CHARACTERIZATION OF CYCLIC-GMP-AMP SIGNALING AND NOVEL GENE NETWORKS IN THE *VIBRIO* SEVENTH PANDEMIC ISLANDS OF EL TOR *VIBRIO CHOLERAE*

By

Geoffrey Blaine-Graessley Severin

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biochemistry and Molecular Biology– Doctor of Philosophy

2020

ABSTRACT

CHARACTERIZATION OF CYCLIC-GMP-AMP SIGNALING AND NOVEL GENE NETWORKS IN THE *VIBRIO* SEVENTH PANDEMIC ISLANDS OF EL TOR *VIBRIO CHOLERAE*

By

Geoffrey Blaine-Graessley Severin

Vibrio cholerae is the causative agent of the diarrheal disease cholera, for which there have been seven pandemics. The first six pandemics (1817-1923) have been attributed to strains of the classical biotype while the seventh pandemic (1961- current) is being perpetuated by circulating strains of the El Tor biotype. It is hypothesized that El Tor's acquisition of two unique genomic islands of unknown origins, Vibrio Seventh Pandemic Island 1 & 2 (VSP-1 & 2), potentiated its displacement of classical strains in both environmental and clinical reservoirs prior to the seventh pandemic. Despite their connection to the pandemic evolution of the El Tor biotype and the likelihood they encode a wealth of novel biological functions the majority of the ~ 36 ORFs in the VSP islands have remained uncharacterized. The works presented in the thesis represent examples of our collective efforts to understand the unique traits afforded to the El Tor biotype by the VSP islands.

In 2012, ten years after the initial discovery of the VSP islands, the cyclic di-nucleotide synthase DncV was identified in VSP-1 and shown to synthesize the novel second messenger cyclic-GMP-AMP (cGAMP). Despite this significant discovery, cGAMP remained an orphan second messenger in bacteria for six years. In Chapter 2, I present our identification of the first cGAMP signaling network in bacteria by connecting the synthesis of cGAMP to the allosteric activation of the VSP-1 encoded patatin-like phospholipase *vc0178*, now named CapV. This finding helped catalyze the revelation that homologous cyclic-dinucleotide signaling networks

contribute to bacteriophage-immunity and are distributed widely across the bacterial phyla.

dncV and *capV* represent just two of the ~ 11 ORFs encoded in VSP-1, therefore we hypothesized that cGAMP signaling may extend to include additional genes within the island. Bioinformatic analysis of VSP island genes performed in collaboration with the laboratory of Eva Top at the University of Idaho predicted that *dncV* was likely to share a biological pathway with the VSP-1 encoded putative deoxycytidylate deaminase *vc0175*, we have now named DcdV. In Chapter 3, I describe our ongoing efforts to understand the biological function of DcdV and the discovery of its novel negative regulator DifV.

The work presented in this thesis has made fundamental contribution to understanding the pandemic evolution of the El Tor biotype and the utility homologous genes afford to their bacterial host while providing a roadmap for further exploration of the biological role of the VSP-1 and VSP-2 islands in pandemic *V. cholerae*.

This thesis is dedicated to my wife, Christina, and children, Allie and Bobo.
I love you.

ACKNOWLEDGEMENTS

Thank you, Christopher Waters Ph.D. for being the best boss I've ever had. You inspire people to be their best; for themselves and for others. I would also like to thank my guidance committee; George Mias Ph.D., Lee Kroos, Ph.D., Beronda Montgomery Ph.D., and Christoph Benning Ph.D for their thoughtful ideas and critiques. Additionally, I have had the pleasure of performing research under the guidance of both Beronda and Christoph, and I am extremely grateful for those experiences.

I have been fortunate to work with amazing collaborators during my time at MSU. I owe a special thank you to Wai-Leung Ng Ph.D., Miriam Ramliden, Brendan O'Hara, and Lisa Hawver Ph.D. from Tufts University, Atul Khataokar Ph.D. and Matthew Neiditch Ph.D. from Rutgers University, Clinton Elg and Eva Top Ph.D. from the University of Idaho, and Kenny Wang Ph.D. from Christoph's lab here at MSU. I have enjoyed sharing ideas, failures, and successes with you and I'm looking forward to seeing all the cool stories to come.

Thank you, Lijun Chen Ph.D. and Tony Schillmiller Ph.D. from the Mass Spectrometry Core at MSU. Over the last 7 years you have been incredibly generous with your time; graciously training (and retraining) me and troubleshooting countless technical issues, most of which could be immediately traced back to me.

Since 2013, 5128 BPS has been my research home and the members of the Waters Lab have been my research family. One special feature of this family (from past, present and no doubt future members) is the shared investment each person has in one another's success. Here I've been fortunate to work with wonderful trainees, all of whom have made tremendous contributions to my understanding of research, science communication, and friendship. Of these mentees I want to especially thank Christopher "Rhoadie" Rhoades and Macy Pell both of whom

toiled away, for years, exploring my half-baked ideas. We have learned an incredible amount about the VSP islands from their ability to dust themselves off and get back on the hypothesis horse time and time again.

Two of my graduate student colleagues also deserve special recognition; Brian “Bird” Hsueh and Nico “Bert” Fernandez Ph.D. Thank you, Brian for your cheerful attitude, dependability, kindness, and reigning in my meandering mind. Nico, you are wise for your years, incredibly focused, organized, kind, and generous. I’m very thankful that I got to sit and share ideas with the two of you day after day.

One of the greatest benefits of returning to Michigan was being close to our families. Thank you in particular to my in-laws; Julia and Peter Munoz, my parents; Blaine and Kathryn Severin, and brother and sister in-law; Gregory Severin and Regan Quick-Severin. Your support and advice have been priceless and I look forward to the day when we can return to our old routines and Sunday dinners!

Thank you to my children; Allie and Bobo. You guys keep life in perspective.

Finally, I owe any and all of my success to my wife, Christina. Thank you for encouraging me to become a scientist and making countless sacrifices to make it possible.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES x

KEY TO ABBREVIATIONS xii

Chapter 1: Introduction 1
1.1: Preface 2
1.2: The Evolution of El Tor *Vibrio cholerae* and the Seventh Cholera Pandemic 2
1.3: Non-cGAMP cdN signaling in bacteria 5
1.4: Discovery of bacterial cGAMP 6
1.5: The functions of cGAMP in *V. cholerae* biology 8
1.6: Transcriptional and Allosteric Regulation of DncV and CapV 9
1.7: Bacterial cGAMP signaling outside of *Vibrio* 11
1.8: Definition of the CD-NTase family of oligonucleotide synthases 12
1.9: cGAMP signaling in metazoans 15
1.10: CD-NTase family of Enzymes and Bacteriophage Defense 16
1.11: The expanding role of DncV beyond CBASS 20
1.12: Expanding the known functions of VSP-1 in *V. cholerae* 22

Chapter 2: Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae* 23
2.1: Preface 24
2.2: Chapter Body 24

Chapter 3: Identification of DcdV, a VSP-1 encoded deoxycytidylate deaminase, and its allosteric regulator DifV 25
3.1: Preface 26
3.2: Abstract 26
3.3: Introduction 27
3.4: Materials and Methods 32
3.4.1: Bacterial Strains, Plasmids, and Growth Conditions. 32
3.4.2: Growth Curve Assays. 36
3.4.3: Fluorescence Microscopy and Analysis. 36
3.4.4: In-vitro Nucleic Acid Deamination Assay 36
3.4.5: Protein Stability Assay 38
3.4.6: Determination of Uracil in DNA 39
3.5: Results 39
3.5.1: VSP-1 Island Gene Correlogy Study Indicates *dncV* and *dcdV* may share a functional connection 39
3.5.2: Ectopic Expression of *dcdV* Leads to Conditional Cell Filamentation 44
3.5.3: DifV is encoded immediately 5' of the *dcdV* locus in VSP-1 47
3.5.4: Elucidating the Mechanism of DifV Mediated DcdV Inhibition 51
3.5.5: DcdV induced filamentation requires both conserved NK-family features and the CDA active site 53

3.5.6: The CDA domain of DcdV utilizes both dCMP and dCTP substrates and may not be allosterically regulated by dTTP 58
3.5.7: The DCD domain increases the genomic incorporation of dUMP 59
3.6: Discussion 62

