phosphorylate PhoP to induce genes required for Salmonella to survive in macrophage phagolysosome [104]. Importantly, this log-phase ATR also provides cross protection against other environmental challenges such as oxidative stress, heat shock, and high osmolarity, but not vice versa.

The other major ATR system in Salmonella is the stationary-phase ATR, which is induced by exposing stationary-phase cells to below pH 5 [103]. Forty eight ASPs are induced in this ATR and only five of them overlap with those induced in the log phase induced acid tolerance [104]. This system is σ5-independent and is not affected by mutation in genes such as fur and phoP. One example of genes induced by this response is OmpR [105]. Two known genes induced by OmpR are the OmpC and OmpF porins. Upon acid stress, OmpR is activated by phosphorylation from the phosphate donor acetyl phosphate. Although OmpR is required for optimal function of the stationary-phase ATR, how it is induced and what OmpR-dependent genes are involved in acid tolerance are not known [105]. The Salmonella CadC is a mild repressor of ompR transcription. In fact, unlike in E. coli and Vibrio, the Salmonella CadC regulates many genes apart from the
cadBA operon. Proteins downregulated by cadC include proteins involved in glycolysis (PfkA, PfkB, FbaB, and STM4519), energy production (AtpD), and stress response (Tig and HslU). Proteins up-regulated by CadC include outer membrane proteins such as OmpC and OmpF. However, it is not clear whether the CadC regulation is direct or indirect [101].

The hdeA/B genes are absent from the Salmonella genome, as confirmed by Southern blotting [98]. Other chaperones that might be involved in the acid stress response of Salmonella have not yet been identified.
Table 1: Comparison of decarboxylase systems in different enteric bacteria.

<table>
<thead>
<tr>
<th></th>
<th>Glutamate decarboxylase system</th>
<th>Arginine decarboxylase system</th>
<th>Lysine decarboxylase system</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gadA/B gadC gadE</td>
<td>adiA adiC adiY</td>
<td>cadA cadB cadC</td>
<td></td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>+     +    +</td>
<td>+    +    +</td>
<td>+    +    +</td>
<td>[5]</td>
</tr>
<tr>
<td>Vibrio cholera</td>
<td>-     -    -</td>
<td>-    -    -</td>
<td>+    +    +</td>
<td>[88]</td>
</tr>
<tr>
<td>Vibrio vulnificus</td>
<td>-     -    -</td>
<td>-    -    -</td>
<td>+    +    +</td>
<td>[90]</td>
</tr>
<tr>
<td>Shigella flexneri</td>
<td>+     +    +</td>
<td>+    +    +</td>
<td>-    -    -</td>
<td>[95, 97]</td>
</tr>
<tr>
<td>Salmonella typhimurium</td>
<td>-    -    -</td>
<td>+    +    +</td>
<td>+    +    +</td>
<td>[98, 99, 100]</td>
</tr>
</tbody>
</table>

“+”: gene present; “-”: gene not found or encodes a nonfunctional protein
1.2 The Stringent Response and ppGpp

The stringent response in bacteria is a collective term for the growth arrest and adaptive changes taking place in cellular metabolism and physiology to maintain survival under stress conditions, in particular, nutritional starvation [106-110]. The cellular signals resulting from starvation are sensed by RSH (Rel/Spo Homologue) proteins which then synthesize ppGpp (nucleotide guanosine 5'-diphosphate 3'-diphosphate) from ATP and GTP. ppGpp is the signaling molecule that then triggers a plethora of downstream effects that amounts to the stringent response. The RSH proteins are conserved among many Gram-positive and Gram-negative bacteria species, hence so are ppGpp and the stringent response.

1.2.1 Synthesis of ppGpp by RSH Proteins

The domain arrangement of the RSH family protein is schematically depicted (Figure 7A). The protein can be divided into two domains, NTD (N-terminal domain) and CTD (C-terminal domain) [109]. The NTD encodes a ppGpp hydrolase, a linker region, and a ppGpp synthetase. The RSH hydrolase subdomain encodes a Mn$^{2+}$-dependent pyrophosphohydrolase with a conserved His-Asp (HD) motif [109]. The CTD mostly plays a regulatory role in balancing the hydrolysis and synthesis activity of RSH proteins to avoid a futile cycle of ppGpp production.

In *Escherichia coli*, two RSH proteins have been identified to synthesize ppGpp, RelA and SpoT [111, 112]. RelA is an 84 kDa monofunctional protein that can only synthesizes ppGpp [112]. SpoT is a 79 kDa bifunctional protein that can both synthesize and hydrolyze ppGpp [111].
Figure 7: Domain arrangements and structure of RSH family proteins.

A) The domain arrangements of RelA and SpoT in *E. coli*, Relmtb (*Mycobacterium tuberculosis*), and Relseq (*Streptococcus equisimilis*) are drawn to scale. Domain boundaries are defined based on descriptions in [113-119]. ACT stands for aspartate kinase-chorismate mutase-TyrA. TGS stands for ThrRS, GTPase and SpoT.

B) The crystal structure of the N-terminal domain (NTD) of Relseq [PDB ID: 1VJ7] is shown and the domains are coloured as in (A). NTD can exist in two state, a hydrolase OFF/synthetase ON state, and a hydrolase ON/synthetase OFF state [113]. A Mn$^{2+}$ ion (gray sphere) binds to the hydrolase domain. The disordered loops are traced as dashed line and highlighted with a red arrow. This figure was prepared using *PyMOL* (DeLano 2002).
There are several structural differences between RelA and other bi-functional RSHs (Figure 7A). First, the HD motif required to coordinate Mn$^{2+}$ for hydrolysis is mutated in RelA [113]. Therefore, RelA does not have hydrolase activity. Secondly, the RelA synthase domain contains an acidic triad of residues (ExDD) instead of the conserved basic RxKD triad conserved in bi-functional RSH proteins [114]. The RxKD is important in bifunctional RSHs as the ExDD mutation results in the drastic reduction of ppGpp synthesis [115]. In addition, RelA seems to exist as dimers in its inactive form [120]. The dimerization is mediated by residues 455-538 and 550-682 and inhibits the production of ppGpp under amino acid starvation [120]. The CTD of RelA contains a TGS (ThrRS, GTPase and SpoT) domain and an ACT (aspartate kinase-chorismate mutase-TyrA) domain, a proposed ligand-binding domain that primarily binds amino acids such as serine [116, 117, 121].

RelA synthesizes ppGpp under amino acid starvation by sensing uncharged tRNA in the A-site of the 70S ribosome (Figure 8) [122, 123]. The association of RelA with the ribosome is mediated by RelA CTD (residues 455-744), presence of mRNA and the N-terminus of L11 subunit of the 50S ribosome [120, 123]. RelA interacts with mRNA, but there is so far no evidence for its direct binding to L11. It has been suggested that RelA hops among the ribosomes to stimulate the synthesis of ppGpp. This proposal is based on the observation of the increased rate of ppGpp synthesis with an increase in the concentration of ribosomes and correlation of ppGpp synthesis with the release of RelA, rather than release of deacylated tRNA [123]. Finally, overexpression of RelA has been found to complement a G791 substitution in 16S ribosome that abolishes its protein synthesis function [124]. Thus, RelA seems to function also in stimulating ribosome function independent of its role in ppGpp synthesis [124].

In contrast, SpoT synthesizes ppGpp by sensing stresses such as fatty acids shortage, iron and/or carbon starvation, and phosphate starvation (Figure 8) [110, 125-127]. The mechanism by which it does so is studied in more details in fatty acid starvation. ACP (acyl carrier protein) is an important cofactor in lipid metabolism [126, 128]. To be functional, ACP must be modified post-translationally by a 4'-phosphopantethein prosthetic group (4'PP) to become holo-ACP [128]. The functional ACP then carries the fatty acid intermediates through the successive enzymatic reactions of the fatty acid synthesis elongation cycle. Under fatty acid starvation, changes in the ratio of unacylated ACP to acylated ACP leads to the binding of functional ACP to the TGS domain of SpoT [128]. This interaction shifts the balance of the two SpoT activities towards
ppGpp synthesis [109]. The specificity of this interaction is also determined by the NTD of SpoT because the chimeric fusion of RelA NTD and SpoT TGS does not bind ACP [123, 128]. When ppGpp is made, it inhibits the expression of the *fabHDG* operon that is responsible for fatty acid synthesis [129]. Phosphate starvation is sensed by SpoT as well. The presence of ppGpp then induces the expression of genes in the *pho* regulon involved in the metabolism and transport of inorganic phosphate Pi and Pi compounds [125, 130].

*E. coli* strains deleted of both RelA and SpoT are called ppGpp\(^0\) strains [111]. They show various growth defects, most noticeably being polyauxotrophic for different amino acids [111, 131]. The ppGpp\(^0\) phenotypic suppressors are called M\(^+\) mutants. The mutations map exclusively to RNAP rpoB/C and D subunit genes (discussed in Section 1.2.2) [111, 131].

Studies of two other RSH proteins, Relseq (*Streptococcus equisimilis*) and Relmtb (*Mycobacterium tuberculosis*), also gave structural insights into the regulatory role of RSH CTD in balancing the hydrolase and synthetase activities of these proteins (Figure 7A). It has been shown that a deletion in the CTD inhibits both hydrolase and synthase activities in Relseq, while the same deletion inhibits only the synthetase activity in Relmtb [118, 119]. Furthermore, full length Relseq which synthesizes ppGpp in *Streptococcus equisimilis*, degrades all ppGpp in *E. coli*. Chimeric proteins of Relseq NTD to RelA CTD restore the ppGpp synthetase activity in *E. coli* [118]. Although the underlying mechanism for the species-specific function remains unknown, it is clear that the CTD serves an important regulatory function.

The crystal structure has been solved for Relseq NTD (residues1~385) (Figure 7B) [113]. The hydrolase domain (5-159) is predominantly helical, with a short \(\beta\)-hairpin between two \(\alpha\) helices to form part of the active site. The synthetase domain (176-371) consists of a five-stranded \(\beta\)-sheet surrounded by five \(\alpha\) helices in a sandwich-like arrangement. The synthetase and hydrolase active sites are more than 30 Å apart. The crystal structures revealed two states of the protein, a hydrolases OFF/synthetase ON state and a hydrolase ON/synthetase OFF state (Figure 7B). The hydrolase and the synthetase domains are joined by an overlapping 3-helix bundle denoted as C3HB [113]. When the \(\text{Mn}^{2+}\) ion is present, a group of residues including the Hist80Asp81 doublet in the NTD and C3HB coordinate the \(\text{Mn}^{2+}\) ion to switch to the hydrolases OFF/synthetase ON state. On the other hand, when ppGpp binds to C3HB, it induces a conformational change which then traps the ppGpp in the hydrolase domain and thus switches to the hydrolase ON/synthetase
OFF state [113]. The NTD-CTD boundary for both SpoT and Relseq is a solvent accessible region that might act as a hinge [118]. Therefore, the conformational changes between the two states most likely result from a CTD-mediated intramolecular signal transmission [113]. Lastly, similar to the oligomerization of RelA, Relmtb forms trimers [119]. The trimerization is mediated by the NTD, which can form trimers on its own. However, it is regulated by the CTD because NTD-only trimers are easily dissociable. The physiological role of the Relmtb trimer remains unknown [119].

1.2.2 Cellular Adaptations Mediated by ppGpp in the Stringent Response

Once ppGpp is synthesized, it then triggers the stringent response [132]. It exerts its effects via diverse mechanisms in a vast number of cellular processes, including transcription, translation, DNA replication, and phospholipids synthesis (Figure 8).
Figure 8: Schematic diagram of the stringent response in *E. coli*.

This diagram shows that RelA and SpoT sense different environmental stresses to induce the production of ppGpp. Examples of proteins with which ppGpp directly interacts are shown. The major cellular processes that are either inhibited or activated by ppGpp in the stringent response are also included. This figure is adapted from [108].
The most well-studied downstream effect of ppGpp is on its modulation of *E. coli* RNAP (RNA polymerase). Although it is well-known that ppGpp directly binds to the $\beta$ and $\beta'$ subunits of RNAP, the details of this interaction initially examined in the co-crystal structure of *Thermus thermophilus* RNAP with ppGpp has been questioned [133, 134]. Vrentas et al. argued that the preparation of RNAP for binding study in Artsimovich et al. was undersaturated with the $\omega$ subunit and it was known that RNAP lacking the $\omega$ subunit cannot be inhibited by ppGpp *in vitro* [133, 134]. Mutagenesis studies performed by Vrentas et al. showed that substitutions at the ppGpp binding site identified in the *E. coli* RNAP crystal structure did not reduce transcription inhibition by ppGpp either in the presence or absence of DksA (discussed below) [133]. It also did not alter the effect of ppGpp on promoter complex life time [133]. Therefore, to date the ppGpp binding site on RNAP still remains undetermined.

There are several models that attempt to explain the modulation of RNAP by ppGpp. Two of the main models are the kinetic model and the sigma factor competition model [108, 109]. For the kinetic model, ppGpp binding either decreases RNAP open complex stability, traps the RNAP in a closed complex, or affects binding to a promoter which leads to a more readily occurring dissociation of the stable open complex from the promoter. ppGpp may also increase pausing during transcriptional elongation. For the sigma factor competition model, the binding of ppGpp releases RNAP from the $\sigma^{70}$-RNAP complex from the promoters. Consequently, alternative sigma factors including $\sigma^e$, $\sigma^{32}$ and $\sigma^{54}$ that are outcompeted by $\sigma^{70}$ for binding to RNAP under normal conditions can now bind to RNAP. The transcription of genes under the control of these sigma factors is then activated [108, 135].

An important co-factor for ppGpp is a protein called DksA (DnaK suppressor A), as the overexpression of DksA complements *dnaK* mutation to restore thermotolerance to bacteria [109]. DksA is a 17 kDa protein that cooperates with ppGpp to affect RNAP activity *in vivo*. Structurally, it is similar to GreA and GreB, both of which are well-characterized transcriptional elongation factors [109]. DksA acts on RNAP by inserting its N-terminal coiled-coil finger domain through the RNAP secondary channel [136]. This insertion leads to long-range structural changes within RNAP that alter interaction with the -6 to +6 region at $\sigma^{70}$ promoters and thus enhances ppGpp’s effect on gene transcription [109, 137]. In ppGpp$^0$ strains DksA overproduction can completely compensate for positive and negative regulation with respect to amino acid auxotrophy, cell-cell aggregation, motility, filamentation, stationary phase.
morphology and RpoS accumulation [109]. So far, there is no global pattern for the type of genes that requires both DksA and ppGpp. The exact mechanism for how the pair functions together still remains unknown. However, it has been known that DksA is absolutely required for the direct inhibition of ribosome transcriptions and direct activation of amino acid promoters (both are discussed below) [138-141].