Chapter 4: Concluding remarks 68
4.1: Conclusion and Significance 69
4.1.1: DncV – cGAMP – CapV 69
4.1.2: DcdV and DifV 70
4.1.3: Correlogy and VSP island gene networks 71
4.1.4: Origins and exclusivity of the VSP islands 71
4.2: Future Directions 72
4.2.1: Further characterization of El Tor *V. cholerae* CBASS 72
4.2.2: Characterization of the *dncV* - *dcdV* gene network 76

BIBLIOGRAPHY 79

LIST OF TABLES

Table 3.1: Strains and descriptions 34

Table 3.2: Plasmids and descriptions 35

LIST OF FIGURES

Figure 1.1: Schematic of the El Tor *V. cholerae* N16961 VSP Islands. 4

Figure 1.2: Summary of cdN synthesizing enzymes, products, and function. 6

Figure 1.3: Cartoon depiction of the homologous *capV* and *dncV*-like operons found in the mammalian commensal strain of *E. coli ECOR31*. 11

Figure 1.4: Expansion of the Cyclic-oligonucleotide Lexicon. 13

Figure 3.1: Cytidine deamination reaction and simplified de novo thymidine triphosphate biosynthesis pathway. 30

Figure 3.2: VSP-1 schematic and predicted gene networks (MRS). 41

Figure 3.3: VSP-2 schematic and predicted gene networks (MRS). 42

Figure 3.4: Ectopic expression of *dcdV* leads to cell filamentation in the absence of VSP-1 in El Tor *V. cholerae*. 45

Figure 3.5: Ectopic expression of *dcdV* leads to cell filamentation in *E. coli* that is alleviated by addition of a single copy cosmid containing VSP-1. 46

Figure 3.6: The negative regulator of DcdV induced filamentation, DifV, is encoded 5' of *dcdV* within *orf1* 48

Figure 3.7: DcdV induced filamentation is not controlled by Quorum Sensing. 50

Figure 3.8: DifV inhibition of DcdV induced filamentation is unlikely to occur at the level of translation or protein stability. 51

Figure 3.9: Phyre2 model of DcdV and conserved features of the N-terminal NK domain and C-Terminal DCD domain. 54

Figure 3.10: Conserved features of the N-terminal NK domain and C-Terminal DCD domain are required for DcdV induced filamentation in *E. coli*. 55

Figure 3.11: *E. coli* lysates containing DcdV preferentially deaminate dCMP and dCTP substrates. 56

Figure 3.12: dTTP does not inhibit DcdV deamination activity in *E. coli* lysates. 57

Figure 3.13: Ung activity is not required for DcdV induced filamentation in *E. coli*. 59

Figure 3.14: Flow chart for detecting incorporation of dU in Δ*ung E. coli* genomic DNA. 60

Figure 3.15: DcdV activity increases genomic dU incorporation in Δ*ung E. coli*. 61

Figure 3.16: Model for hypothetical DcdV activity in phage immunity. 65

KEY TO ABBREVIATIONS

10-MTHF	10-Formyltetrahydrofolate
5,10-MTHF	5,10-Methylenetetrahydrofolate
5-MTHF	5-Methyltetrahydrofolate
ABS	absorbance
ATP	adenosine triphosphate
BER	Base Excision Repair
BSA	bovine serum albumin
cAAG	cyclic-AMP-AMP-GMP
CBASS	cyclic-oligonucleotide-based antiphage signaling system
CDA	cytosine deaminase
c-di-AMP	cyclic di-AMP
c-di-GMP	cyclic di-GMP
cdN	cyclic di-nucleotide
cGAMP	cyclic-GMP-AMP
DAC	diadenylate cyclase
DCD	deoxycytidylate deaminase
dCMP	deoxycytidine monophosphate
dCTP	deoxycytidine triphosphate
DGC	diguanylate cyclase
dTMP	thymidine monophosphate
dTTP	thymidine triphosphate
dUMP	deoxyuridine monophosphate
dUTP	deoxyuridine triphosphate

GMP	guanosine monophosphate
GTP	guanosine triphosphate
HCD	high-cell density
iNOS	inducible nitric oxide system
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LCD	low-cell density
MSHA	mannose-sensitive hemagglutinin
NK	nucleoside/nucleotide kinase
NMP	nucleotide monophosphate
NTP	nucleotide triphosphate
OD	optical density
ORF	open reading frame
PBS	phosphate buffered saline
PDE	phosphodiesterase
QS	quorum sensing
RBS	ribosome binding site
TLD	thymine-less death
UDG	uracil-DNA glycosylase
UNG	uracil-N-glycosylase
VSP	*Vibrio* Seventh Pandemic Island
WT	Wild-type

Chapter 1: Introduction

1.1: Preface

Contents of this chapter were published in Cell Host & Microbe in 2019 (Severin, G. B. & Waters, C. M. Pyrimidines and Cyclic Trinucleotides Join the Second Messenger Symphony. *Cell Host Microbe* **25**, 471–473 (2019)). Per the publisher Cell Press, "As an author, you (or your employer or institution) may do the following: Include the article in full or in part in a thesis or dissertation (provided that this is not to be published commercially)."

Contents of this chapter were published in "Microbial Cyclic Di-Nucleotide Signaling" in 2020 (Ramliden, M. S., Severin, G. B., O'Hara, B. J., Waters, C. M., Ng, WL. (2020) Microbial Cyclic GMP-AMP Signaling Pathways. In: Chou SH., Guiliani N., Lee V., Römling U. (eds) Springer, Cham). Per the publisher Springer, "Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis."

1.2: The Evolution of El Tor *Vibrio cholerae* and the Seventh Cholera Pandemic

The Gram-negative γ-proteobacterium *Vibrio cholerae* is the causative agent of the diarrheal disease cholera and is estimated to afflict ~2.9 million individuals resulting in ~95,000 deaths annually [1]. Since 1821, there have been seven recorded cholera pandemics which have been caused by two *V. cholerae* biotypes; classical and El Tor. While the first six pandemics were caused by classical *V. cholerae* strains, the current (7^{th}) pandemic, which began in 1961, is being perpetuated by circulating strains of El Tor *V. cholerae*. The seventh has lasted longer than any of the previous pandemics and has recently afflicted vulnerable populations across the globe including Yemen, Haiti, and Mozambique due to poor sanitation resulting from armed conflict and natural disasters [2]. Despite both biotypes sharing a common ancestor, the El Tor biotype appears to have globally displaced the classical biotype in both endemic and clinical populations

as the etiological agent driving modern day cholera [3]. El Tor's mysterious supplantation of the classical biotype has been a curiosity in the cholera field and this global "evolution in action" event provides a unique opportunity for exploring pandemic bacterial evolution.

In 2002, Dziejman et al. constructed a genomic microarray based on the recently published genome of the El Tor *V. cholerae* strain, N16961, which they used to make genomic comparisons between a variety of *V. cholerae* isolates [4]. Interestingly, they found that <1% of the gene content represented in the microarray (which accounted for >93% of the known N16961 genes) differed between the El Tor and classical strains tested. The greatest difference between the biotypes appeared to be El Tor's acquisition of two unique genomic islands which spanned loci *vc0175-vc0185* and *vc0490-vc0497* in the N16961 genome that were named the **V**ibrio **S**eventh **P**andemic Islands **1** and **2** (VSP 1 & 2) (Figure 1.1), respectively [4]. The boundaries of the canonical El Tor VSP-2 were later expanded to cover the region containing *vc0490-vc0516* based on PCR interrogation of the island boundaries [5]. Interestingly, VSP-2 appears to be undergoing intense selective pressures as variant lineages are occasionally found to have gene deletions, additions, and rearrangements [6–8].

Through genomic analysis of unique clinical and environmental isolates of *V. cholerae* from the past hundred years, Hu et al. [3] postulates that the pandemic evolution of El Tor, from a nonpathogenic bacterium (1900's) to global scourge (1961), involved six critical events. While the first four events included formulaic acquisitions of pathogenicity islands and virulence factors similar to those found in the classical biotype, the 5th stage (1925-1954) was marked by El Tor's acquisition, from an unknown origin,VSP-1 and 2 [4]. The 6th stage (1954-1960) is noteworthy for both a cessation in genomic rearrangements and horizontal gene transfer and the accumulation of 12 novel SNPs, which have no obvious relevance to pathogenicity. For this

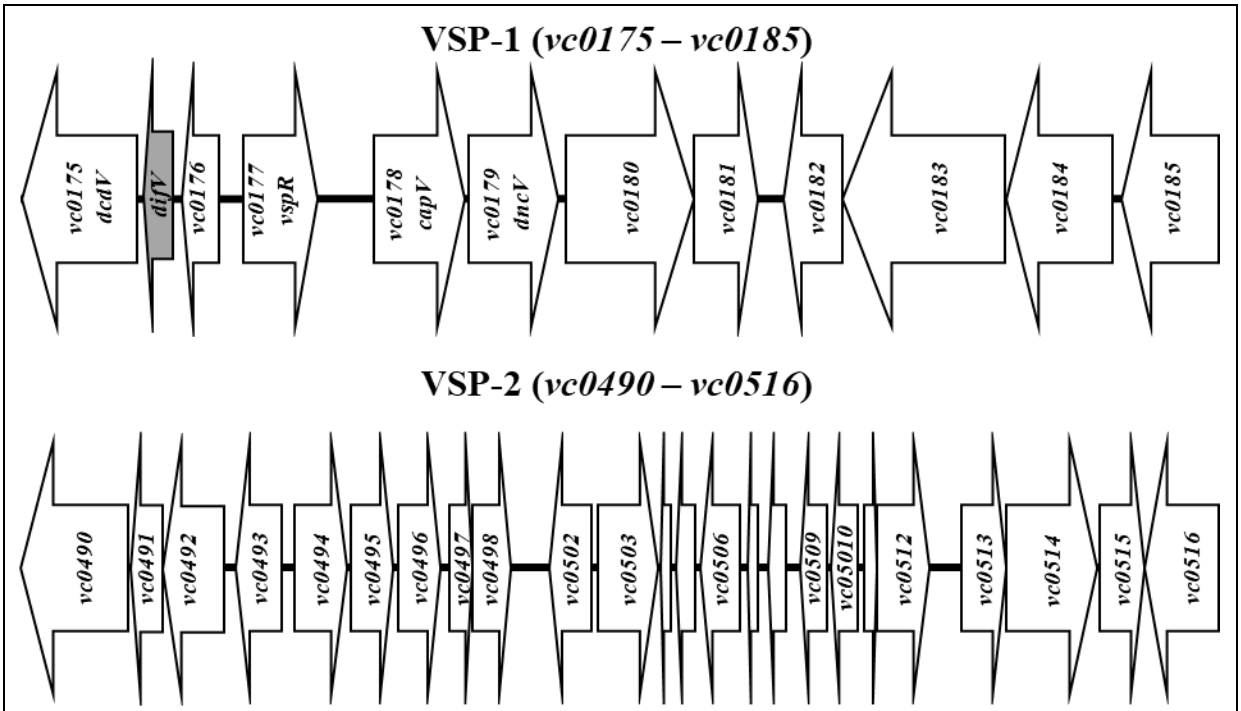

Figure 1.1: Schematic of the El Tor *V. cholerae* N16961 VSP Islands.
Schematic depiction of VSP-1 and VSP-2 based on the annotated El Tor V. cholerae N16961 genome. VSP-1 is ~14 kb in length covering *vc0175 – vc0185*. The previously unannotated VSP-1 ORF *difV* is labeled in gray.VSP-2 is ~25 kb in length covering *vc0490 – vc0516*. Not to scale.

reason, it is hypothesized that the acquisition of VSP-1 and 2, which represent the largest genetic differences between the two biotypes [4], is responsible for potentiating the pandemic evolution of El Tor and displacement of classical *V. cholerae*. However, the fitness advantage realized by El Tor through maintenance of these islands is just beginning to be understood.

Despite their perceived relevance to El Tor's rise in global prominence, VSP-1 (~14 kb, 11 ORFs) and VSP-2 (~25 kb, 24 ORFs) remain largely uncharacterized (Figure 1.1). Primary sequence and modeling predictions suggest that these genes encode a diverse array of functions (transcriptional regulators, c-di-GMP metabolizing enzymes, pilins, chitinases, etc.) all of which may have profound effects on bacterial fitness. Additionally, partial and intact homologous islands are present in other bacteria of clinical and environmental origins suggesting that the

utility of maintaining these genes is not limited to El Tor *V. cholerae* but broadly applicable to prokaryotes generally. Uncovering the function of the genes on VSP-1 and VSP-2 is a central to our understanding the evolution of *V. cholerae* pandemics. Interestingly, one of the first concrete functions shown to be encoded on VSP-1 was synthesis of the novel cyclic di-nucleotide (cdN) 3',3'-cyclic GMP-AMP (cGAMP) [9].

1.3: Non-cGAMP cdN signaling in bacteria

In order to understand the novelty of cGAMP and other cdNs to be discovered later, I will first provide a broader perspective of cdN signaling in bacteria. Bacteria use cdNs as second messengers to regulate various cellular processes at the level of transcription,translation, and protein activity [10–13]. In 1987, the first cdN, c-di-GMP, was identified in the lab of Dr. Moshe Benziman while investigating the regulation of cellulose biosynthesis in *Gluconacetobacter xylinus* (previously referred to as *Acetobacter xylinum*) [14]. It would be another decade before it was recognized that c-di-GMP is nearly ubiquitously used in bacteria to regulate diverse behaviors such as motility, biofilm formation, stress response, and cell cycle progression, often within the same organism [10, 12, 13] (Figure 1.2). C-di-GMP is synthesized by diguanylate cyclase (DGC) enzymes possessing a GGDEF domain and degraded by phophodiesterases (PDE) possessing either an EAL or HD-GYP domain. The second cdN, c-di-AMP, was first identified in 2008 as a product of the sporulation checkpoint protein DisA in *Bacillus subtilis* and its *Thermotoga maritima* homolog, and the family of c-di-AMP synthases is referred to as diadenylate cyclases (DAC) [15]. c-di-AMP is now known to play a critical role in sensing DNA damage, cell wall homeostasis, and adaptation to changes in osmolarity [11] (Figure 1.2). While c-di-AMP signaling networks may be less frequently distributed, they are often essential in