The most noticeable manifestation of ppGpp-mediated effect on transcription is the inhibition of ribosome production [142]. The binding of ppGpp to RNAP alters the initiation pathway that traps RNAP, reduces the ability of RNAP to form an open complex on rRNA promoters rnpP1 and P2 and decreases the stability of the open complex [138, 139]. The inhibitory effect requires DksA [140]. As a result, the production of ribosomes is halted during the stringent response [142]. This effect is the most noticeable mark of the stringent response. In fact, the stringent response is classically defined as a rapid downregulation of rRNA biosynthesis and ribosome production during amino acid starvation.

In addition, ppGpp, with DksA, regulates the level of σ^s [143]. The σ^s factor is important for general stress resistance in *E. coli* in stationary phase because it is required to induce the expression of genes involved in protection from low pH, oxidative stress, high temperature and osmotic shock [143]. The *rpoS* gene (encodes σ^s) is transcribed during exponential growth. However, the cellular concentration of σ^s is kept low because an adaptor protein RssB directs σ^s to ClpXP for degradation. When ppGpp is made, it activates the transcription of anti-adaptor proteins IraP and IraD, which bind and inhibit RssB [144-146].

ppGpp also activates the transcription of the sigma factor σ^E [146]. σ^E (RpoE) coordinates the cellular response to the presence of misfolded proteins in the periplasm or outer membrane. ppGpp enhances σ^E activity by increasing the level of the *rpoE* mRNA. In addition, σ^E-mediated transcription initiation of certain genes is stimulated directly by ppGpp and DksA [147, 148].

There is *in vitro* evidence that certain amino acid biosynthesis promoters (P_{argI}, P_{thrABC}, P_{livF} and P_{hisG}) are activated by combining ppGpp and purified DksA [109, 139, 141, 149]. ppGpp also regulates the half-life of mRNAs by directly inhibiting the transcription of *pcnB* [150]. PcnB encodes a polyA-polymerase that adds poly-A tails to mRNA and targets them for degradation. Thus, ppGpp preserves mRNA in *E. coli* under nutrient starvation.
Promoters of the genes inhibited by ppGpp share common features. There is a GC-rich discriminator sequence between -5 to +2 which forms DNA supercoils [151]. Cells with mutations in this region are no longer susceptible to stringent control [152]. The activity of this discriminator is also affected by the -35 and -10 sequences and the length of the linker between them [151]. Promoters repressed by ppGpp usually have a 16 bp linker in contrast with the 17 bp consensus [153].

In addition to its effect in transcription, ppGpp affects translation. As discussed below, it inhibits protein synthesis by promoting the formation of an inactive 100S ribosome, interfering with activity of the elongation factors EF-Tu and EF-G, and modulating of the function of the initiation factor IF2 [132, 148, 150, 154].

During stationary phase, a ribosome modulation factor (RMF) is expressed [155]. It specifically associates with the 100S ribosome, which is formed by dimerization of 70S ribosome and has no translation activity. The expression of RMF is induced by ppGpp, and not by σ8. However, although mutants deleted with the rmf gene are not able to form 100S ribosome in stationary phase, induced ppGpp expression in log-phase results in a transcription of RMF but not the 100S ribosome [155]. This result suggests that there are other factors affecting 100S formation [155]. ppGpp also interacts with EF-Tu and EF-G. The EFTu-ppGpp complex seems to increase the fidelity of proofreading in protein biosynthesis by slowing aminoacyl-tRNA incorporation [148]. ppGpp binds to the same site as GTP on initiation factor 2 (IF2) with a Kd of approximately 2 μM, which is five times stronger than that of ppGpp binding to EF-G (Kd 10 μM) [154]. It was proposed that, as ppGpp level increases and GTP level reduces, ppGpp partially competes out GTP for binding IF2, which reduces the efficiency of fMet-tRNAi to the 30S subunits. The inhibition in translation possibly feeds back to affect the transcriptional machinery [154, 156].

ppGpp also affects DNA replication. In E. coli, in vitro evidence has shown ppGpp inhibits the initiation of DNA replication by reducing the synthesis of primers by DNA primase DnaG, which forms a complex with the DnaB helicase [157]. In contrast, ppGpp inhibits DnaG in Bacillus subtilis by causing temporary stalling of DNA replication upon nutrient starvation in both leading and lagging strands [158]. DNA replication resumes with the addition of amino acids [158]. It was proposed that the rapid and reversible replication arrest serves to maintain
genomic stability by preventing deleterious consequences associated with replication in starving cells [158].

Other cellular processes that are affected by ppGpp in *E. coli* include phospholipid synthesis, purine nucleotide biosynthesis, biofilm formation and accumulation of polyP (hundreds of inorganic phosphate) polymers. ppGpp interacts and inhibits PlsB (glycerol-3-phosphate acyltransferase), which is required for phospholipid synthesis [159]. ppGpp regulates purine nucleotide biosynthesis by inhibiting key enzymes such as IMP (inosine monophosphate dehydrogenase) and adenylosuccinate synthetase, an essential enzyme to synthesize AMP [132, 160, 161].

Recently, ppGpp was found to regulate biofilm formation in *E. coli* with a second messenger Cyclic-di-GMP which is implicated in biofilm control [162]. The biofilm is secreted by *E. coli* cells to protect the bacteria from physical or chemical stress and is mostly composed of exopolysaccharides [163]. The *pgaABCD* operon responsible for biofilm formation is repressed by high level of ppGpp. Addition of low doses of translation inhibitors such as chloramphenicol or tetracycline prompts SpoT to degrade ppGpp, leading to the derepression of *pgaA*, which encodes an outer-membrane protein. However, for optimal expression of the *pgaABCD* operon, a delicate balance between ppGpp and c-di-GMP is required [163].

(pp)ppGpp induces accumulation of polyP polymers by inhibiting the phosphatase PPX which hydrolyzes polyP [164]. The level of PolyP increases to 20 mM in stationary phase cells. PolyP may promote ribosomal protein degradation by the Lon protease [165]. Mutants that lack PPK, the polyphosphate kinase PPK that synthesizes polyP fail to adapt to stress and do not survive in stationary phase [164].

Recently, a high-throughput transcription profiling experiment showed that the global downstream effects of ppGpp signalling can be grouped into two loops [131]. One is an independent feed-back loop controlled by LRP at low ppGpp level, and the other is a feed-forward loop regulated mainly through RpoS (σ^s) at high ppGpp level. LRP works with ppGpp to induce genes involved in amino acid metabolism, replenishing depleted amino acids under mild short-term starvation conditions. RpoS induces the expression of genes that prepare the cell for survival under prolonged starvation conditions. The two loops together balance biosynthetic and survival processes as cells experience starvation [166].
The signaling effects of ppGpp in the stringent response have also been examined in a variety of bacteria, including *B. subtilis*, *Pseudomonas aeruginosa*, *S. typhimurium*, *Streptomyces coelicolor*, and *Francisella. tularensis*. In *B. subtilis*, ppGpp has been found to inhibit GTP biosynthesis [167], ribosome transcription [168] and transcription repressors for genes required for stress response [169]. It also activates GTPases and the synthesis of second messenger molecules that induce the expression of proteins involved in stress resistance such as antibiotic synthesis, flagellar production and sporulation [109, 167, 170]. In addition, the effects of ppGpp often contribute to the virulence of pathogenic bacteria [106]. ppGpp either inhibits the repressors, or promotes the binding of transcription activators of genes that encode crucial virulence factors involved in processes such as biofilm formation and bacteria transmission [106, 171].

In conclusion, it can be seen that ppGpp critically regulates the adaptation of bacteria cells to different stresses. The inhibition of LdcI by ppGpp thus extends ppGpp’s regulatory network to acid stress.

### 1.3 MoxR Family AAA+ ATPase

#### 1.3.1 Introduction to AAA+ ATPases

The AAA+ ATPase (ATPase associated with diverse cellular activities) proteins are a subfamily of the P-Loop NTPase superfamily which accounts for approximately 10~18% of all cellular proteins in prokaryotes, eukaryotes and viruses [172]. The P-Loop NTPases carry out diverse functions in the cell mainly by inducing conformational changes in other molecules using energy from ATP or GTP hydrolysis. P-loop NTPases have two conserved sequence motifs, Walker A (Gx4GK[S/T] where x is any residue) which binds to the β and γ phosphate of the bound NTP, and Walker B (hhhh[D/E]) where h is hydrophobic residue), which binds to the Mg$^{2+}$ ion [172-175].

The structure of the AAA+ domain is well studied. The classical AAA+ domain is represented by the AAA+ ATPase replication factor C (RFC) that functions as a clamp loader in DNA replication in eukaryote (Figure 9) [176, 177]. It consists of an N-terminal subdomain and a C-terminal α-helical subdomain [173, 175, 177]. The N-terminal subdomain is of an α/β/α Rossman fold. The parallel β sheet is made up of five stands arranged in the sequence of β2-β3-β4-β1-β5
Structural diversity in the N-terminal subdomain can come either from the number and position of the five or more α-helices that connect this central β-sheet or from insertions between strands (Figure 9). The C-terminal subdomain is an α-helical bundle domain. The AAA+ domain connects to other domains through a linker region N-terminal to the first helix of the AAA+ domain. This linker region has a conserved motif (either Gly-Gly or h-Gly, where h is any hydrophobic residue) that contributes to the adenine-ring-binding pocket [175].

Both the conserved Walker-A and -B motifs, which are responsible for the binding and hydrolysis of ATP, are in the N-terminal α/β/α subdomain. The Walker A motif directly interacts with the phosphates of ATP, most importantly via its Lys residue. The Walker-B motif hydrolyzes ATP. The Asp residue coordinates the Mg$^{2+}$ ion required for ATP hydrolysis, and the Glu residue activates water for the hydrolysis reaction. After the C-terminal end of the Walker-B motif, there is a second region of homology (SRH) which contains two important elements for AAA+ function: the Sensor 1 motif and arginine finger at the N- and C- terminus, respectively. The Sensor 1 element is present in all AAA+ proteins. It is located in the loop connecting β4 to α4, which physically places it between the Walker A and B motifs. Sensor 1 interacts with Walker B and with the γ-phosphate of the bound ATP. In AAA+ oligomers, the arginine finger from one subunit serves to complete the ATP binding site of an adjacent subunit. In the C-terminal α helical bundle subdomain, there is a Sensor 2 motif that also participates in nucleotide binding. A conserved Arg residue in this motif interacts directly with the γ-phosphate of ATP.

In their biologically active form, AAA+ ATPases assemble into oligomers, mostly hexamers. As a consequence, the ATP-binding sites coordinated by the structural elements described above are exposed at the interface between subunits. Binding and hydrolysis of ATP then leads to conformational changes that are transmitted to either substrate proteins or to the interacting partner proteins. The oligomers often also have a central cavity or pore that has a role in the function of several AAA+ chaperones [175].

The AAA+ ATPases are found in all branches of life and serve essential roles in diverse cellular processes, such as DNA replication [177], protein degradation and disaggregation [178-180], cell apoptosis [181], homologous recombination [182] and biosynthesis of chlorophyll [183, 184]. Their diversity in function and structure has made them a very important family to study.
Figure 9: The classical AAA+ ATPases domain structure.

The classical AAA+ domain structure is represented by the AAA+ ATPase domain of RFC (replication factor C) [PDB ID: 1SXJ, [177]]. The αβα domain is colored in purple, and the α-helical domain is colored in green. The N- and C-termini are labeled. This figure was prepared using PyMOL (DeLano 2002).
1.3.2 Introduction of RavA and ViaA

1.3.2.1 RavA

RavA (Regulatory ATPase Variant A) belongs to the RavA subfamily of the MoxR AAA+ family. The MoxR family of AAA+ ATPases has been poorly studied. Members of this family have been found in multiple bacteria and archaea species such as in *Oligotropha carboxidovorans*, *Rhizobium leguminosarum*, and *Francisella tularensis*, where they are required for the oxidation of carbon monoxide (CO) as a carbon source, maintaining cell envelop integrity and stress resistance, respectively [185-187].

The *rava* gene is located upstream of the *viaA* gene in the same operon (Figure 10A). The two genes have a 7 bp overlap at the 3' end of *rava* (Figure 10A). RavA is a 56 kDa protein that exists as hexamer in its active form. RavA consists of three domains: the N-terminal AAA+ ATPase domain (residue 1-306), a discontinuous triple helical domain (residue 307-330, 442-497), and a novel LARA domain (LdcI associated domain of RavA) (residues 331-437) (Figure 10B). The AAA+ domain is made up of the αβα subdomain (residues 1-192) and an α-helical subdomain (residue 226-306). They are linked by a 32-residue helical segment (193-225). The RavA hexamer has a different AAA+ domain arrangement from other hexamers such as HslU. In the AAA+ domain of the HslU hexamer, the all-α subdomain is located on top (left) of the αβα subdomain so each subunit has a complete AAA+ on its own. In contrast, the all-α subdomain in RavA which contains the sensor 2 motif required for ATP hydrolysis is located on the right of the αβα subdomain. As a result, the complete ATP hydrolysis site in RavA hexamer is composed of the αβα subdomain of one subunit and the α-helical domain from its neighbouring subunit. Overall, the RavA hexamer shows a clockwise orientation, while the HslU hexamer is in a counterclockwise orientation (Figure 10 C) [3].

The triple helical domain is stabilized by hydrophobic interactions localized at the interface between the three helices. The LARA domain is named so because this domain mediates the interaction of RavA with LdcI, the inducible lysine decarboxylase. The LARA domain has a compact anti-parallel β-barrel-like structure consisting of six β-strands and one α-helix. The domain contains many basic residues on the surface including Arg340, Arg347, Arg348, Arg398, Lys400, Lys409 and Arg423 (Figure 10D) [3]. RavA and LdcI have been shown to form a large cage-like structure (3.3 MDa) as a result of five RavA hexamers bridging two LdcI decamers [63]. This interaction reduces the inhibition of LdcI by ppGpp both *in vitro* and *in vivo* [3].
exact mechanism of this modulation is currently under investigation.