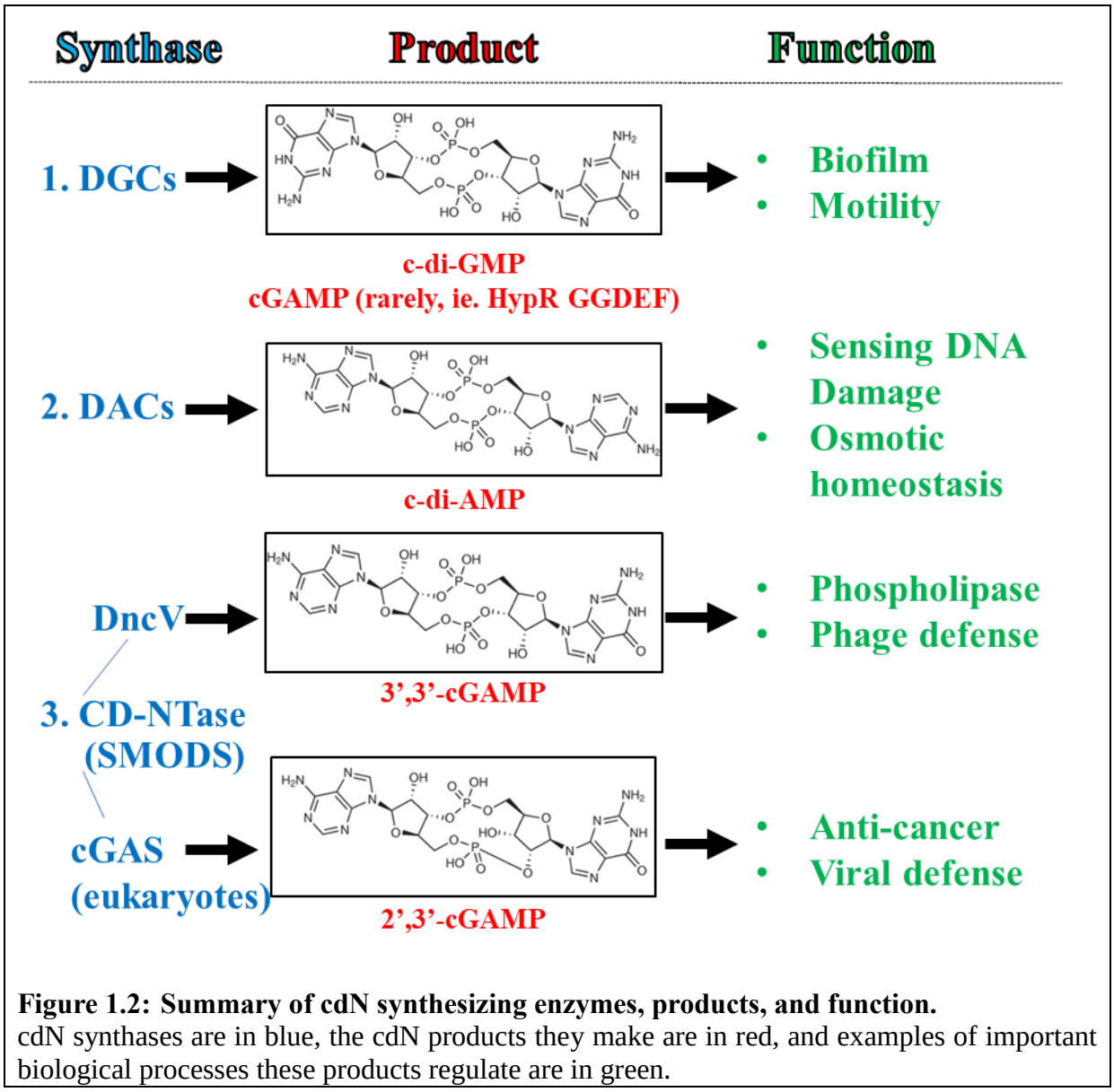

Figure 1.2: Summary of cdN synthesizing enzymes, products, and function.
cdN synthases are in blue, the cdN products they make are in red, and examples of important biological processes these products regulate are in green.

Gram-positive bacteria. cGAMP, identified in 2012 and described below, represents the third cdN to be characterized.

1.4: Discovery of bacterial cGAMP

In 2012, while dissecting the virulence regulon of El Tor *V. cholerae*, Davies et al. found an interesting regulatory network leading to VSP-1 involving ToxT, the master regulator of virulence, the sRNA TarB, and the previously uncharacterized VSP-1 encoded transcriptional

regulator *vc0177* (named *vspR* for ***V****ibrio* **S**eventh **P**andemic **r**egulator). The pro-virulence activity of ToxT leads to the transcription of TarB, which in turn represses the production of VspR [9]. Deletion of *vspR* resulted in increased expression of a number of adjacent genes in VSP-1 including *vc0179* (renamed *dncV* for dinucleotide cyclase *Vibrio*). Together, these data informed a model that suggested ToxT, through inhibition of VspR production, induces the expression of genes in VSP-1 including *dncV* during *V. cholerae* host colonization.

DncV was predicted to be an enzyme belonging to the nucleotidyl transferase superfamily of enzymes that transfer nucleotide monophosphates to a hydroxyl group found on a diverse array of biomolecules [16]. Hypothesizing DncV would utilize nucleotide substrates, Davies et al. incubated purified DncV protein with individual and mixed cocktails of nucleotide triphosphates and analyzed the reaction products using thin-layer chromatography. Incredibly, DncV was shown to synthesize both of the known bacterial cyclic-di-nucleotide (cdN) second messengers cyclic-di-GMP (c-di-GMP) and cyclic di-AMP (c-di-AMP) when provided only GTP or ATP. This result identified DncV as the first non-canonical DGC or DAC to synthesize c-di-GMP or c-di-AMP, respectively, as well as the first c-di-AMP synthase identified in a Gram-negative bacterium. Even more remarkably, when provided with a substrate cocktail containing equimolar concentrations of ATP and GTP, DncV produced little c-di-AMP or c-di-GMP, but instead preferentially synthesized the novel hybrid cdN 3',3'-cyclic GMP-AMP (cGAMP) (Figure 1.2) [9].

The initial implications of cGAMP's discovery were limited to *V. cholerae,* but it was soon discovered that some species of δ-proteobacteria catalyze cGAMP synthesis from canonical DGC c-di-GMP synthesizing enzymes called hybrid, promiscuous (Hypr) GGDEF proteins (Figure 1.2) [17]. However, it is important to understand that Hypr-GGDEF synthases are not

homologous to DncV, and thus they only have functional similarity. Efforts to understand the role of cGAMP in *V. cholerae* laid foundational work to a paradigm shift in cdN signaling, discussed later in this chapter, revealing that c-di-GMP, c-di-AMP and cGAMP were just the tip of the cdN iceberg.

1.5: The functions of cGAMP in *V. cholerae* biology

As mentioned, *dncV* is encoded at locus *vc0179* in the VSP-1 island of El Tor *V. cholerae* isolates responsible for the current cholera pandemic and absent in the classical biotype that predominated in the previous six pandemics (Figure 1.1). RNA-Seq of El Tor cultures following 15 min of ectopic DncV expression revealed nearly 90 genes were differentially regulated by at least 2-fold [9]. This scale of transcriptomic influence suggested cGAMP was a global transcriptional regulator, similar to other cdNs, and these genes could be grouped into three major biological functions: shifting fatty acid anabolism to catabolism, down regulation of MSHA pilus, and reduction in chemotaxis gene expression [9]. Additionally, increased expression of *dncV* also limited *V. cholerae* chemotaxis through low-density agar, even though individual cells appeared to remain flagellated and motile [9].

As discussed in more depth in Chapter 2, only one direct protein target of cGAMP has been identified so far in bacteria, which we discovered in El Tor *V. cholerae* [18]. We found that in the El Tor biotype of *V. cholerae,* overproduction of cGAMP is toxic, leading to growth inhibition, and eventually cell death [18]. Genetic suppressor analysis determined that this cGAMP toxicity is dependent on the phospholipase CapV, encoded at the locus *vc0178* directly upstream of *dncV* on VSP-1 [18]. We found that cGAMP binds directly to CapV and activates its phospholipase activity, resulting in degradation of cell membrane phospholipids and release of free fatty acids [18]. Activation of CapV phospholipase activity *in vitro* is specific to cGAMP as

other related dinucleotides required up to one thousand-fold higher concentrations in order to activate it [18]. Once activated, CapV degrades phosphatidylethanolamine and phosphatidylglycerol, the two most common classes of phospholipid in the *V. cholerae* cell membrane, and releases fatty acids into the cytosol [18]. Such sustained activation or over-activation of CapV by high intracellular levels of cGAMP ultimately leads to membrane damage, growth arrest, and cell death [18]. Thus, we found that CapV and DncV form the first functional cGAMP signaling pathway discovered in bacteria, and this pathway regulates membrane lipid degradation (Figure 1.2). Additionally, unpublished data from our laboratory and collaborators suggests that changes in global transcription previously associated with *dncV* activity, including virulence and decreased chemotaxis, and impacts of *dncV* on bacterial motility may primarily be due to pleotropic effects resulting from CapV-dependent membrane stress (W. Ng, personal communication). In contrast to c-di-GMP in *V. cholerae*, which encodes dozens of enzymes responsible for the synthesis and degradation of c-di-GMP [19] and multiple effectors including transcriptional regulators, riboswitches, and protein complexes, our discovery of CapV as the predominant receptor for cGAMP suggests the network is discrete, insulated, and targeted to allosterically regulate a select number of effectors proteins rather than global transcriptional regulons.

1.6: Transcriptional and Allosteric Regulation of DncV and CapV

Like DncV, CapV expression is controlled in part by the repressor VspR and is therefore hypothesized to be induced by the virulence regulator ToxT [9], though this regulation has not been directly shown. However, another *V. cholerae* virulence regulator, ToxR, was also found to bind DNA upstream of *capV*, thereby repressing its transcription [20]. It is not surprising that the cGAMP-CapV pathway is highly regulated, as over activation of the pathway results in cell

death [18]. However, it is not known how these two regulatory systems interact to control CapV expression, nor is it known when the cGAMP-CapV pathway is induced in the cell.