1.3.2.2 ViaA

ViaA contains a VWA (Von Willebrand Factor Type A) domain (Figure 10D) [63]. Proteins containing this domain are found in both prokaryotes and eukaryotes, but their functions are much better studied in eukaryotes [188]. The VWA domain mediates protein-protein interaction by undergoing conformational changes after binding divalent metal cations. Extracellular proteins that contain VWA domains often are receptors or integrins that involve in extracellular signaling and cell adhesion. Intracellular proteins that contain a VWA domain are found to function mostly in multiprotein complexes in transcription, DNA repair, ribosomal and membrane transport. Proteins with both AAA+ and VWA domains play essential roles in protoporphyrin IX biosynthesis in chloroplasts [188].

As a representative, the VWA domain structure (residues 224-433) of the complement protein C2a is shown (Figure 10E) [189, 190]. Complement proteins are a system of over 30 plasma proteins that sensitize pathogens for phagocytosis or for direct lysis. This domain has a central β sheet with one antiparallel edge strand, amphipathic α helices that lie against each β sheet face. The VWA domain contains a conserved metal-ion-dependent adhesion site (MIDAS) motif (DXSXS, T and D, where X represents any amino acid residue) at the carboxyl-terminus of the β sheet [191, 192]. The MIDAS motif binds divalent ions and is required for the function of VWA-domain proteins [190-192]. In C2a, it consists of Asp240, Ser242, Ser244, Thr317 and Asp356 and binds to a Mn^{2+} ion (Figure 10E) [190]. In ViaA, the MIDAS motif is constituted by D327, S329, S331, T393 and D421. Proteins containing the VWA domain in prokaryotes are not well studied.
Figure 10: Structure and domain arrangements of RavA and ViaA.

A) The overlapping region between the 3' end of ravA and the 5' end of viaA in the ravA-viaA operon [63]. B) The crystal structure of RavA monomer with the domains labelled [PDB ID: 3NBX, [3]]. C) Comparison between RavA and HslU hexamer shows that RavA ATPase domain exhibits a clockwise orientation and that of HslU is in a counter-clockwise orientation [3]. D) Domain arrangements of RavA and ViaA [63]. E) The crystal structure of the VWA domain of the complement protein C2a [2I6Q, [189]]. A Mn²⁺ ion (gray sphere) is coordinated by the MIDAS motif consisted of Asp240, Ser242, Ser244, Thr317 and Asp356. The VWA domain structure was prepared using PyMOL (DeLano 2002).

Note: Figure 10 A) is reprinted from [63] © the American Society for Biochemistry and Molecular Biology.
1.4 Thesis Rationale

Work presented in this thesis follows the \textit{in vitro} study of the inhibition of LdcI by ppGpp in our lab. The exciting discovery was the result of the finding of ppGpp in the LdcI decamer X-ray structure. Not only does this interaction provide the first direct link between the stringent response and acid resistance, but this is one of the examples where ppGpp regulates protein activity via direct binding during the stringent response. In addition, RavA belongs to the poorly characterized MoxR family of AAA+ ATPase protein. The interaction of RavA and LdcI is an important clue in deciphering its function. In this thesis, I first studied the details of the interplay between ppGpp, LdcI and RavA \textit{in vivo}. Secondly, triggered by the observation that GadB in the glutamine decarboxylase pathway migrates to the membrane upon acid stress, I used fluorescence imaging to study the localization of the components of the lysine decarboxylase pathway, RavA and ViaA under different stress conditions.
2 Antagonistic Regulation of LdCI by ppGpp and RavA

Reprinted (adapted) with permission from:


©2011 European Molecular Biology Organization


Note: I performed experiments presented in Figure 13, Figure 14 and Figure 15.
2.1 Introduction

Acid stress response in *Escherichia coli* enables the bacteria to survive in acidic environment such as in the stomach (pH 2.5). It is consisted of both decarboxylase-based and chaperone-based systems. The lysine decarboxylase pathway (AR4), is one of the four major decarboxylase-based systems [5]. It provides protection for *E. coli* at moderate acidic conditions (pH 5). It is composed of LdcI (CadA), the lysine decarboxylase, and CadB, the lysine-cadaverine antiporter. The expression of the *cadBA* operon is induced by its transcription activator CadC in the presence of lysine and at low pH (pH 5).

The X-ray crystal structure of LdcI has been solved (Figure 5A) [2]. In its active form, LdcI exists as decamer composed of five LdcI dimers. LdcI is a member of the Fold-Type I family of PLP-dependent enzymes, members of which include AdiA and SpeF. Each LdcI monomer can be divided into three domains: an N-terminal wing domain (residues 1-129), a core domain (residues 130-563) and a C-terminal domain (CTD) (residues 564-715). The core domain is made up of a linker region (130-183) and two subdomains: a PLP binding subdomain (PLP-SD) (184-417) and subdomain 4 (SD4) (418-563) (Figure 11A). The N-terminal wing domain interacts with the SD4 domain of the neighbouring monomer to contribute to the decamer formation. The active site is located in the core domain. Dimerization completes and buries the PLP-bound active site, which is accessible from a narrow cleft formed by the dimer interface. The lower part of the cleft is composed of residues from the SD4 and CT domains of one monomer, and the upper part of the cleft is composed of residues from the PLP-SD of the other monomer. Entry to the active site is lined with negative charged residues which can interact with positively charged lysine substrates.

In the active site, the PLP cofactor is covalently bound to the ε-amino group of K367 via a Schiff base to form the internal aldimine. The interaction is mediated by hydrogen bonding between polar residues such as T220, S221, S364 and H366 with the phospholate group in PLP. In addition, there is also stacking interaction between H245 and A332 to the pyridoxal ring of PLP.
Figure 11: Inhibition of LdcI by ppGpp in vitro.

A) The top ring of an LdcI decamer is shown as a cartoon representation, where each monomer has a different color and the bottom ring is shown in grey. The Wing domain (residues 1-129) is shown in red. The Linker region (residues 130-183) is shown in orange. The PLP-SD subdomain (residues 184-417) is shown in yellow. The SD4 subdomain (residues 418-563) is shown in green. The CTD domain (residues 418-563) is shown in blue. A close-up view of the residues in neighbouring dimers involved in decamerization. The N-terminal wing domain (red) and the SD4 subdomain (green) mediate the intra-ring interaction. L89 which is mutated in the study is boxed.

B) LdcI decamer is shown with binding of ppGpp. Each monomer has a different color and the bottom ring is shown in grey. A close-up view of one of the ppGpp binding sites is shown. The left-hand monomer is colored orange and the right hand monomer is colored grey (residues for this monomer are indicated with a prime symbol). The position of the C5’ and C3’ carbon atoms on the ppGpp ribose ring as well as the phosphate residues are indicated in italics. Positions of R206 and R97 which are mutated in the study are boxed [2].

C) LdcI activity was inhibited by ppGpp in vitro as shown by a TNBS (2, 4, 6-trinitrobenzensulfonic acid) assay [63]. The activity of LdcI (25 nM) was measured from pH 4.5 to 8.5 in the absence or presence of 100 μM of nucleotides (GTP, GDP, ppGpp or pppGpp) under low salt (25 – 135 mM NaCl). Error bars represent the standard deviation for the measurements of three replicates.

D) Wild type and mutant LdcI proteins (1 μM) were separated at pH 6.5 by size-exclusion chromatography using a Superose 6 HR 10/30 column. LdcI R97A and R206S eluted as decamers, similar to WT LdcI. LdcI L89R mutants eluted as dimers.

E) The three LdcI mutants are not inhibited by ppGpp. Lysine decarboxylation activity for WT LdcI and the LdcI mutants (25 nM) was measured using the TNBS-activity assay in the absence and presence of 100 μM ppGpp. Rates are relative to the maximum rate obtained at pH 5.5 (shown in C) and error bars indicate the standard deviation of the average of three replicates. The top panel shows data at pH 5.0 and the bottom panel shows data at pH 6.5.
A surprising and exciting discovery was made with the crystallization of LdcI that the small alarmone, ppGpp, directly binds to the LdcI decamer. ppGpp is an important signalling molecule for the bacterial stringent response. The stringent response describes the cellular response to, in particular, amino acid starvation, which is sensed by RelA to synthesize ppGpp from GTP and ATP. SpoT, another protein that both synthesizes and hydrolyzes ppGpp senses fatty acids, iron or carbon shortages. The downstream effects of ppGpp encompass a global change of transcription pattern where genes involved in normal metabolism are turned off and genes involved in stress resistance are turned on [108].

The binding pocket for ppGpp is located at the interface between neighboring LdcI dimers (Figure 11B). It is composed of residues from SD4 and CTD of one monomer (colored light orange in Figure 11B) and the Wing domain, PLP-SD, and SD4 of the neighboring monomer (colored grey and labeled with a prime symbol). Interaction with ppGpp is mediated by both electrostatic interactions, which are formed between positively charged residues including R565, R585, and R206 with the phosphate groups on ppGpp, and stacking interaction, which are formed between R97’ and L564 with the guanosine base on ppGpp (Figure 11B). In total, there are 10 ppGpp binding sites per LdcI decamer, five of which are high affinity sites (K_d ~13 nM) and the other five of which are lower affinity sites (K_d ~685 nM).

Binding of ppGpp to LdcI has then been shown to strongly inhibit LdcI activity in vitro. The activity of LdcI is low below pH 5.0 and peaks at pH 5.5 (Figure 11C). The LdcI activity was reduced by approximately 10 fold at pH greater than 5.0 in the presence of ppGpp.

The effect of ppGpp on LdcI in vitro was further investigated by using three LdcI mutants: LdcI R97A, LdcI R206S and LdcI L89R. The R97A and R206S LdcI mutants are not inhibited by ppGpp as the mutations are part of the ppGpp binding site. The L89R mutant has the mutation at the interaction site between neighbouring dimers. Therefore, this mutant cannot form decamers and cannot be inhibited by ppGpp.

The LdcI R97A and LdcI R206S proteins eluted as decamers, and the LdcI L89R eluted as dimers at pH 6.5 by size exclusion chromatography (Figure 11D). For all three of them, the LdcI activity was not affected by the presence of ppGpp, as measured by a TNBS (2, 4, 6-trinitrobenzenesulfonic acid) assay. In addition, the activity of the LdcI L89R dimers was 5-fold
lower than that of WT LdcI and the other two mutants (Figure 11E). Overall, these *in vitro* data presented above demonstrated that LdcI is inhibited by ppGpp.

RavA, a MoxR-family chaperone like AAA+ ATPase, has been found to interact with LdcI [3, 63]. RavA is induced in stationary phase by σ^S under aerobic conditions and exists as a hexamer in its active form (Figure 12A). The exact function of RavA is still under investigation. By both EM and pull-down experiments, RavA was found to interact with LdcI to form a cage-like structure (3.3 MDa). The cage is composed of five RavA hexamers bridging two LdcI decamers (Figure 12B).

Docking of the crystal structures of RavA hexamer and LdcI decamer onto the EM cage structure suggested that the LARA domain near the C-terminus of RavA mediates the interaction with LdcI (Figure 10 and Figure 12B). There are two binding constants of RavA to LdcI, 18 nM and 1.22 µM, which likely reflect the two RavA legs in a hexamer that binds to LdcI. This result is also confirmed by SPR (surface Plasmon resonance) experiments, showing that RavAΔLARA (residue 336~433 is deleted) does not bind LdcI [3, 63].

It was shown by ITC (isothermal titration calorimetry) that, while the ATPase activity of RavA is not affected by ppGpp, addition of ppGpp does not affect the activity of LdcI when the RavA-LdcI complex is pre-formed (Figure 12 C and D). In contrast, when RavA is added after pre-incubating LdcI with ppGpp, RavA is not able to reduce the inhibition of LdcI by the alarmone (Figure 12 E). In addition, the activity of LdcI after preincubating with RavA ΔLARA was still strongly inhibited by ppGpp (Figure 12 D). Overall, the data above demonstrated that RavA can antagonize the inhibition of LdcI by ppGpp *in vitro* [3].
Figure 12: RavA antagonizes the inhibition of LdcI by ppGpp in vitro.

(A) RavA hexamer structure is shown [3].

B) Crystal structures of LdcI decamer (gray) with ppGpp (blue) binding and RavA hexamer (black) are docked onto the EM cage structure using the VEDA software. LdcI dimer interacting with RavA is colored in red for the upper monomer and green for the lower monomer.

(C) The ATPase activity of RavA is not affected by ppGpp, as measured by ITC experiments in the presence of different nucleotides. Error bars represent the standard deviation of the average of three replicates.

(D) Pre-formation of RavA-LdcI complex reduces the inhibitory effect of ppGpp. LdcI activity was measured by ITC in the presence of RavA or RavAΔLARA and/or ppGpp. ppGpp was added after RavA to LdcI.

(E) RavA does not affect LdcI inhibition by ppGpp when LdcI has been pre-incubated with ppGpp. The LdcI activity is measured by ITC. RavA was added after ppGpp to LdcI.

The ITC experiments in (C), (D) and (E) were performed using a MicroCal VP-ITC calorimeter at 25°C with a stirring speed of 310 rpm in a buffer consisting of 100 mM sodium MES, pH 6.5, 1 mM TCEP, 1 mM ATP, 0.25 mM MgCl₂, and 0.1 mM PLP. For each experiment, a single injection of 75 µL was made and the initial rates were calculated in ORIGIN 7.0 using ΔH_{apparent} for L-lysine of -3161 cal mol⁻¹ [3].
In order to understand the physiological significance of the inhibition of LdcI by ppGpp, an *in vivo* study on the interaction between RavA, ppGpp and LdcI activity is described in this chapter. First, the MG1655 ΔcadBA and MG1655 ΔrelAΔspoTΔcadBA strains were made by a λ Red recombination system [193]. Plasmids expressing either WT LdcI or the three LdcI mutants (R96A, R206S and L89R) were then transformed into the two knockout strains to generate complementation strains. These complementation strains were then tested in a media shift assay to show that ppGpp inhibits the activity of LdcI and RavA reduces the inhibition *in vivo*.