Quorum Sensing (QS) plays an important role in the regulation of *dncV* and *capV* expression. QS is a process by which bacteria monitor inter- and intra-species population densities by secreting and detecting molecules called autoinducers. The concentrations of autoinducers in the environment help dictate whether an individual bacterium adopts a gene expression profile that is better suited to existence as an individual, called low-cell density (LCD), or one that facilitates participation in group behaviors that are beneficial to life in a dense bacterial population, called high-cell density (HCD). In *Vibrios,* QS controls the expression of around 10% of the genome [21] and unpublished observations demonstrate that the HCD QS state leads to the expression of *dncV* and *capV* (W-L. Ng, personal communication). Being included in the HCD regulon suggests that DncV and CapV are present as *V. cholerae* transitions to and participates in an environment where individual bacterium behave, to some extent, in altruistic behaviors [22].

As described, the catalytic activity of CapV is finely tuned to the presence of cGAMP [18]. Additionally, DncV has been co-crystallized with folates derived from the cell, and both 5-methyltetrahydrofolic acid (5-MTHF) and 5-methyltetrahydrofolate diglutamate can inhibit DncV activity [23]. This inhibition is attributed to changes in the active site conformation, preventing DncV from synthesizing cGAMP. While the purpose of this regulation by folates and its relevance in *V. cholerae* is not understood, when considered in the context of QS it is likely DncV and CapV are present at HCD and the cellular abundance of folate species dictates, in part,

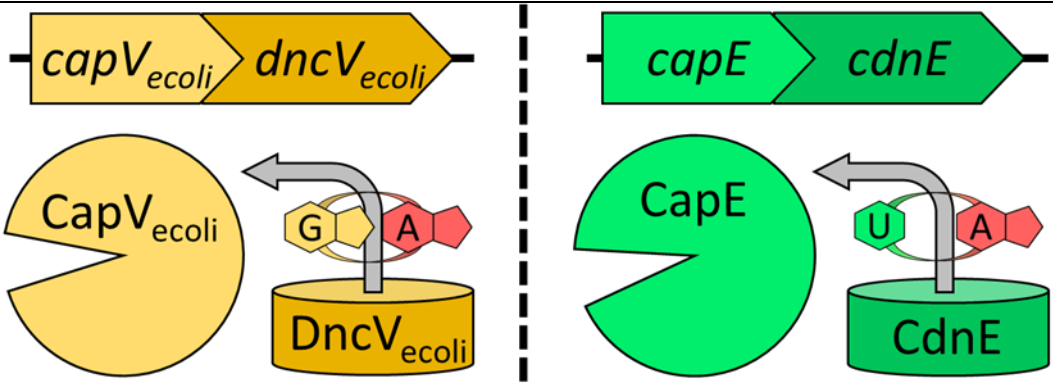

Figure 1.3: Cartoon depiction of the homologous *capV* and *dncV*-like operons found in the mammalian commensal strain of *E. coli ECOR31*.
Left: The more conserved pair of enzymes, $dncV_{ecoli}$ and $capV_{ecoli}$, predominantly synthesize and respond to cyclic-GMP-AMP, respectively. *Right:* The more degenerate pair, *cdnE* and *capE*, predominantly synthesize and respond to cyclic-UMP-AMP, respectively.

the catalytic activity of DncV, and thus activation of CapV.

1.7: Bacterial cGAMP signaling outside of *Vibrio*

The Hypr-GGDEF found in some δ-proteobacteria species have been characterized to synthesize cGAMP for the regulation of riboswitches best characterized in *Geobacter metallireducens* and *Geobacter sulfurreducens* [17, 24] (Figure 1.2). In these species, ligand-free riboswitches contain a terminator loop that prevents transcription of nearby downstream genes. When cGAMP is bound, the terminator structure is disrupted allowing transcription of downstream genes. *G. metallireducens* contains at least 11 riboswitches that specifically bind cGAMP to control transcription of 17 genes. Prominent among the genes regulated by cGAMP in *Geobacter spp.* are pilins and cytochromes associated with exoelectrogenesis which can enable the reduction of insoluble metal complexes to act as electron acceptors [17, 24]. In these bacteria, cGAMP and c-di-GMP control opposing energy generating strategies with cGAMP promoting transient association with iron particles while c-di-GMP stimulates biofilm formation on larger metal surfaces [25]. Therefore, in addition to regulation of phospholipid degradation in

V. cholerae, cGAMP has been suggested to play an integral role in regulating exoelectrogenesis in these bacteria although this control is exerted by a Hypr-GGDEF DGC enzyme rather than an ortholog of DncV (Figure 1.2).

While *Geobacter spp*. use Hypr-GGDEF enzymes to synthesize cGAMP, there are also homologs of DncV present in other bacteria. The animal-commensal *E. coli* strain ECOR31, for example, has a horizontally transferred genomic island encoding the homolog $DncV_{ECOR31}$, which also synthesizes 3',3'-cGAMP [26]. Overexpression of $DncV_{ECOR31}$ leads to a reduction in the steady-state mRNA level of the transcriptional regulator *csgD*, resulting in downregulation of the red dry and rough (rdar) biofilm morphotype commonly expressed by *E. coli* strains [26]. This overexpression also inhibited swimming and swarming motility through inhibition of flagellin production [26].

1.8: Definition of the CD-NTase family of oligonucleotide synthases

In 2015 an in-depth domain analysis using DncV as a template showed that similar proteins likely are produced by many bacterial phyla [27]. This clustered analysis connected DncV with similarly organized "**S**econd **M**essenger **O**ligonucleotide or **D**i-nucleotide **S**ynthetase" (SMODS). A similar bioinformatic analysis came to the same conclusion that DncV orthologs are widespread and named this class of enzymes **c**GAS/**D**ncV-like **n**ucleo**t**idyltransfer**ases** (CD-NTases) [28]. Central to the research of Whiteley et al. in 2019 was the distribution and conservation of the catalytic activity of the DncV. DncV is structurally similar to cGAS (the metazoan cdN synthase that makes 2'3'-cyclic GMP AMP, discussed later), but distinct from the DGC and DAC domains that synthesize c-di-GMP and c-di-AMP, respectively. Therefore, CD-NTases (and SMODS) refer to a third distinct family of