2.2 Materials and Methods

2.2.1 Generation of the ΔcadBA Strains

2.2.1.1 Preparation of the Kan<sup>R</sup> Knockout Cassette from pKD4

The kanamycin-resistance cassette (Kan<sup>R</sup>) used for gene knockout was amplified from the pKD4 plasmid (Table 2) according to the standard PCR protocol for Pfx polymerase purchased from Invitrogen. The primers used are cadBA-KO-F and cadBA-KO-R (Table 3). All PCR products were purified according to the Invitrogen PCR purification protocol, digested with DpnI (from NEB—New England Biolab) at 37°C for three hours to remove any remaining plasmid templates and re-purified using the Invitrogen PCR purification kit. The product concentration was determined by the A<sub>260</sub>/A<sub>280</sub> ratio measured with a BioRad SmartSpec™ 3000.
<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Strain</th>
<th>Features</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pKD3</td>
<td>BW25141</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, Cam&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pKD4</td>
<td>BW25141</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, Kan&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pCP20</td>
<td>DH5α</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>[194]</td>
</tr>
<tr>
<td>pET3a</td>
<td>DH5α</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, ori&lt;sub&gt;pBR322&lt;/sub&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pET22b</td>
<td>DH5α</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, ori&lt;sub&gt;pBR322&lt;/sub&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pET322</td>
<td>DH5α</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, ori&lt;sub&gt;pBR322&lt;/sub&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>sfGFP-pCOLADuet-1</td>
<td>DH5α</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, ori&lt;sub&gt;pBR322&lt;/sub&gt;</td>
<td>[195]</td>
</tr>
<tr>
<td>mRFP-pRSETB</td>
<td>DH5α</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>[196]</td>
</tr>
<tr>
<td>YFP-pRSETB-CKAR</td>
<td>BL gold</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>[197]</td>
</tr>
<tr>
<td>sfGFP-pKD3</td>
<td>BW25141</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, Cam&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>mRFP-pKD3</td>
<td>BW25141</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, Cam&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>YFP-pKD3</td>
<td>BW25141</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, Cam&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>pST39</td>
<td>DH5α</td>
<td>Amp&lt;sup&gt;R&lt;/sup&gt;, ori&lt;sub&gt;pSC101&lt;/sub&gt;</td>
<td>[198]</td>
</tr>
<tr>
<td>pT7POL26</td>
<td>DH5α</td>
<td>Kan&lt;sup&gt;R&lt;/sup&gt;, ori&lt;sub&gt;pSC101&lt;/sub&gt;</td>
<td>IPTG inducible promoter</td>
</tr>
</tbody>
</table>
Table 3: List of primers used in the thesis work.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cadBA-KO-F</td>
<td>GATCCCTGATGTGTTAATAAAACCTC-AAAGTTCTCAGTTAGTGAGGTGGCTAGCTGCTTC</td>
</tr>
<tr>
<td>cadBA-KO-R</td>
<td>ACCTTTAAGGGTATAGCGGCAGCTACG-GCTGACGGTATGCACCATATGATATCTCCCTCTAG</td>
</tr>
<tr>
<td>k2rc-R</td>
<td>GCAGTTCATTTCAAGGGAACGAGAG</td>
</tr>
<tr>
<td>kt-new-F</td>
<td>GGAATCATGGCATTGGCCG</td>
</tr>
<tr>
<td>cadC-ver-F</td>
<td>TCAACCGCTTTTAATTATTCCG</td>
</tr>
<tr>
<td>yjdL-ver-R</td>
<td>GTAACGTTAAAAACAGAAGAAGCAT</td>
</tr>
<tr>
<td>pcadBA-BamHI-F-2</td>
<td>TACTATGGATCCCTTGTACCTATCTCGAC-AAATTTCTT</td>
</tr>
<tr>
<td>pcadBA-NsiI-R-2</td>
<td>GAGACTATGCAATTATTTTTTGTCTTT-CTTCTTTCAATAACC</td>
</tr>
<tr>
<td>pKD3-mRFP-SphI-F</td>
<td>ACTTTGACCTAGCAGCTAGGCTAAAGGAAGAGAG</td>
</tr>
<tr>
<td>pKD3-mRFP-AflII-R</td>
<td>ATGGTTCAACGTAGTTGCTTGAAGAGACGTTGAG</td>
</tr>
<tr>
<td>pKD3-sfGFP-SphI-F</td>
<td>ACTTTGACCTAGCAGCTAGGCTAAAGGAAGAGAG</td>
</tr>
<tr>
<td>pKD3-sfGFP-SphI-R</td>
<td>ATGGTTCAACGTAGTTGCTTGAAGAGACGTTGAG</td>
</tr>
<tr>
<td>pKD3-YFP-SphI-F</td>
<td>ACTTTGACCTAGCAGCTAGGCTAAAGGAAGAGAG</td>
</tr>
<tr>
<td>pKD3-YFP-AflII-R</td>
<td>ATGGTTCAACGTAGTTGCTTGAAGAGACGTTGAG</td>
</tr>
<tr>
<td>LdcI-sfGFP-F</td>
<td>CCGCTATACCGTTAAGGTTATGAAGAAAGGAGGCAAAAAATGGGTAAAAGGAGAGAG</td>
</tr>
<tr>
<td>LdcI-sfGFP-R</td>
<td>GTGTTCCACTGTTAAGGTTATGAAGAAAGGAGGCAAAAAATGGGTAAAAGGAGAG</td>
</tr>
<tr>
<td>LdcI-6Gly-sfGFP-F</td>
<td>CCGCTATACCGTTAAGGTTATGAAGAAAGGAGGCAAAAAATGGGTAAAAGGAGAG</td>
</tr>
<tr>
<td>CadB-mRFP-F</td>
<td>CGGCGAGGACACTCAATGGATAACCACACACCGCCTAACCACACTCATGAGGAGAGGAGAGGAGAGAG</td>
</tr>
<tr>
<td>CadB-mRFP-R</td>
<td>CCAACACATGGGCAAAAAATGGTAAGAGGAGGAGGAGGCTCGAGAAAAATGGGTAAAAGGAGAGAG</td>
</tr>
<tr>
<td>CadB-6Gly-mRFP-F</td>
<td>CGGCGAGGACACTCAATGGATAACCACACACCGCCTAACCACACTCATGAGGAGAGGAGAGGAGAGAG</td>
</tr>
<tr>
<td>CadC-sfGFP-F</td>
<td>TTATGTGTTACCTTATCTCGACAAAAATTTTTTTTCTTACGAAATGGGTAAAAGGAGAGAGAG</td>
</tr>
<tr>
<td>CadC-sfGFP-R</td>
<td>GAACATACAAAAACTCAACAAACAATAATTTTCCTCGAGCATAAATCCATATGATATATCTCCCTCTTAA</td>
</tr>
<tr>
<td>CadC-6Gly-sfGFP-F</td>
<td>TTATGTGTTACCTTATCTCGACAAAAATTTTTTTTCTTACGAAATGGGTAAAAGGAGAGGAGAGGAGAGAG</td>
</tr>
<tr>
<td>RavA-mRFP-F</td>
<td>CCAGCTTACAGGAGTGGGTGCCAGAGTTCCCGAGGAGCACAAGATATGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGA</td>
</tr>
<tr>
<td>RavA-mRFP-R</td>
<td>CCAGCTTACAGGAGTGGGTGCCAGAGTTCCCGAGGAGCACAAGATATGAGGAGGAGAGAGAGAGAGAGAGAGAGAGA</td>
</tr>
<tr>
<td>Sequence Name</td>
<td>Nucleotide Sequence</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>RavA-6Gly-mRFP-F</td>
<td>CCAGCCTACAGGATGTCGGTCACAGATTCGCCAGGCAACAAATGCCGG GAGAGCGGGGGGTGGGAATGGCCTCCTCCAGAG</td>
</tr>
<tr>
<td>ViaA-YFP-F</td>
<td>TACCCGGGATGCGAAGCCCGCTGCTAGCAGCTGGCGGCGAATGGTGAGCAAGGGCGAGGGGGGGGGGAGGAGGGGGTGGAATGGTGAGCAAGGGCGAGG</td>
</tr>
<tr>
<td>ViaA-6Gly-YFP-F</td>
<td>CAGGTTCAGTGAGATGAGAATGGCCGCTGCTGCATCGAGAGACCACATATGAATATCTCCCTTAAGGCGGGTGGGAATGGTGAGCAAGGGCGAGG</td>
</tr>
<tr>
<td>LdcI-200ext-up-F</td>
<td>GGTCGTATTAACGCAATATGAGCAGCAGCCGCCTGCTAGCAGCTGGCGGCGAATGGTGAGCAAGGGCGAGG</td>
</tr>
<tr>
<td>LdcI-200ext-down-R</td>
<td>CGAGCAGCGGCCTGGCAAGGC</td>
</tr>
<tr>
<td>CadB-200ext-up-F</td>
<td>ATGACCTGCCCTGATGATCCT CATTCAGGCTTACATCGAGA</td>
</tr>
<tr>
<td>CadB-200ext-down-R</td>
<td>TCTGGTAAAAGGTAAAACAGAAATAACACAAAAGTTTCTGAAATAACACAAAAGTTTCTG</td>
</tr>
<tr>
<td>CadC-200ext-up-F</td>
<td>TCTGGTAAAAGGTAAAACAGAAATAACACAAAAGTTTCTGAAATAACACAAAAGTTTCTG</td>
</tr>
<tr>
<td>CadC-200ext-down-R</td>
<td>AAATAACACAAAAGTTTCTGGATTCATATCTCCTCCTTTGAGCG</td>
</tr>
<tr>
<td>RavA-200ext-up-F</td>
<td>GCTGGAGAA AAAGCGTGTGTAAGCAGGCATCGAACAAG</td>
</tr>
<tr>
<td>RavA-200ext-down-R</td>
<td>GCTGGAGAA AAAGCGTGTGTGAAGCAGGCATCGAACAAG</td>
</tr>
<tr>
<td>ViaA-200ext-up-F</td>
<td>CCCACAAGGGCATCGAACAAG</td>
</tr>
<tr>
<td>ViaA-200ext-down-R</td>
<td>TGAGCTTTTCCCTCCATGGG</td>
</tr>
</tbody>
</table>
2.2.1.2 DY330 Transformation

The DY330 transformation procedure was adapted from published protocol [193]. Overnight LB culture of DY330 strain (Table 4) was diluted 100 fold in 50 mL LB and was grown at 30°C with shaking to O.D.\textsubscript{600} = 0.4 ~ 0.6. The cell culture was then incubated at 42°C for 15 min to induce the \(\lambda\) Red recombination system, after which it was placed in ice-water slurry for 10 min. The cooled cell culture was pelleted and concentrated 100 fold by re-suspending in 500 \(\mu\)L ice-cold sterile water. The cell suspension was then washed three times with 500 \(\mu\)L ice-cold sterile water. After the final wash, the cell pellet was made into 5\(\times\)100 \(\mu\)L aliquots stored at \(-80^\circ\)C that were sufficient for five electroporation reactions (~10\(^8\) cells / reaction).

Purified \(Kan^R\) cassette (~ 400 ng) was mixed with 100 \(\mu\)L aliquot of DY330 electro-competent cells in a pre-cooled electroporation cuvette (0.1 cm pathlength). Electroporation was performed by using a Bio-Rad Gene Pulser set at 1.8 kV, 25 \(\mu\)F with Pulse controller of 200 Ohms. Cells were recovered in 1mL LB at 30°C for one and half hours. Cells were then selected on LB/Kan (25 \(\mu\)g/mL Kan) agar plates at 30°C overnight. Colonies were verified by colony PCR using \textit{Taq} polymerase from \textit{Fermentas}. Four primers were used for three sets of reactions: cadC-ver-F and yjdL-ver-R; cadC-ver-F and k2rc-R; kt-new-F and yjdL-ver-R (Table 3). The PCR products were sequenced at the TCAG Sequencing Facility at Sickkids Hospital, Toronto.
### Table 4: List of *E. coli* strains used in the thesis work.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG1655</td>
<td>$F^{-} rph^{-}$</td>
<td>[131]</td>
</tr>
<tr>
<td>MG1655 ΔrelA ΔspoT</td>
<td>MG1655ΔrelAΔspoT</td>
<td>[131]</td>
</tr>
<tr>
<td>MG1655 ΔcadA ΔcadB</td>
<td>MG1655 ΔcadA ΔcadB</td>
<td>This work</td>
</tr>
<tr>
<td>DY330</td>
<td>W3110 ΔlacU169 gal490 λcl1857 Δ(cro-bioA)</td>
<td>[193]</td>
</tr>
<tr>
<td>BW25141</td>
<td>lacI&lt;sup&gt;q&lt;/sup&gt; rrrB&lt;sub&gt;T14&lt;/sub&gt; ΔlacZ&lt;sub&gt;W136&lt;/sub&gt; ΔphoBR580 hsdR514 ΔaraBAD&lt;sub&gt;RM33&lt;/sub&gt; ΔrhaBAD&lt;sub&gt;LD78&lt;/sub&gt; galU95 endA&lt;sub&gt;BT333&lt;/sub&gt; uidA(ΔMluI)::pir&lt;sup&gt;+&lt;/sup&gt; recA1</td>
<td>[194]</td>
</tr>
<tr>
<td>DH5α</td>
<td>$F, \phi 80$ lacZDM15, D(lacZYA&lt;sup&gt;−&lt;/sup&gt; argF)U169, deo&lt;sup&gt;R&lt;/sup&gt;, recA1, endA1, hsdR17(rk&lt;sup&gt;−&lt;/sup&gt;, mk&lt;sup&gt;+&lt;/sup&gt;), phoA, supE&lt;sub&gt;44&lt;/sub&gt;, I&lt;sup&gt;+&lt;/sup&gt;, thi-1, gyrA96, relA1, λ&lt;sup&gt;+&lt;/sup&gt;</td>
<td>This work</td>
</tr>
<tr>
<td>BL21 Gold</td>
<td>$F^-, ompT, hsdS(r_B^-, m_B^-), dcm^-, TetR, gal, endA, Hte$</td>
<td></td>
</tr>
</tbody>
</table>
2.2.1.3 P1 Transduction

The O/N LB culture of DY330 ΔcadBA::Kan^R cells was diluted 100 fold in LB with 5 mM CaCl_2 and 10 mM MgSO_4. The mixture was incubated at 30°C for 1 hour with shaking until it showed slight turbidity. The culture was then lysed with 50 μL of P1 phage at 30°C for two hours. Lysed culture was then mixed with 150 μL chloroform. After centrifuging at 5000 x g for 10 min at 4°C, the phage lysate (supernatant) was stored at 4°C.