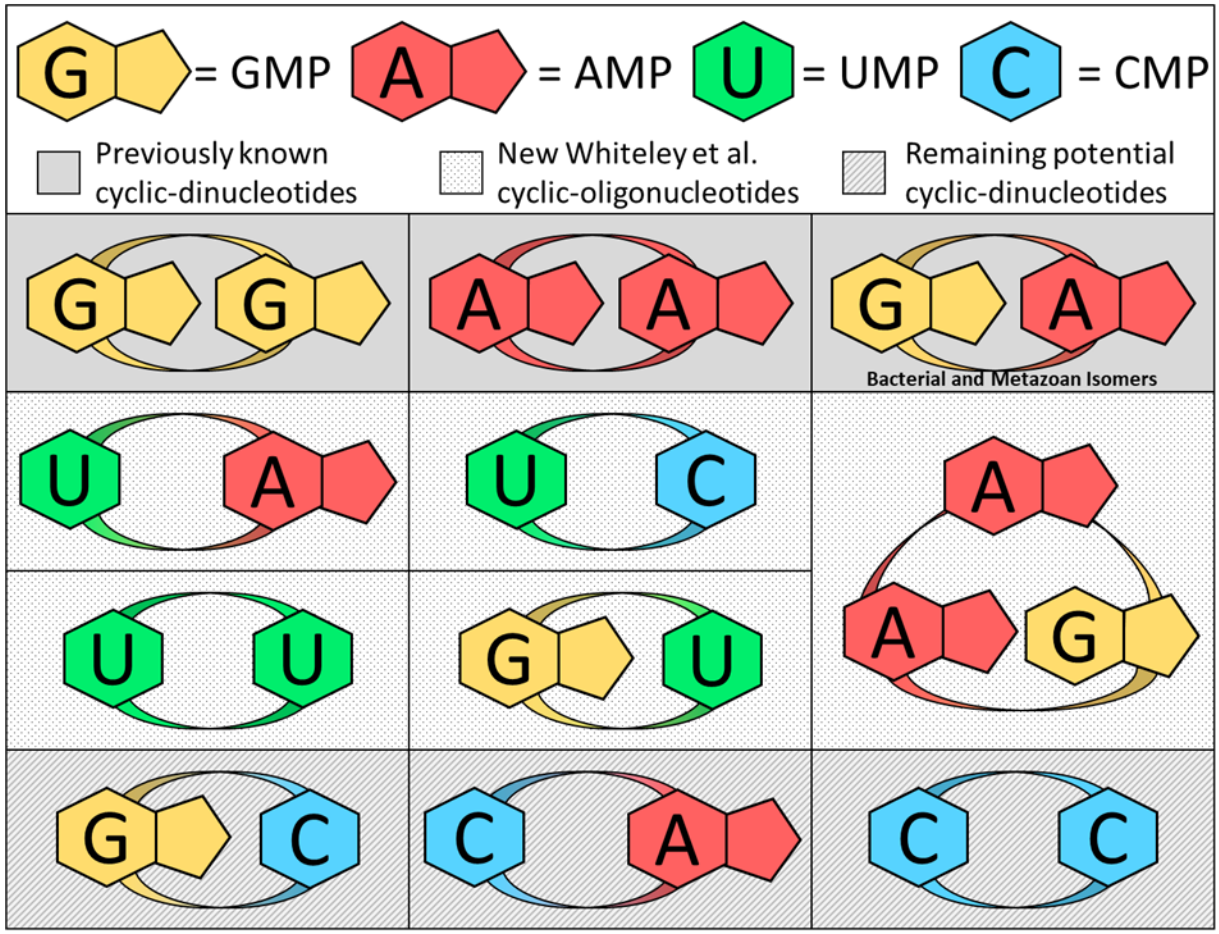

Figure 1.4: Expansion of the Cyclic-oligonucleotide Lexicon.
The four previously known cdN molecules were all composed of purine nucleotides (c-di-GMP, c-di-AMP, and both the bacterial 3'3' cyclic-GMP-AMP and metazoan 2'3' cyclic-GMP-AMP). With the identification of the bacterial CD-NTase family of enzymes, Whiteley et al. has expanded this signaling lexicon to include cyclic-UMP-AMP, cyclic-UMP-CMP, cyclic-di-UMP, cyclic-GMP-UMP, and the cyclic trinucleotide cyclic-AMP-AMP-GMP. Three possible cyclic dinucleotides remain to be discovered; cyclic-GMP-CMP, cyclic-CMP-AMP, and cyclic-di-CMP.

cdN synthases (Figure 1.3).

Whiteley et al. noticed that an animal commensal strain of Escherichia coli (ECOR31) harbors a duplication of *capV* and *dncV* and named the more conserved gene pair $capV_{ecoli}$ and $dncV_{ecoli}$, referred to previously as $DncV_{ECOR31}$ above, and the more degenerate copy *capE* and *cdnE* (Figure 1.3) [28]. In vitro evaluation of nucleotide synthase activity revealed that while DncV*ecoli* primarily produced 3'3'-cGAMP, analogous to *V. cholerae* DncV [9], the more

divergent CdnE generated the hybrid cdN cyclic-UMP-AMP (cUMP-AMP). Incredibly, the phospholipase activities of CapV*ecoli* and CapE were only activated *in vitro* by the cdN produced by their neighboring synthase (Figure 1.3) [28].

Pyrimidine-containing cdNs had been conspicuously absent from the four known cdN second messengers (c-di-GMP, c-di-AMP, 3'3'-cGAMP, and 2'3'-cGAMP [discussed below]), and it was hypothesized that their limited capacity for hydrogen bonding and base stacking interactions made uracil and cytosine intrinsically less favorable substrates for regulating critical signaling networks [29]. This discovery of cUMP-AMP represented the first naturally synthesized pyrimidine containing cdN and cUMP-AMP responsive effector protein, and it also demonstrates the insulated signaling capacity of cdN networks originating from homologous synthases [28].

Building on the shoulders of the prescient 2015 bioinformatic analysis by Burroughs et al.[27], Whiteley et al. performed an expansive bioinformatic search for CD-NTases and discovered that homologs are widely distributed across nearly all bacterial phyla and encoded in greater than 10% of the available bacterial genomes [28]. Eight CD-NTase clades were identified based on sequence alignments of the more than 5,600 unique enzymes identified. To understand the enzymatic repertoire of CD-NTases in these clades, Whiteley et al. purified 66 candidate enzymes and evaluated their capacity to catalyze nucleotide-based chemistry in vitro. 16 of the most active CD-NTases synthesized oligomeric nucleotide derived products including seven cdN species (such as cyclic-di-UMP), cyclic-AMP-AMP-GMP (cAAG), as well as several nucleotide products of unknown composition and structure (Figure 1.4) [28]. These observations greatly expanded the diversity of naturally synthesized nucleotide-based signaling molecules and suggest they may regulate undefined biological functions in bacteria and their metazoan host.

Additionally, like DncV and CdnE, many of these enzymes appear to be encoded in conserved operons within mobile genetic elements, suggesting they are not part of the core genomes of bacterial species but rather newly acquired adaptive traits [28]. However, the function of these new signaling systems is not well understood.

1.9: cGAMP signaling in metazoans

In addition to cGAMP being used by bacteria as a signaling molecule, it is also a potent regulator of the immune system in metazoan organisms. cGAMP is the only cdN that has been observed in metazoans and its role in innate immunity regulation stimulating defense against viral infection and cancer has been well characterized (Reviewed in [30–32]) (Figure 1.2). The eukaryotic cGAMP synthase, cGAS, shares close structural similarity to DncV while retaining no sequence similarity [33]. cGAS is located in the cytoplasm of eukaryotic cells, and its cGAMP synthesis activity is stimulated by binding to double-stranded DNA, which is triggered during a viral infection, mitochondrial stress, or genome instability [33–35]. The dsDNA binding site of cGAS is structurally analogous to the folate binding pocket found in DncV, however ligand occupancy of these allosteric sites leads to dichotomous effects on the catalytic activity of each enzymes; with dsDNA activating cGAS and folates repressing DncV [23]. Unlike DncV, which produces 3',3'-cGAMP with a phosphodiester ring containing two 3'-5' bonds, cGAS synthesizes the isomer 2',3'-cGAMP with a 2'-5',3'-5' linkage [36] (Figure 1.2). The structural similarity of cGAS to DncV is evident as one amino acid change in cGAS switches this enzyme from producing 2',3'-cGAMP to bacterial 3',3'-cGAMP [36]. The difference in the structure of these two molecules significantly impacts binding to eukaryotic receptors.

cGAMP synthesized by cGAS binds to the eukaryotic cdN receptor STING to induce Type I interferon production [37]. STING bound to a cdN was originally discovered as the signal

essential for upregulation of Type I interferons by the invasive bacterial pathogen *Listeria monocytogenes*, although in this case STING was recognizing cyclic-di-AMP secreted by the invading bacteria [38, 39]. It is now appreciated that activation of the STING pathway by *L. monocytogenes* inhibits bacterial clearance as this activation drives an anti-viral immune response [40]. In other words, c-di-AMP activation of STING is a manipulation of innate immune activation by *L. monocytogenes* to generate a more favorable environment for its replication and spread. While STING can bind c-di-AMP, c-di-GMP, and bacterial 3',3'-cGAMP, it has the strongest affinity for eukaryotic 2',3'-cGAMP [28, 37, 40, 41].

STING is not the only eukaryotic receptor of cdNs, as the enzyme RECON, an oxidoreductase, can bind to and be inhibited by bacterial cGAMP and c-di-AMP, but it does not bind tightly to eukaryotic 2,'3'-cGAMP or c-di-GMP [40, 42]. Inhibition of RECON by cdNs leads to activation of the inducible nitic oxide system (iNOS) and production of reactive oxygen species, which inhibit bacterial growth. Furthermore, binding of cdNs by RECON antagonizes signaling through STING by decreasing the concentration of free cdNs in the cell [40]. Therefore, bacterial cGAMP is a potent modulator of the immune system through multiple pathways.