A 150 μL of O/N LB culture of each of the MG1655 and MG1655ΔrelAΔspoT strains (Table 4, obtained from [131]) was pelleted and the cell pellet was re-suspended in 150 μL (300 μL for MG1655ΔrelAΔspoT) fresh MC buffer (10 mM MgSO_4 and 5 mM CaCl_2). 80 μL of phage lysate of DY330 ΔcadBA::Kan^R (160 μL for MG1655ΔrelAΔspoT) was then added and the mixture was incubated for 20 min at 30°C. 230 μL (460 μL for MG1655ΔrelAΔspoT) of 1 M sodium citrate (Na_3C_3H_5O(COO)_3), a chelator for divalent ions, was then added to each mixture to stop transduction. Cells were then pelleted, re-suspended in 1 mL LB with 5 mM Na_3Citrate, and recovered at 30°C for 1 hour. The ΔcadBA::Kan^R transductants were selected overnight on LB/Kan agar plates with 5 mM sodium citrate at 30°C. Several colonies were then re-streaked sequentially on fresh LB/Kan plates for two days, and on LB plates for the third day to eliminate remaining phages. The colonies were verified by PCR as described in Section 2.2.1.2.

2.2.1.4 Elimination of Kan^R Cassette

The pCP20 plasmid (Table 2) encodes a FLP recombinase under a thermal inducible promoter. It is used to remove the Kan^R cassette in the ΔcadBA::Kan^R strains [194]. pCP20 carries an ampicillin-resistance (Amp^R) marker, a chloramphenicol-resistance (Cm^R) marker and shows a temperature-sensitive replication. The pCP20 plasmid was transformed into the two ΔcadBA::Kan^R MG1655 strains by electroporation performed similarly to that described in Section 2.2.1.2 except for the following changes. Cells were grown to an O.D._600 of 0.4 ~ 0.6 in 25 mL SOB medium (5.0 g/L yeast extract, 20.0 g/L tryptone, 400 mM NaCl, 100 mM KCl, 40 mM MgCl_2, and 40 mM MgSO_4) with kanamycin (25 μg/mL). The washing step was performed with ice-cold 10% (v/v) glycerol. After electroporation, the cells were recovered in SOC (SOB + 20 mM glucose) and selected overnight on LB/Amp (100 μg/mL Amp) plates at 30°C. Several Amp^R colonies were then re-streaked on LB plates to grow at 37°C O/N for induction of FLP and curing of pCP20. Loss of all antibiotic resistance was tested on plates with corresponding
antibiotics. The colonies were verified by colony PCR for the removal of the Kan\textsuperscript{R} cassette as described in Section 2.2.1.2.

### 2.2.2 Generation of Complementation Strains

The following complementation strains were made: MG1655 (\textit{ΔrelAΔspoT} \textit{ΔcadBA} \textit{pcadBA}), MG1655(\textit{ΔrelAΔspoT} \textit{ΔcadBA} \textit{pcadBA-LdcI-R97A}), MG1655(\textit{ΔrelAΔspoT} \textit{ΔcadBA} \textit{pcadBA-LdcI-R206S}), and MG1655(\textit{ΔrelAΔspoT} \textit{ΔcadBA} \textit{pcadBA-LdcI-L89R}).

The full length \textit{cadBA} operon was cloned into the p322 plasmid generated in the following steps. The multiple cloning site (MCS) from pET22b (Table 2) was used to replace the MCS of pET3a (Table 2) at \textit{BglII} and \textit{Bpi} restriction sites. Short complementary oligonucleotides were generated containing the \textit{SmaI} cut site sandwiched between \textit{EcoRI} and \textit{SacI} in the MCS of pET22b. Both this short oligonucleotide and pET3a were digested with \textit{EcoRI} and \textit{SacI} and, subsequently, ligated to generate p322 (Table 2). The \textit{cadBA} operon, including 401 base pairs upstream of the \textit{cadB} gene which is its promoter region, was PCR amplified and the PCR product was digested with \textit{NdeI} and \textit{BamHI} and cloned into the p322 to generate the pcadBA plasmid. The p322 plasmid alone, the p322 plasmid expressing WT LdcI, LdcI R97A, LdcI R206S or LdcI L89R was transformed into MG1655 WT and MG1655 \textit{ΔrelAΔspoT} strains by electroporation.

### 2.2.3 Media Shift Assay

#### 2.2.3.1 Media Shift Assay to Test the Inhibition of LdcI by ppGpp

The complementation strains made above were grown to an O.D.\textsubscript{600} of between 0.4 and 0.5 with shaking at 37°C in defined rich media [200]: 40 mM MOPS-MES pH 5, 1.32 mM K\textsubscript{2}HPO\textsubscript{4}, 0.2 mM ACGU (Adenine, Cytosine, Guanine and Uracil), 50 ng/mL of each of the 20 amino acids, 0.1% (w/v) D-glucose and 30 mM lysine. At mid-log phase, cells were pelleted by centrifugation at 4000 g at 22°C for 10 min, washed once in the post-shift media (5 mM MES pH 5.0, 1.32 mM K\textsubscript{2}HPO\textsubscript{4}, 0.2 mM ACGU, and 0.001% (w/v) D-glucose), pelleted and finally re-suspended to an O.D.\textsubscript{600} of 1.0 in post-shift media containing 30 mM L-lysine. The cells were incubated at 37°C with shaking for 2 hours and samples were withdrawn every half an hour to measure O.D.\textsubscript{600}, pH, LdcI levels by Western blot and the concentration of ppGpp.
The extraction of ppGpp from the cell culture was performed following published protocol in the following steps [201]. At O.D.₆₀₀ ≈ 0.5, cell samples were lysed by rapidly pouring into 23 M formic acid to reach a final concentration of 1 M. They were then frozen at -80°C in liquid nitrogen, thawed and extracted for 30 min at 4°C and pelleted. The supernatant which contained the nucleotides were diluted 20 fold and passed over an anion-exchange Q-sepharose column and eluted with 1 M ammonium formate. The eluate was put in 100 Da-cutoff dialysis tubing (Spectra/Por) and was dialyzed against 1 M sucrose for 40 hours, with changing to fresh 1M sucrose every 12 hours. The dialyzed sample was then lyophilized for 12 to 16 hours. The nucleotides were then resuspended in 500 µL water. ppGpp was detected and the levels measured by Agilent 1100 HPLC using anion-exchange chromatography with a Waters Spherisorb S5 SAX (4.6 by 250 mm) column [201].

2.2.3.2 Media Shift Assay to Test the in vivo Effect of RavA

Four strains were used in the media shift assay: MG1655 pST39-RavA/pT7pol26, MG1655 pST39-RavAΔLARA/pT7pol26, MG1655 pST39-RavAK52Q/pT7pol26 and MG1655ΔcadBA (Table 2). The strains were made by sequentially transforming the pT7POL26 plasmid (Table 2) and the pST39-RavA, pST39-RavAΔLARA and pST39-RavAK52Q plasmids into MG1655 WT strain by electroporation.

The media shift assay was performed as described above (Section 2.2.3.1) with the following modifications: O/N culture of each of the four strains was grown in 500 mL defined rich media at 37°C with initial O.D.₆₀₀ ≈ 0.01. When the O.D.₆₀₀ reached 0.2, 0.2 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) was added. The cells were then incubated at 30°C for 1 hr to induce the expression of RavA (WT or mutants) until the final O.D.₆₀₀ was approximately 0.4. Cells were then harvested by pelleting at 5000 × g at 22°C for 10 min, and washed in the post-shift media with 30 mM lysine. Samples were withdrawn every 30 min for up to 2 hrs to measure O.D.₆₀₀, pH change, and levels of LdcI and RavA by Western blot.

2.3 Results

2.3.1 Inhibition of LdcI by ppGpp in vivo

The media shift assay was developed to test the inhibition of LdcI by ppGpp in vivo (Figure 13). LdcI catalyzes the decarboxylation of lysine (+1 charge at pH 7) to cadaverine (+2 charge at pH
7). As more cadaverine is being produced and transported into the media as a result of the activity of both LdcI and CadB, the pH of the weakly buffered post-shift media (5 mM MES) is expected to increase. Thus, the activity of LdcI in cells was reflected indirectly as an increase in pH.

At first, cells were grown in the defined rich media. The media was strongly buffered at pH 5.0, supplemented with all the amino acids and with an additional 30 mM L-lysine. This condition induced the abundant expression of LdcI. After shifting the cells to the post-shift minimal media, amino acids starvation induced the stringent response and the production of ppGpp. The activity of LdcI and the levels of ppGpp were followed as a function of time (Figure 13A). As expected, the ΔcadBA strain did not show an increase in media pH because it lacks the LdcI system. The mutant strains where LdcI is not inhibited by ppGpp (ΔcadBA pcadBA-R97A, ΔcadBA pcadBA-R206S, and ΔcadBA pcadBA-L89R) increased pH at a faster rate than the WT strains (MG1655 WT and ΔcadBA pcadBAWT). The O.D.₆₀₀ stayed constant post-shift, and so did the levels of LdcI across all strains throughout the time course. In addition, ppGpp was only detected post-shift and its level remained relatively stable throughout the time course and across all strains within the errors of measurement (Figure 13 B and C). Overall, these results showed that the difference in the rate of pH increase reflects a difference in LdcI activities.
Figure 13: Inhibition of LdcI by ppGpp in vivo.

A) Wild type and ΔcadBA knockout strains were grown in defined rich media containing 30 mM L-lysine at pH 5.0 until mid-log phase (pre-shift), then the cells were shifted to minimal media also containing 30 mM L-lysine weakly buffered at pH 5.0 (post-shift). The O.D.₆₀₀ of the cells is shown in the top panel, the pH of the culture media is shown in the middle panel, and the amount of ppGpp extracted from the cultures is shown in the bottom panel. The insert shows an HPLC trace demonstrating the induction of ppGpp in the ΔcadBA pcadBA-R97A strain post-shift. These experiments were repeated three times.

B) Western blot shows the LdcI level at indicated time point during the shift assay. 15 µg of total cell extract was loaded in each lane.

C) Quantification of the LdcI level from the Western blot in (B) at the indicated time points shows that LdcI level stayed constant pre- and post-shift. The quantification was performed with the QuantityOne software.
To further confirm that ppGpp accounts for the difference in the *in vivo* activity between the LdcI wild type and mutants, the assay was repeated with the complementation strains in a ΔrelA ΔspoT ΔcadBA background which is not able to produce ppGpp [111]. It can be seen that while the O.D.₆₀₀ and the LdcI levels stayed constant, strains expressing WT LdcI increased pH at the same rate as the LdcI mutants that do not bind to ppGpp (Figure 14). Therefore, results from the above assay showed that LdcI is indeed inhibited by ppGpp *in vivo*.

In the shift assays, it should be noted that the activity of the LdcI ppGpp binding site mutant (LdcI R97A and LdcI R206S) was similar to the dimer mutant (LdcI L89R) *in vivo*, while they had approximately five-fold higher activity than the latter *in vitro* (Figure 11E). This shows that the inhibition of WT LdcI *in vivo* was not as complete as *in vitro* (Figure 11E). This observation can be attributed to either a sensitivity issue with the given experimental design, or that there are other rate-limiting factors such as saturation of the activity of CadB [2].
Figure 14: Media Shift Assay on $\Delta r e l A \Delta s p o T \Delta c a d B A$ strains.

(A) A shift experiment from defined rich media into weakly buffered, lysine-supplemented media was performed on $\Delta r e l A \Delta s p o T \Delta c a d B A$ strains expressing either the $p322$ vector, or $p322$ with WT $pcadBA$, $pcadBA-R97A$, $pcadBA-R206S$, or $pcadBA-L89R$. Upon shift, the $\Delta r e l A \Delta s p o T \Delta c a d B A p322$ strain did not change the pH while all the other strains showed identical shifts in pH. These experiments were repeated twice. (B) Western blot for LdcI at the pre-shift and post-shift timepoints. 15 µg of total cell extract was loaded in each lane and LdcI was detected with $\alpha$-LdcI antibody. (C) The amount of LdcI from the Western blot in (B) was quantified and plotted as a bar graph. The levels of LdcI for most of the strains remained relatively constant pre- and post-shift. The quantification was performed with the QuantityOne software.
2.3.2 Antagonistic Regulation of RavA and ppGpp on the Activity of LdcI

Four strains were used in the assay: MG1655ΔcadBA, MG1655 + RavA, MG1655 + RavAΔLARA, and MG1655 + RavA (K52Q). RavA (K52Q) has a mutation on the conserved K52 residue in the Walker A motif that is responsible for ATP binding. Therefore, this RavA mutant cannot bind ATP [173]. The RavA proteins were over-expressed from the pST39 plasmids. As shown in Figure 15, the O.D.₆₀₀ stayed constant post-shift throughout the time course of the experiment. The pH/O.D.₆₀₀ ratio instead of pH alone was used to measure LdcI activity. As the experiments were repeated at least three times on different days, the pH value is normalized against cell numbers in order to make the values comparable across different days. Consistent with the in vitro results, the WT + ravA strain increased the pH of the media at a higher rate than the WT + ravAΔLARA strain, while no pH change was observed for the ΔcadBA cells (Figure 15A). LdcI was more strongly inhibited by ppGpp when RavAΔLARA was overexpressed than when WT RavA was over-expressed because the RavA truncation mutant cannot form a complex with LdcI. The strain overexpressing RavA (K52Q) mutant increased pH faster than the strain overexpressing RavAΔLARA, but slower that the strain overexpressing WT RavA (Figure 15A). This observation thus indicates that the ATPase activity of RavA, in addition to the LARA domain, is also required for modulating the inhibition of ppGpp. In addition, Western blot analysis showed that both LdcI and RavA (WT and mutants) were detected at similar levels pre- and post-shift (Figure 15B).

In conclusion, the above results confirmed that formation of the RavA-LdcI complex reduced the inhibitory effect of the ppGpp on LdcI in vivo.
Figure 15: RavA antagonizes the inhibitory effect of ppGpp on LdcI activity in vivo.