1.10: CD-NTase family of Enzymes and Bacteriophage Defense

By the late summer of 2019 it had been established that DncV synthesized cGAMP in *V. cholerae* [9] activated CapV to induce membrane degradation [18], other CD-Ntase enzymes capable of synthesizing myriad cdNs [28] were part of conserved operons containing effectors proteins [27], and these operons were broadly distributed in mobile genetic elements across greater than 10% of published bacterial genomes [28]. The group of Rotem Sorek made an important observation that these CD-NTase encoding operons were often found close to known

bacteriophage defense mechanism (ie. CRISPR-Cas) [43], which was consistent with the initial bioinformatic analysis performed by Burroughs et. al. in 2015 [27].

Bacteriophage, or phage, defense mechanisms are often found together in genomic islands called "defense islands" [44]. The Sorek lab noticed that of the 637 instances where orthologs of *dncV* and *capV* were found together they were encoded near known phage immunity systems 63% of the time [43]. Additionally, in 96% of these cases, representing 613 of the 637 instances, *capV-dncV* homologs were also encoded as part of a four gene operon that included homologs of the El Tor *V. cholerae* VSP-1 genes *vc0180* and *vc0181* (Figure 1.1). VC0180 is a two domain peptide composed of an E1 (ubiquitin-activating enzyme) and E2 (ubiquitin-conjugating enzyme) while VC0181 encodes a JAB domain (deubiquitinase). While ubiquitination is a well described process of protein modification in eukaryotes canonically associated with marking proteins for degradation(reviewed in [45]), it is not known to be utilized in prokaryotes.

To test the utility of this four gene operon for the purposes of phage defense, the Sorek Lab provided *capV-dncV-vc0180-vc0181* and ~ 800 bp of its preceding 5' regulatory region from El Tor *V. cholerae* and a separate homologous system identified in *E. coli* TW1168 to the naïve laboratory strain of *E. coli* MG1655. When challenged with 10 phage known to normally infect this strain, provision of the *E. coli* derived *capV-dncV-E1/E2-Jab* system imparted resistance to 6 of the 10 phage while the El Tor variant provided resistance to only 2 [43]. This disparity in defense capacity between the homologous systems was attributed to the absence of potential *V. cholerae* regulatory factors required to properly deploy of the El Tor defense system in *E. coli*. Further studies using the *E. coli* operon demonstrated the activity of both CapV and DncV were absolutely required for phage defense while VC0180 and VC0181 play an important accessory

role, though the mechanism of their contribution remains enigmatic [43]. The required activity of CD-NTases and the adjacent effectors in these new phage defense systems led to the classification of these gene networks as **C**yclic-Oligonucleotide-**B**ased **A**ntiphage **S**ignaling **S**ystem (CBASS).

Two additional CBASSs have been described, and while their activities highlight the requirement for CD-NTase activity and a cdN responsive effector, these systems are flexible and accommodate different effector and associated accessory genes other than phospholipases and ubiquitinating enzymes. The simplest CBASS was demonstrated by provisioning a laboratory strain of *Bacillus subtilis* with a two gene CBASS encoded in a strain of *Bacillus cereus* [43]. Acquisition of this two gene operon, composed of putative effector containing four-transmembrane helices and a CD-NTase, was found to afford limited phage resistance to its heterologous host .

The third CBASS system described was found in clinical isolates of *E. coli* and *Pseudomonas aeruginosa*. This CBASS is composed of four or five genes (depending on the bacterial host) encoding eukaryotic homologs to HORMA domain proteins and Trp-13, the endonuclease NucC, and the novel CD-Ntase synthase CdnC [46]. In eukaryotes HORMA domains participate in the formation of signaling complexes involved in many processes including mitotic spindle assembly and autophagy by recognizing small peptide sequences presented by interacting partners and assembly of these complexes is antagonized by the activity of Trp-13 enzymes [46]. In this system, CdnC requires a protein-protein interaction with the associated HORMA peptide to synthesis cyclic-AMP-AMP-AMP (cAAA). Prior to assembling the HORMA:CdnC complex, the HORMA domain must first adopt a “closed” conformation which forms when bound to a short peptide sequence [46]. Despite this necessity, the activating

peptide sequence remains cryptic and is hypothesized to be a phage protein or a cellular peptide not typically available in healthy cells [46]. cAAA synthesized by HORMA activated CdnC then stimulates the endonuclease activity of NucC, which indiscriminately degrades both the host and invading phage genomic DNA [46]. The inappropriate formation of the HORMA:CdnC complex is prevented by the antagonistic activity of the associated Trip-13 protein which prevents "open" conformation HORMA from interacting with CdnC, thus keeping the CBASS system in an inactive state [46].

There are a number of mechanisms that bacteria deploy to thwart viral insults including surface receptor modulation to avoid phage detection, restriction modification systems that target invading phage genomes for degradation, and interference with the assembly of virions (reviewed in [47]). The common strategy employed by CBASSs, thus far, is termed abortive replication, an altruistic behavior where an infected bacterium performs an act of cellular suicide to curtail infection of its kin. The connection of *capV-dncV* to abortive replication though CBASS was consistent with our observation that activation of CapV by cGAMP led to membrane degradation and cell death [18, 43]. However, the mechanism that governs the appropriate deployment of CBASSs remains an outstanding question in the field.

Interestingly, the activity of CBASSs have only been observed in the context of heterologous hosts [43, 46]. In the case of the El Tor *V. cholerae* CBASS, numerous attempts to test the phage defense activities of CapV and DncV have failed to recapitulate phage immunity in the native background (W. Ng, C. Waters, and P. Kranzusch; personal communications). Presumably this is a consequence of the paradoxical nature of phage defense experiments, where the necessity to use successful viral predators to challenge a host means these phage have already evolved the capacity to evade the phage defenses being tested. The evolving conflict between

phage and bacteria is an example of an evolutionary arms race and it is reasonable to presume the CBASS systems are no exception (review in [48, 49]). The reason CBASS systems may be found in such diverse gene arrangements in ~10% of bacterial genomes is their modularity, in both effector and cdN composition, affords rapid counter evolution to neutralizing strategies made by their phage predators.

1.11: The expanding role of DncV beyond CBASS

Bioinformatic analysis of the co-occurrence of gene networks within the VSP islands across bacterial genomes, undertaken in collaboration with the laboratory of Eva Top at the University of Idaho and describe in Chapter 4, revealed that along with the previously described CBASS system, *dncV* also co-occurs with the VSP-1 encoded putative deoxycytidylate deaminase *vc0175*, renamed *dcdV*.

Deoxycytidylate deaminases (DCD), part of the cytidine deaminase (CDA) family of enzymes, play a vital role in de novo pyrimidine biosynthesis in many organisms by catalyzing the deamination of deoxycytidine monophosphate (dCMP) to form deoxyuridine monophosphate (dUMP) [50]. dUMP is later methylated by thymidine synthase using 5,10-methylenetetrahydrofolate (5,10-MTHF) to form thymidine monophosphate (dTMP) [50]. This process is critical to the de novo synthesis of thymidine triphosphate (dTTP) which is required for the replication of DNA genomes. DCD enzymes are also utilized by some phage to coopt nucleotide biosynthesis to provide the appropriate ratio of deoxynucleotide (dNTP) substrates to facilitate expeditious replication of their viral genome, which often differs in G + C content from their host [51].

Interestingly, in metazoans the CDA enzyme APOBEC3 has been well studied for its anti-viral activity in innate immunity (reviewed in [52]). APOBEC3 is expressed in a number of

cell types [53], and during a retroviral infection it interacts with viral capsid proteins and viral-RNA. These interactions allow APOBEC3 to stowaway in assembling virions from which it is later released into the cytoplasm of a new host cell during a subsequent infection. APOBEC3 deaminates dC substrates in ssDNA, and this activity serves to neutralize retro-viral infections by corrupting the minus-strand DNA transcribed by the reverse transcriptase. Unlike APOBEC3, most prokaryotic CDA enzymes primarily operate on free nucleotides.