A) The effect of RavA overexpression on LdcI activity in the cell. The ΔcadBA strains and WT cells overexpressing RavA, RavAΔLARA or RavA (K52Q) were grown to log-phase in defined rich media buffered at pH 5. RavA, RavAΔLARA, or RavA (K52Q) was induced and cells were then shifted to minimal media weakly buffered at pH 5, supplemented with 30 mM lysine and no amino acids. The O.D.₆₀₀ of the cells is shown in the top panels, while the pH of the culture media is shown in the bottom left panel. The bottom right panel shows the increase in pH/O.D.₆₀₀ normalized to the value at 0+ (right after shift). Error bars represent the standard deviations of at least three replicates.

B) The levels of LdcI, RavA, RavAΔLARA, and RavA(K52Q) pre- and post-shift were determined by Western blot analysis at the indicated time points in the media shift assays. The RavA(K52Q) gel was run separately from the other three. RavA is present after IPTG induction for all strains. Endogenous expression of RavA is too low in log-phase cells to be detected.
2.4 Discussion

The inhibition of LdcI by the direct binding of ppGpp is the first piece of evidence that links the acid stress response to the stringent response. LdcI is then another protein that directly binds ppGpp, in addition to many others including *E. coli* RNAP, translational factors IF-2, and the DNA primase DnaG (Section 1.2) [138, 156, 157].

The inhibition of LdcI by ppGpp is confirmed by both *in vitro* and *in vivo* methods. The *in vitro* TNBS assay demonstrated that ppGpp inhibits the decameric form of LdcI *in vitro* at between pH 6 to pH 8, which is physiologically relevant because the cytoplasmic pH is 2 units higher than the extracellular environment in acidic environment (e.g.pH 5) [5]. At pH lower than 5, LdcI cannot bind and be inhibited by ppGpp likely because of the protonation of the negatively charged phosphate groups on ppGpp (pKa 6.52 for 3' β phosphate, and 6.82 for 5' β phosphate).

Physiologically, this inhibition may reflect the cell’s decision, when facing the dilemma of both acid stress response and stringent response, to conserve amino acids during nutritional starvation (Figure 16). Thus, ppGpp is used to fine-tune the bacterial acid stress response via allosteric inhibition. Under the setting of the media shift assay, the intracellular concentration of LdcI is approximately 20 to 30 µM, which corresponds to high concentration as defined in the *in vitro* assays. At this concentration and intracellular pH of approximately 7.0, the majority of LdcI in the cell is decameric and is inhibitable by ppGpp. The concentration of ppGpp increases to approximately 1 to 2 mM upon amino acid starvation. As the binding constant of ppGpp to LdcI is in the nM range, it was expected that the LdcI be completely inhibited during the stringent response given that the cytoplasmic pH is also close to neutral [202]. Although WT LdcI protein did show a lower activity than the ppGpp binding site mutants *in vivo*, the incomplete inhibition of the protein suggests that additional regulatory systems of LdcI may be present.

RavA binding to LdcI reduces the effect of ppGpp inhibition both *in vitro* and *in vivo*. The LARA domain of RavA binds to LdcI and the ATPase domain of RavA is required for this antagonistic effect. In addition, both RavA and ppGpp have similar binding constants to LdcI: $K_d$ of 0.02 – 1 µM for RavA and $K_d$ of 0.01 – 0.7 µM for ppGpp [2]. Based on these observations, it can be speculated that RavA carries out this regulation either by blocking the access to the
ppGpp binding site in LdcI, or by inducing a local conformational change in LdcI that reduces its ppGpp binding affinity. Alternatively, ppGpp might cause a conformational change in LdcI to reduce RavA binding to the decarboxylase. In conclusion, RavA may play a more important role in regulating the LdcI inhibition by ppGpp in either stationary phase or under anaerobic conditions when it is highly expressed. Detailed study of the function of RavA and the RavA-LdcI cage will greatly improve the understanding of the RavA-LdcI interaction.
Figure 16: Physiology significance of the inhibition of LdcI by ppGpp.

When nutrients are sufficient (beige background), LdcI decamers (pink flower-like shapes) actively decarboxylize lysine to decrease cytoplasmic proton concentration at pH 5. Upon starvation (white background), LdcI is inhibited by ppGpp, possibly for cells to conserve lysine to maintain cell survival.
3 Cellular Localization of Ldcl, CadB, CadC, RavA and ViaA

Note: I performed all the experiments described in this chapter.
3.1 Introduction

In the acid resistance pathway 2 (AR2) identified in *E. coli*, the glutamate decarboxylase pathway enables the bacteria to survive under extreme acidic conditions (pH 2~3). The structure of the glutamate decarboxylase GadB was solved at two pHs, pH 4.6 in its active form, and pH 7.6 in its inactive form (Figure 3) [10]. *Capitani, G., et al.* found that at pH 4.6, the N-terminus of each monomer in the hexameric GadB changes from disordered structures at pH 7.6 to a helix (Figure 3). GadB was found to localize more to a membrane fraction by Western blotting and this experiment further suggests that the conformational change in GadB is involved in the migration of GadB to the membrane, possibly to interact with GadC, the glutamate-GABA antiporter. The lysine decarboxylase pathway (AR4) is composed of the inducible lysine decarboxylase (LdcI) and the lysine-cadaverine antiporter CadB. Therefore, the observation from GadB prompted us to investigate whether LdcI in the lysine decarboxylase pathway shows similar membrane localization upon acid stress.

In an effort to study the localization of the different proteins involved in the lysine-dependent acid stress response, I have endogenously tagged LdcI (CadA), CadB, CadC, RavA and ViaA with different fluorescent proteins (GFP/RFP/YFP) to observe their potential re-localizations under different stress conditions. LdcI (CadA) and CadC were tagged with a GFP variant called superfolder GFP (sfGFP) [203, 204]. The sfGFP folds well with its fusion partners, is resistant to degradation or denaturation and also fluoresces in the periplasm. CadB and RavA were tagged with a monomeric RFP (mRFP) [196], which is a RFP variant that matures fast and fluoresces in monomeric form in the cell. ViaA was tagged with a monomeric YFP (mCitrine), the sequence of which is from a pRSETB-CKAR plasmid used in a FRET (fluorescence resonance energy transfer) study [197]. These tagged strains will be very useful tools for studying the localization behaviour of each protein under aerobic, anaerobic and acid stress conditions.
3.2 Materials and Methods

3.2.1 Overall Strategy for Endogenously Tagging *ldcl, cadB, cadC, ravA and viaA*

The overall strategy for endogenously tagging is shown in Figure 17. The DNA sequences of sfGFP, mRFP and YFP were first cloned into a pKD3 plasmid upstream of the chloramphenicol cassette (Table 2). The cassette is flanked by two *flp* recombinase target sites (FRT) and can be removed via site-specific recombination by Flp recombinase (Figure 17). The *GFP/RFP/YFP-cat* cassette was then amplified by PCR with primers that contain homologous sequence to the C-terminus end of the genes to be tagged. The PCR products were then transformed into DY330 strains which encodes an endogenous λ Red recombination system to be homologously recombined into the C-termini of the corresponding genes. The tags were then transferred to the MG1655 WT strain by P1 phage transduction. Lastly, the *cat* cassette was removed by Flp recombinase.
Figure 17: Schematic depiction of the cloning of sfGFP, mRFP and YFP into the pKD3 plasmid.

A) The pKD3 plasmid contains a $cat^R$ cassette flanked by Flp recombinases sites (FRT) sites. B) Each of the fluorescent protein genes is cloned with the restriction enzyme indicated. H1 and H2 on the forward and reverse primers, respectively, indicate the 40 bp DNA sequences homologous to the C-termini of the five tagged genes.
3.2.2 Cloning sfGFP, mRFP and YFP into the pKD3 Plasmid

The plasmid for sfGFP was obtained from Dr. Robert Sauer at Massachusetts Institute of Technology (Table 2) [195]. The mRFP plasmid was obtained from Dr. John Brumell’s lab at Sickkids Hospital, Toronto (Table 2). The YFP plasmid was purchased from Addgene (Table 2) [197].

The DNA sequences of sfGFP, mRFP and YFP were cloned into a pKD3 plasmid using either SphI (for sfGFP) or SphI and AflIII (for mRFP and YFP) to generate pKD3-sfGFP, pKD3-mRFP and pKD3-YFP plasmids. Only SphI was used for cloning sfGFP into pKD3 because sfGFP DNA sequence contains an internal AflIII restriction site (Figure 17B).

3.2.3 Preparation of the DNA Cassettes for Tagging ldcl, cadB, cadC, ravA and viaA

The DNA cassettes used for tagging by homologous recombination for each of the five genes were amplified by PCR from pKD3-sfGFP, pKD3-mRFP or pKD3-YFP plasmids following a standard protocol for Phusion polymerase from Fermentas. The primers and plasmids used are listed in Table 2 and Table 3. The schematic diagram for the cassettes is also given in the Table 5. The concentration of the PCR products was between 70~100 ng/μL as measured by Nanodrop1000 spectrophotometer from ThermoScientific.
Table 5: Primers and plasmids used for tagging *ldcI, cadB, cadC, ravA* and *viaA*.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primers used</th>
<th>Plasmid template</th>
<th>Cassette generated*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cadA</strong></td>
<td>LdcI-sfGFP-F, LdcI-sfGFP-R</td>
<td>pKD3-sfGFP</td>
<td>5' 40bp sfGFP FRT cat FRT 40bp 3'</td>
</tr>
<tr>
<td></td>
<td>LdcI-6Gly-sfGFP-F, LdcI-6Gly-sfGFP-R</td>
<td>pKD3-sfGFP</td>
<td>5' 40bp 6Gly sfGFP FRT cat FRT 40bp 3'</td>
</tr>
<tr>
<td><strong>cadB</strong></td>
<td>CadB-mRFP-F, CadB-mRFP-R</td>
<td>pKD3-mRFP</td>
<td>5' 40bp mRFP FRT cat FRT 40bp 3'</td>
</tr>
<tr>
<td></td>
<td>CadB-6Gly-mRFP-F, CadB-6Gly-mRFP-R</td>
<td>pKD3-mRFP</td>
<td>5' 40bp 6Gly mRFP FRT cat FRT 40bp 3'</td>
</tr>
<tr>
<td><strong>cadC</strong></td>
<td>CadC-sfGFP-F, CadC-sfGFP-R, CadC-6Gly-sfGFP-F, CadC-6Gly-sfGFP-R</td>
<td>pKD3-sfGFP</td>
<td>5' 40bp sfGFP FRT cat FRT 40bp 3'</td>
</tr>
<tr>
<td><strong>ravA</strong></td>
<td>RavA-mRFP-F, RavA-mRFP-R</td>
<td>pKD3-mRFP</td>
<td>5' 40bp mRFP FRT cat FRT 40bp 3'</td>
</tr>
<tr>
<td></td>
<td>RavA-6Gly-mRFP-F, RavA-6Gly-mRFP-R</td>
<td>pKD3-mRFP</td>
<td>5' 40bp 6Gly mRFP FRT cat FRT 40bp 3'</td>
</tr>
</tbody>
</table>

*40bp refers to 40 bp DNA sequences homologous to the 3' of the corresponding gene*
3.2.4 DY330 Transformation

The DY330 transformation procedure is adapted from a published protocol and seeks to optimize electroporation efficiency [193]. A total of 10 electroporation reactions were carried out to make 10 tagging strains for the five genes. Each electroporation reaction was performed as follows: O/N DY330 LB culture were diluted 100 fold in 25 mL LB and were grown at 30°C to O.D.₆₀₀= 0.4~0.6 (mid-log phase) in an 125 mL flask. Cells were then incubated at 42°C for 15 min to induce the λ Red recombination system. Immediately afterwards, the flask was cooled down in an ice-water slurry for 10 min. The cell culture was then pelleted at 500 × g for 15 min at 4°C. The pellet was then re-suspended and washed for three times, each time with 20 mL ice-cold water at 500 × g for 15 min at 4°C. Lastly, the pellet was re-suspended in 100 μL ice-cold sterile water. This was enough for one electroporation reaction (~10¹⁰ cells / reaction).

Subsequently, 3 μL of purified Cam⁺ cassettes (≥70 ng/μL) was mixed with one 100 μL-aliquot of DY330 electro-competent cells. Eletroporation was performed by using a Bio-Rad Gene Pulser set at 1.8 kV, 25 μF with Pulse controller of 200 Ohms in a pre-cooled electroporation cuvette (0.1 cm). Half of the cells were recovered in 1 mL LB at 30°C for 3 hours and plated on LB/Cam (34 μg/mL) plates. The other half was plated after overnight recovery at 30°C. After incubating at 30°C for two days, colonies of varying sizes were picked and re-streaked on LB/Cam plates. Colony PCR was performed to verify the correction insertion using *Phusion* polymerase from *Fermentas*. The primers used were as following (Table 3): LdcI-ext200-up-F and LdcI-ext200-down-R for LdcI, cadB-ext200-up-F and cadB-ext200-down-R for CadB, cadC-ext200-up-F and cadC-ext200-down-R for CadC, ravA-ext200-up-F and ravA-ext200-down-R for RavA, and viaA-ext200-up-F and viaA-ext200-down-R for ViaA.

3.2.5 P1 Transduction

Overnight LB cultures of each of the 10 tagged DY330 strains were each diluted 100 fold in LB with 0.005 M CaCl₂ and 0.01 M MgSO₄. The mixture was incubated at 30°C for 1 hour with shaking when slight turbidity was observed. 50 μL of P1 phage was then added and the culture was incubated at 30°C for 2~4 hours for cell lysis. Lysed culture was then transferred to a 15 mL falcon tube and mixed with 150 μL chloroform by vortexing. After centrifuging at 5000 x g for 10 min at 4°C, the supernatant (phage lysate) was stored at 4°C.
The WT MG1655 strain was obtained from Traxler et al. [131]. 150 μL of O/N LB MG1655 cultures grown at 37°C was spun down. The cell pellet was re-suspended in 150 μL fresh MC buffer (0.01 M MgSO₄ and 0.005 M CaCl₂). 80 μL of phage lysate of the tagged DY330 strains was then added and the mixture was incubated for 20 min at 37°C with shaking. 230 μl of 1 M sodium citrate (Na₃C₅H₅O (COO)₃), a chelator for divalent ions, was then added to each mixture to stop transduction. Cells were then pelleted, re-suspended in 1 mL LB with 5 mM Na₃Citrate, and incubated at 37°C with shaking for 1 hour for recovery. The DY330 Cam⁰ transductants were selected on LB/Cam agar plates with 5mM citrate after O/N growth at 37°C. Several colonies were then re-streaked sequentially on fresh LB/Cam plates for two days and on LB plates for the third day to eliminate remaining phages. Colonies were verified by colony PCR with the same primers used for verification in DY330 strain as described in Section 3.2.4.

3.2.6 Elimination of Cam⁰ Cassette

The pCP20 plasmid (Table 2) is used to remove the Cam⁰ cassette in the 10 tagged MG1655 strains [194]. It was transformed into the tagged MG1655 Cam⁰ strains by electroporation. For curing and removing the Cam⁰ marker, several Amp⁰ Cam⁰ colonies were incubated at 42°C in LB media for 5 to 6 hours and then streaked on LB plates to grow O/N at 37°C. This simultaneously induced the expression of FLP and cured pCP20. Loss of all antibiotic resistances was then confirmed by absence of growth on both Amp and Cam LB plates. Removal of the cam⁰ cassette was verified by colony PCR with the same primers used for verification in DY330 strain as described in Section 3.2.4.

3.2.7 Western Blot

After the genes have been successfully tagged, LdcI, RavA and ViaA proteins were also verified by Western blot.

For LdcI, MG1655 ldcI-sfGFP, MG1655 ldcI-6Gly-sfGFP, and WT MG1655 strains were grown to O.D₆⁰₀ ≈ 0.5 (mid-log phase) at 37°C in defined rich media (40 mM MOPS-MES pH 5, 1.32 mM K₂HPO₄, 0.2 mM ACGU, 50 ng/mL of each of the 20 amino acids, 0.1% (w/v) D-glucose and 30 mM lysine). They were then lysed by sonication. The total lysate, supernatant and pellet fractions were taken for each strain. Protein concentration of each sample was then
determined by Bradford reagent and equal amounts of total protein were loaded to perform a Western blot using either α-LdcI or α-GFP antibody.

For RavA and ViaA, 50 mL of each of the MG1655 rava-mRFP, MG1655 rava-6Gly-mRFP, MG1655 viaA-YFP, MG1655 viaA-6Gly-YFP and WT MG1655 strains was grown at 37°C under anaerobic conditions to either O.D.\textsubscript{600} ≈ 0.5 (mid-log phase) (5 hrs) or stationary phase (24 hrs). Cells samples were lysed by sonication and Western blotted using either α-RavA or α-ViaA antibody.

### 3.2.8 Fluorescence Microscopy

Fluorescence images of the 10 tagged strains were taken using either Nikon ECLIPSE 80i (for LdcI and CadB) or a Zeiss LSM510 confocal microscope (RavA, ViaA). For observing LdcI, cells were first grown in LB media for 2 hrs to an O.D.\textsubscript{600} ≈ 0.5, and then shifted to defined rich media at pH 5 (40 mM MOPS-MES pH 5, 1.32 mM K\textsubscript{2}HPO\textsubscript{4}, 0.2 mM ACGU, 50 ng/mL of each of the 20 amino acids, 0.1% (w/v) D-glucose and 30 mM lysine). The images for LdcI were taken at 30 mins after the acid shift. The images for CadB were taken after growing cells directly in the same defined rich media (pH 5) to O.D.\textsubscript{600} ≈ 0.5 (mid-log phase). The images for RavA and ViaA were taken with cells grown under the same conditions for obtaining samples for Western blotting as described above. Before observation, cells were lightly fixed in 4% formaldehyde for 10 min.

### 3.2.9 Media Shift Assay

The \textit{in vivo} activity of LdcI-(6Gly)-sfGFP, CadB-(6Gly)-mRFP and CadC-(6Gly)-sfGFP were tested by a media shift assay using WT MG1655 and MG1655 ΔcadBA strains as controls. The assay was performed twice for each strain. Cells were grown to an O.D.\textsubscript{600} of between 0.4 and 0.5 with shaking at 37°C in 100 mL defined rich media (40 mM MOPS-MES pH 5, 1.32 mM K\textsubscript{2}HPO\textsubscript{4}, 0.2 mM ACGU, 50 ng/mL of each of the 20 amino acids, 0.1% (w/v) D-glucose and 30 mM lysine) [200]. At mid-log phase, cells were pelleted by centrifugation at 4000 g at 22°C for 10 min, washed once in the post-shift media (5 mM MES pH 5.0, 1.32 mM K\textsubscript{2}HPO\textsubscript{4}, 0.2 mM each of adenine, cytosine, guanine and uracil, and 0.001% (w/v) D-glucose), pelleted again as above and finally re-suspended to an O.D.\textsubscript{600} of 1.0 in post-shift media containing 30 mM L-
lysine. The cells were incubated at 37°C with shaking for 2 hours and samples were withdrawn every 30 mins to measure O.D.\textsubscript{600} and pH [201].

### 3.3 Results

There were 10 strains generated in total: MG1655 \textit{ldcI}\textsubscript{-sfGFP}, MG1655 \textit{ldcI-6Gly-sfGFP}, MG1655 \textit{cadB-mRFP}, MG1655 \textit{cadB-6Gly-mRFP}, MG1655 \textit{cadC-sfGFP}, MG1655 \textit{cadC-6Gly-sfGFP}, MG1655 \textit{ravA-mRFP}, MG1655 \textit{ravA-6Gly-mRFP}, MG1655 \textit{viaA-YFP}, MG1655 \textit{viaA-6Gly-YFP}. For each of the 5 proteins, they were tagged either with or without a 6 Glycine linker upstream of the fluorescent proteins.

#### 3.3.1 \textit{LdcI} (CadA)

\textit{LdcI} has been successfully tagged in MG1655 WT strain and the colony PCR behaved as expected (Figure 18). Using the primers as indicated, the expected size for the PCR product was 499 bp in untagged strains, 2250 bp after insertion of the cassette and 1290 bp after removal of \textit{Cam}^R marker (Figure 18). The expected molecular weight of \textit{LdcI} is 81.2 kDa and that of \textit{LdcI-sfGFP} is approximately 108 kDa (GFP is 26.7 kDa). The Western blot shows that, \textit{LdcI (6Gly)} sfGFP (~108 kDa) migrates at a higher position than WT \textit{LdcI} (81.2 kDa) according to the expected size (Figure 19A). The \textit{LdcI-sfGFP} fusion protein was detected by both α-\textit{LdcI} and α-GFP antibody, but the WT \textit{LdcI} was only detected by the α-\textit{LdcI} antibody. In addition, the WT \textit{LdcI} proteins are present almost exclusively in the supernatant fraction (Figure 19A). In contrast, the \textit{LdcI-sfGFP} and \textit{LdcI-6Gly-sfGFP} fusion proteins are present exclusively in the pellet fraction (Figure 19A), indicating that the fusion proteins are insoluble. Moreover, by fluorescence imaging, the \textit{LdcI} fusion proteins were observed to form foci structures (Figure 19B). For the majority of the cells, they contained only one green focus located either at one pole of the cell (approximately 90% of single-focus cells), in the middle, or at quarter-cell position. Only a few cells contain two foci, either at both poles, or one at one pole while the other at quarter-cell position. This localization pattern conforms to the formation and inheritance pattern of inclusion bodies described in \textit{Lindner et al.}, further confirming that the \textit{LdcI} fusion proteins form aggregates (detailed discussion in Section 3.4) [205].

On the other hand, the media shift assay showed that the sfGFP tag, in spite of promoting aggregation, did not affect the function of \textit{LdcI} (Figure 19C). \textit{LdcI} catalyzes the decarboxylation
of lysine (+1 at pH 7) to cadaverine (+2 at pH 7). As more cadaverine is being produced and transported into the media as a result of LdcI activity, the pH of the weakly buffered post-shift media (5mM MES) increases. Thus, the activity of LdcI in cells was reflected indirectly as an increase in pH. In Figure 19, it shows that while the O.D.₆₀₀ stayed constant throughout the time course, the MG1655 ldcI-sfGFP strains increased the pH at the same rate as WT MG1655. The MG1655ΔcadBA strain did not increase the pH of the media because it lacks a lysine decarboxylase system.
Figure 18: Endogenous tagging of LdcI with sfGFP.

A) Schematic diagram showing the steps involved in tagging the C-terminal of LdcI (CadA) with sfGFP and the verification of the cloning by colony PCR. The forward- and reverse-primers used are listed and their annealing positions are shown with arrows. The expected sizes of the products are indicated on the diagram. The MG1655 ldcI-sfGFP and MG1655 ldcI-6Gly sfGFP are tagged in the same fashion except that the latter has a 6 Glycine linker (indicated in bracket) upstream of sfGFP. B) The PCR results are shown with the expected size of the products indicated with colored arrows. Arrows are colored according to the expected sizes as labeled in A). Brown arrow indicates the 499-bp product. Black arrow indicates the 2250-bp product, and blue arrow indicates the 1290-bp product.
A) Western blot showing expression of Ldcl-sfGFP (108 kDa) and WT Ldcl (81.2 kDa) in different lanes. Lanes 1, 4, 7: WT MG1655; Lanes 2, 5, 8: MG1655 ldlc-sfGFP; Lanes 3, 6, 9: MG1655 ldlc-6Gly-sfGFP.

B) Fluorescence images of MG1655 ldlc-sfGFP and MG1655 ldlc-6Gly-sfGFP showing green fluorescence in corresponding bacterial cells.

C) Growth curves and pH changes over time for different strains: WT MG1655, MG1655 ldlc-lacBA, MG1655 ldlc-sfGFP, and MG1655 ldlc-6Gly-sfGFP.
**Figure 19: Verification of MG1655 ldcI-(6Gly)-sfGFP strains.**

A) Western blot results for the expression of LdcI-(6Gly)-sfGFP fusion protein. MG1655 (WT), MG1655 ldcI-sfGFP, MG1655 ldcI-6Gly-sfGFP strains were grown in to log-phase in defined rich media at pH 5 at 37˚C. The cells were then harvested and lysed. Equal amount of total protein was loaded for each of the total lysate (T), supernatant (S) and pellet (P) from each strain. LdcI was detected on 7% SDS-PAGE gel with α-LdcI antibody and GFP was detected on 10% SDS-PAGE gel with α-GFP antibody. The expected molecular weight (MW) is 82 kDa for WT LdcI, 26.7 kDa for GFP, and 108.7 kDa for LdcI-6Gly-sfGFP.

B) Fluorescence images for MG1655 ldc-6Gly-sfGFP. Cells were first grown in LB media for 2 hrs to an O.D.₆₀₀ ≈ 0.5, and then shifted to defined rich media (pH5, 30 mM lysine and 0.2% glucose). Cell samples were taken at 30 mins after the acid shift and fixed in 4% formaldehyde for 10 min. Fluorescence images were taken with Nikon Eclipse 80i microscope. Black arrow indicates cells with one green focus at the pole position. Gray arrow indicates cells with two foci. White arrow indicates cells with one focus at quarter cell position. Brown arrow indicates cells with one focus at the middle position.

C) The LdcI-(6Gly)-sfGFP has the same activity as WT LdcI as tested by the media shift assay. MG1655 WT, MG1655 ldcI-sfGFP and MG1655 ldcI-6Gly-sfGFP strains were grown to log-phase in defined rich media buffered at pH 5. They were then shifted to minimal media weakly buffered at pH 5 and supplemented with 30 mM lysine and with no amino acid. The O.D.₆₀₀ of the cells is shown in the top panels, while the pH of the culture media is shown in the bottom panel. Error bars represent the standard deviations of two experimental replicates.
3.3.2 CadB

CadB has been successfully tagged in MG1655 WT strain and the colony PCR behaved as expected (Figure 20). Using the primers as indicated, the expected size for the PCR product was 709 bp in untagged strains, 2460 bp after insertion of the cassette and 1522 bp after removal of $Cam^R$ cassette. In vivo, CadB fluorescence is located along the cell membrane (Figure 21A). In the media shift assay, the MG1655 cadB-mRFP strains behaved the same as WT MG1655 (Figure 21B). This observation shows that the mRFP tag did not affect the function of CadB at least to an extent that can be detected under the current experiment design.
Figure 20: Endogenous tagging of CadB with mRFP.

A) Schematic diagram showing the steps involved in tagging the C-terminal of CadB with mRFP and the colony PCR verification. The forward- and reverse-primers used are listed and their annealing positions are shown with arrows. The expected sizes of the products are indicated on the diagram. MG1655 cadB-mRFP and MG1655 cadB-6Gly-mRFP are tagged in the same fashion except that the latter has a 6 Glycine linker (indicated in bracket) upstream of mRFP. B) The PCR results are shown with the expected size of the products indicated with colored arrows. Arrows are colored according to the expected sizes as labeled in A). Brown arrow indicates the 709-bp product. Black arrow indicates the 2460-bp product, and blue arrow indicates the 1522-bp product.
**MG1655 cadB-mRFP**

**MG1655 cadB-6Gly-mRFP**

**B**

![Graph](image)

- O.D. 650
- pH
- Time (hr) vs. O.D. 650
- Time (min) vs. pH

Legend:
- WT MG1655
- MG1655 ΔcadBA
- MG1655 cadB-mRFP
- MG1655 cadB-6Gly-mRFP
Figure 21: Verification of MG1655 cadB-(6Gly)-mRFP strains.

A) Fluorescence images for MG1655 cadB-6Gly-mRFP strains. These images were taken after growing the corresponding strains to mid-log phase in defined rich media (pH 5, 30 mM lysine and 0.2% glucose) at 37 °C. Cell samples were fixed in 4% formaldehyde for 10 min. Images were taken using Nikon Eclipse 80i microscope. The CadB proteins fluoresce along the cell membrane.

B) The tagging of CadB with mRFP did not affect the activity of the lysine decarboxylase system. MG1655 WT, MG1655 cadB-mRFP and MG1655 cadB-6Gly-mRFP strains were grown to log-phase in defined rich media buffered at pH 5. They were then shifted to minimal media weakly buffered at pH 5 and supplemented with 30 mM lysine and with no amino acid. The O.D.600 of the cells is shown in the top panels, while the pH of the culture media is shown in the bottom left panel. Error bars represent the standard deviations of two experimental replicates.
3.3.3 CadC

CadC has been successfully tagged in MG1655 WT strain and the colony PCR behaved as expected (Figure 22). Using the primers as indicated, the expected size for the PCR product was 500 bp in untagged strains, 2251 bp after insertion of the cassette and 1281 bp after removal of *Cam^R* cassette. In the media shift assay, the MG1655 cadC-sfGFP strains increased the pH of the media at a slightly slower rate (from pH 5.6 to pH 7.3) than WT (from pH 5.6 to pH 7.5) (Figure 23). This result suggests that tagging with sfGFP slightly weakened the activity of CadC in its induction of *cadBA* operon. No fluorescence was observed under the microscope probably due to the very low concentration of CadC in the cell (approximately 20 copies per cell) [56].
Figure 22: Endogenous tagging of CadC with sfGFP.

A) Schematic diagram showing the steps involved in tagging the C-terminal of CadC with sfGFP and the colony PCR verification. The forward- and reverse-primers used are listed and their annealing positions are shown with arrows. The expected sizes of the products are indicated on the diagram. MG1655 \textit{cadC-sfGFP} and MG1655 \textit{cadC-6Gly-sfGFP} are tagged in the same fashion except that the latter has a 6 Glycine linker (indicated in bracket) upstream of sfGFP. B) The PCR results are shown with the expected size of the products indicated with colored arrows. Arrows are colored according to the expected sizes as labeled in A). Brown arrow indicates the 500-bp product. Black arrow indicates the 2251-bp product, and blue arrow indicates the 1281-bp product.
Figure 23: Verification of MG1655 *cadC-(6Gly)-sfGFP* strains.

The tagging of CadC with sfGFP affected the activity of the lysine decarboxylase system. MG1655 WT, MG1655 *cadC-sfGFP* and MG1655 *cadC-6Gly-sfGFP* strains were grown to log-phase in defined rich media buffered at pH 5. They were then shifted to minimal media weakly buffered at pH 5 and supplemented with 30 mM lysine and with no amino acid. The O.D.₆₀₀ of the cells is shown in the top panels, while the pH of the culture media is shown in the bottom left panel. Error bars represent the standard deviations of two replicates.
3.3.4 RavA

RavA has been successfully tagged in MG1655 WT strain and the colony PCR behaved as expected (Figure 24). Using the primers as indicated, the expected size for the PCR product was 749 bp in untagged strains, 2500 bp after insertion of the cassette and 1530 bp after removal of \textit{Cam}^R cassette. The insertion of the \textit{mRFP} gene at the 3' end of \textit{ravA} disrupts the \textit{viaA} gene, as represented by the small white stripe (Figure 24). As expected, RavA-(6Gly)-sfGFP, with a 26.7 kDa mRFP, migrated at a higher position (~82 kDa) than WT RavA (55 kDa) in Western blot (Figure 25A). In the MG1655 \textit{viaA-YFP} strains, WT RavA (55kDa) was expressed. However, no fluorescence was observed for the RavA tagged strains (Figure 25B).
Figure 24: Endogenous tagging of RavA with mRFP.

A) Schematic diagram showing the steps involved in tagging the C-terminal of RavA with mRFP and the colony PCR verification. The forward- and reverse-primers used are listed and their annealing positions are shown with arrows. The expected sizes of the products are indicated on the diagram. MG1655 ravA-mRFP and MG1655 ravA-6Gly-mRFP are tagged in the same fashion except that the latter has a 6 Glycine linker (indicated in bracket) upstream of mRFP. B) The PCR results are shown with the expected size of the products indicated with colored arrows. Arrows are colored according to the expected sizes as labeled in A). Brown arrow indicates the 749-bp product. Black arrow indicates the 2500-bp product, and blue arrow indicates the 1530-bp product.
Figure 25: Verification of MG1655 ravA-(6Gly)-mRFP strains.

A) Western blot for the expression of RavA in the following five strains: MG1655 (WT), MG1655 ravA-mRFP, MG1655 ravA-6Gly-mRFP, MG1655 viaA-YFP, MG1655 viaA-6Gly-YFP. These strains were grown anaerobically either to mid-log phase (after 5 hours) or stationary phase (24 hrs) in LB at 37°C. Purified RavA is included as reference. RavA is detected on a 9% SDS-PAGE gel with α-RavA antibody. The expected molecular weight (MW) is 55 kDa for RavA, and 82 kDa for RavA-6Gly-mRFP proteins. The expected size for RFP is approximately 27 kDa. B) Fluorescence images for MG1655 ravA-6Gly-mRFP. The strains were grown in LB at 37°C anaerobically for 24 hours. Images were taken with confocal microscope (Zeiss LSM510).
3.3.5 ViaA

ViaA has been successfully tagged in MG1655 WT strain and the colony PCR behaved as expected (Figure 26). Using the primers as indicated, the expected size for the PCR product was 587 bp in untagged strains, 2335 bp after insertion of the cassette and 1365 bp after removal of $Cam^R$ marker. The YFP is approximately 27 kDa. The WT ViaA (55 kDa) migrates as expected. Interestingly, despite the disruption of the $viaA$ gene at the 5’ end by the $mRFP$ sequence, RavA-(6Gly)-mRFP strains still expressed a truncated version of ViaA migrating at approximately 55 kDa (Figure 26A). The expression level of ViaA is similar in both mid-log phase and stationary phase under anaerobic growth, which contrasts with the observation in our lab that WT ViaA level peaks at mid-log phase but declines in stationary phase (personal communication with K.W in Dr. Walid Houry lab). Thus, disruption of the $viaA$ gene by the $mRFP$ gene seems to induce overexpression of the truncated ViaA. However, no fluorescence was observed for the ViaA tagged strains (Figure 26B).
Figure 26: Endogenous tagging of ViaA with YFP.

A) Schematic diagram showing the steps involved in tagging the C-terminal of ViaA with YFP and that for the colony PCR verification are depicted. The forward- and reverse- primers used are listed and their annealing positions are shown with arrows. The expected sizes of the products are indicated on the diagram. MG1655 viaA-YFP and MG1655 viaA-6Gly-YFP are tagged in the same fashion except that the latter has a 6 Glycine linker (indicated in bracket) upstream of YFP. B) The PCR results are shown with the expected size of the products indicated with colored arrows. Arrows are colored according to the expected sizes as labeled in A). Brown arrow indicates the 587-bp product. Black arrow indicates the 2335-bp product, and blue arrow indicates the 1365-bp product.
Figure 27: Verification of MG1655 viaA-(6Gly)-YFP strains.

A) Western blot for the expression of ViaA in the following five strains: MG1655 (WT), MG1655 ravA-mRFP, MG1655 ravA-6Gly-mRFP, MG1655 viaA-YFP, MG1655 viaA-6Gly-YFP. These strains were grown anaerobically either to mid-log phase (after 5 hours) or stationary phase (24 hrs) in LB at 37°C. Purified ViaA is included as reference. ViaA are detected on a 9% SDS-PAGE gel with α-ViaA antibody. The expected molecular weight (MW) is 55 kDa for ViaA, 82 kDa for ViaA-YFP proteins. The expected size for YFP is approximately 27 kDa. B) Fluorescence images for ViaA-YFP. The strains were grown in LB at 37°C for 5 hrs anaerobically. Images were taken with by confocal microscope (Zeiss LSM510).
3.4 Discussion

All five genes, ldcI, cadB, cadC, ravA and viaA, were tagged successfully with the FPs in the E. coli genome.

LdcI fusion proteins are insoluble and aggregate to form inclusion bodies, which were shown as green foci by fluorescence imaging (Figure 19B). In Lindner et al., they proposed that inclusion bodies are always inherited by the old pole after each cell division (Figure 28) [205]. Most of the time, only one inclusion body is formed which initially localizes either at one pole, in the middle or at quarter-position of an E. coli cell. However, as shown in Figure 28, since the foci do not move once they are formed, the location of an aggregate after division is determined by its location in the mother cell. As a result, after a maximum of three generations, the inclusion bodies will always be found at the poles of an E. coli cell regardless of the initial position of aggregates formation. Indeed, this is what has been observed for LdcI fusion proteins, that the great majority of cells have one green focus at one pole of a cell, while a number of them have the focus at the middle- or quarter-position (Figure 19B).

Classically, inclusion bodies have been considered as compact non-functional and misfolded protein aggregates such as under heat shock at 42°C. However, recently, it has been found that non-classical inclusion bodies also exist [206-209]. These inclusion bodies are formed by the overexpression of tagged proteins grown at non-heat-shock conditions. They are found to retain 30~100% WT protein activities. This explains the observation that the LdcI-sfGFP proteins still increased the pH of the minimal media post-shift in the media shift assay (Figure 19C). In addition, the LdcI fusion proteins have the same activity as WT LdcI. This is either because the media shift assay is not sensitive enough to detect the reduced activity of LdcI in the inclusion bodies, or there are other rate-limiting factors for the lysine decarboxylase pathway in vivo such as the saturation of CadB.

Strains where CadB has been tagged with mRFP displayed fluorescence along the membrane upon acid induction. This observation conforms to the fluorescence pattern of a CadB-overexpression E. coli strain [210]. This is expected as CadB is abundantly induced at pH 5 in the presence of lysine.
CadC-sfGFP was not observed using fluorescence microscopy, which is likely because the endogenous copy number for CadC is very low (~20 copies per cell) [56]. Furthermore, the activity of the lysine decarboxylase pathway was slightly lower in the CadC-sfGFP strains than in the WT. The sfGFP tag might have affected the folding and/or transport of the periplasmic domain of CadC into the periplasm.

For RavA and ViaA, it has been quantified in our lab that the copy number of RavA is approximately 300 and that of ViaA is approximately 100 under aerobic conditions in stationary phase (personal communication with K.W. in Dr. Walid Houry’s lab). The numbers increased approximately ten fold under anaerobic growth. By comparison, the copy number of LdcI under acid stress is approximately between 7000 and 10000 under acid stress (Figure 13B) [2]. Therefore, combined with the cytoplasmic dispersion, the lower expression level of RavA and ViaA may contribute to the absence of fluorescent signals.
Figure 28: Individual aggregates are located to the old pole through cycles of cell divisions.

Because movement of foci is rarely observed, the location of an aggregate within the cell after division is determined by its location in the mother cell. Those that were at the one- or three-quarter positions are found concentrated around the mid-cell point after division (Step I). Aggregates at the mid-cell are subsequently located in the new pole, with equal probability to be in either of the two cells (Step II). Those that are found in the new-pole end of the cell immediately before division remain in the same pole; however, that pole, having been formed in the previous division event, is now an old pole in the offspring cell (Step III). Note that aggregates can be initially detected at polar, mid-cell, or quarter-cell positions but are eventually located to an old pole. Once there, they are consistently inherited by the old-pole cell after division. Aggregates are indicated by green dots. Red cell ends are old poles, and blue cell ends are new poles [205].

Note: This figure is reprinted from [205] © 2008 by The National Academy of Sciences of the USA.
4 Future Directions

In Chapter 2 of the thesis, I showed that the stringent response signaling molecule, ppGpp, inhibits LdcI in the lysine-dependent acid resistance pathway in *E. coli*. At the same time, this inhibition is reduced if LdcI first interacts with the MoxR-family AAA+ ATPase RavA.

Therefore, it is necessary to understand next the mechanisms behind the antagonistic regulation. To do this, the residues required for the interaction between LdcI and RavA will be mapped. Potential candidates for the residues will be obtained from docking the LdcI and RavA crystal structures onto the EM cage structure (Figure 12 B). In addition, it remains unknown the role of ViaA in the regulation of LdcI by RavA, as ViaA has been shown to interact with RavA [63]. Currently, the exact functions of RavA and ViaA are under investigation in our lab. Together with the proposed experiments, they will further our understanding of the function of the LdcI-RavA cage and its physiological significance in *E. coli*.

In Chapter 3, the components of the lysine-dependent acid resistance pathway, RavA and ViaA have been successfully tagged with different FPs to study their localization behaviour.

The LdcI-sfGFP protein was shown to forms aggregates. However, we cannot tag it at the N-terminus as the N-terminus is involved in LdcI decamerization. One option to solve this problem is to tag LdcI instead with mRFP or mRFP derivatives such as mCherry. The CFP or YFPs are not good alternatives because they are derived from GFP and will likely form aggregates. Another way to solve this problem is to perform an immunofluorescence experiment using a secondary antibody conjugated with Cy3 or FITC for fluorescence imaging of LdcI. This experiment will be used to monitor the potential re-localization of LdcI from cytoplasm to the cell membrane upon acid stress.

Furthermore, once the immunofluorescence protocol is established, the localization behaviour of LdcI will be followed in a number of conditions. First, we will observe it upon acid stress recovery where cells will be grown first at pH 5 and then shifted to pH 7. Secondly, as RavA is more abundantly expressed in anaerobic conditions, we will follow the behavior of LdcI in both WT, ΔravA, ΔravAΔviaA strains under anaerobic conditions. We will also monitor the behaviour of LdcI in the presence of ppGpp upon starvation. Thirdly, the immunofluorescence imaging of
LdcI will be used on the MG1655 cadB-mRFP strains to observe the localization of LdcI and CadB simultaneously under acid stress.

As tagging of CadC at its C-terminus with sfGFP reduces the activity of the lysine-dependent acid resistance pathway, we can try to tag CadC at its N-terminus which is in the cytoplasm. However, as the N-terminus of CadC undergoes a conformational change upon acid induction to bind DNA, whether tagging it to GFP affects CadC activity remains to be tested.

Lastly, no fluorescence has been observed with the RavA and ViaA tagged strains. One way to solve this problem is to try to tag them with CFP and BFP which are more stable and show stronger fluorescence signals. Alternatively, the immunofluorescence method can also be applied to detect RavA and ViaA under acid stress, during stationary phase in both aerobic and anaerobic conditions. Co-localization of LdcI, RavA and ViaA will also be monitored using secondary antibodies with different fluorophores in the immunofluorescence imaging experiment. These experiments will reveal the timing and behaviour of LdcI, RavA and ViaA, which will advance significantly our understanding of the function of RavA, ViaA and the RavA-LdcI cage.
References