Recognizing that *dncV* was recently found to coordinate the CBASS systems discussed previously [43], phage-immunity modules are frequent packaged together in mobile genetic elements [43, 44], and the precedent for CDA enzymatic activity to aid in viral defense [52], we sought to understand the connection between *dncV* and *dcdV*. This study, discussed in depth in Chapter 3, has revealed that *dcdV* encodes a dual domain protein which is capable of deaminating both dCMP and dCTP in a conserved DCD domain. While the specific activity of the less conserved nucleotide/nucleoside kinase-like domain, canonically involved in transferring the γ-phosphate of ATP to dNMP substrates, has yet to be defined the combined activities of these domains enhance the incorporation of genomic dU and lead to a cell filamentation phenotype. Interestingly, filamentation is abrogated in the presence of a previously unknown gene product encoded immediately 5' of *dcdV*, herein named *difV*. In total, our evidence suggests that DcdV and DifV function in concert to abruptly shift a bacterium's deoxynucleotide pools and this process is likely detrimental to its own survival. While much remains to be studied about DcdV, including its predicted connection to DncV, and the regulator role of DifV, we are confident that this system is important to understanding of El Tor *V. cholerae* biology and may contribute to phage defense.

1.12: Expanding the known functions of VSP-1 in *V. cholerae*

The initial aim of my thesis was to understand the role of cGAMP in El Tor *V. cholerae*. This work is highlighted in detail in Chapter 2, where I present our characterization of the first cGAMP regulated effector, CapV, and the consequences of its activation on the cell membrane. Recognition of the DncV–cGAMP–CapV signaling network sparked my curiosity in understanding the other biological functions encoded within VSP-1 and 2 and their contributions to the evolution of Seventh Pandemic Cholera (Figure 1.1). This question was generously funded by BEACON and fueled our collaboration with the lab of Eva Top at the University of Idaho. This collaboration identified functional gene pathways encoded within the VSP islands based on the co-occurrence of island genes within published bacterial genomes. The rapidly developing field of phage defense, and its link to CD-NTase signaling, focused our attention on a predicted gene network between *dncV* and the previously uncharacterized deoxycytidylate deaminase we named *dcdV*. In Chapter 3, I present our research demonstrating the activity of DcdV, and our discovery of a novel inhibitor of this enzyme encoded in VSP-1 we named DifV. Finally, in Chapter 4, I discuss the implications of our findings as they relate to the evolution of pandemic *V. cholerae,* the field of cdN signaling, and address some of the outstanding questions which remain directly related to Chapters 2 and 3 and those with broad implications beyond CapV and DcdV.

Chapter 2: Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*

2.1: Preface

Contents of this chapter, provided as Supplemental Files 1 and 2, were published in Proceeding of the National Academy of Science (PNAS) in 2018 (Severin, G. B*., Ramliden, M. S.*, Hawver, L. A., Wang, K., Pell, M. E., Kieninger, A.-K., Khataokar, A., O'Hara, B. J., Behrmann, L. V., Neiditch, M. B., Benning, C., Waters, C. M., Ng, W.-L. (2018). Direct activation of a phospholipase by cyclic GMP-AMP in El Tor *Vibrio cholerae*. Proceedings of the National Academy of Sciences of the United States of America, 115(26), E6048–E6055).
(*= contributed equally). Per the publisher PNAS, "Authors retain the following rights under the PNAS default license - The right to include the article in the author's thesis or dissertation." I contributed data for Figures (Fig. #) 1B, 2B, 2D, 3D, 4, and 5 in Supplemental File 1. I contributed data for Supplemental Figures (Fig. S#) 2C, 3A, and 4 in Supplemental File 2.

2.2: Chapter Body

Chapter 2 is provided as Supplemental Files 1 and 2.

Chapter 3: Identification of DcdV, a VSP-1 encoded deoxycytidylate deaminase, and its allosteric regulator DifV

3.1: Preface

The works presented in this chapter are a collaboration between me, Brian Y. Hsueh, and Clinton Elg. Mr. Hsueh and I contributed equally to the characterization and identification of DcdV and DifV. Mr. Elg, from the laboratory of Dr. Eva Top at the University of Idaho, developed the Correology software suite and performed the bioinformatic analysis of VSP-1 and 2. Christopher R. Rhoades created the three VSP Island knockout strains.

3.2: Abstract

The El Tor biotype of the Gram-negative bacterial pathogen *Vibrio cholerae* is responsible for initiating and perpetuating the longest cholera pandemic in recorded history (1961-current). Two genetic features that distinguish the El Tor biotype from strains of the classical biotype, responsible for the previous six pandemics, are the presence of two genomic islands VSP 1 & 2. It was recently demonstrated that a four gene operon in VSP-1 (*capV-dncV-vc0180-vc0181*) constitutes an anti-phage defense system, called CBASS, which is coordinated by cyclic-GMP-AMP (cGAMP) synthesized by DncV. Despite the significance of this finding many of the ~36 ORFs encoded in the VSP islands remain uncharacterized. We developed a bioinformatic pipeline to uncover other gene networks within the VSP islands by looking for the co-occurrence of island gene products within bacterial genomes. In addition to the known CBASS system, our analysis predicted *dncV* was involved in a gene network with the putative deoxycytidylate deaminase *vc0175*, renamed here-in as **d**eoxy**c**ytidylate **d**eaminase ***V****ibrio* (*dcdV*). While ectopic expression of *dcdV* in WT El Tor *V. cholerae* revealed no readily distinguishable phenotypes, a strain lacking VSP-1, where *dcdV* is natively encoded, demonstrated reduced growth yield and a filamentous cell morphology. This filamentous phenotype is also inducible in laboratory strains

of *E. coli* ectopically expressing *dcdV* that can be complemented back to wild-type cell morphology by supplying a single copy cosmid containing VSP-1. This complementation was later attributed to a previously unannotated gene product encoded immediately 5' of *dcdV*, named herein as *dcdV* insensitivity factor *Vibrio* (DifV). DcdV is a two domain protein containing a conserved deoxycytidylate deaminase (DCD) C-terminal domain and a putative nucleoside/nucleotide kinase (NK) N-terminal domain. The catalytic activity of the DCD domain performs the unique deamination of both dCMP and dCTP substrates, producing dUMP and dUTP. While the activity of the NK domain remains to be characterized, both the NK and DCD domain promote the incorporation of genomic dU and their combined activity is required to induce cell filamentation. While the biological utility of DcdV and its connection to DncV remain to be elucidated we discuss their relationship in the context of a novel phage-defense system.

3.3: Introduction

Vibrio cholerae, the etiological agent behind the diarrheal disease cholera, is a monotrichous, crescent shaped, Gram-negative bacterium found ubiquitously in marine environments. There have been seven recorded pandemics of cholera, beginning in 1817, the first six of which are believed to have been caused by strains of the classical biotype. The seventh pandemic, which began in 1961 and continues to plague vulnerable populations today, was initiated and perpetuated by circulating strains of the El Tor biotype. Numerous phenotypic and genetic characteristics are used to distinguish the classical and El Tor biotypes [54], but it is thought that El Tor's acquisition of two unique genomic islands of unknown origins, named the Vibrio Seventh Pandemic Islands 1 and 2 (VSP-1 and 2) [4], prior to the start of the seventh

pandemic played a pivotal role in El Tor's evolution to pandemicity and the displacement of the classic biotype in modern cholera disease [3].

Combined, VSP-1 and VSP-2 represent ~36 putative ORFs encoded in ~39kb of genetic material that are typically separated by ~330 kb on the larger of the two El Tor *V. cholerae* chromosomes [4, 5] (Figure 3.2A**) (**Figure 3.3A). While the majority of the genes in these two islands remain to be fully characterized, it has been hypothesized that the biological functions they encode may contribute to environmental persistence [55] and/or the pathogenicity [9] of the El Tor biotype.

The ~26 ORFs encoded within VSP-2 (Figure 3.3A), which include putative condensins, a chitinase, a pseudo-pilin, and numerous transcriptional regulators, have yet to be interrogated for their specific contributions to El Tor fitness. Moreover, the genetic composition of VSP-2 appears to be more fluid across different strains of the El Tor biotype and non-pathogenic *V. cholerae* than VSP-1 [8, 56].

In 2012, the first two genes characterized of the ~11 putative ORFs encoded within VSP-1 were *dncV* (*vc0179*) and a transcriptional repressor named *vspR* (*vc0176*) [9] (Figure 3.2A). At the time DncV represented the first cyclic dinucleotide (cdN) synthase found in any living system to catalyze the formation of the second messenger 3'3'cyclic GMP-AMP (cGAMP). cdN second messengers such as cGAMP, cyclic di-GMP, and cyclic di-AMP are intracellularly synthesized and degraded in response to changes in the bacterium's environment. The intracellular concentration of a cdN dictates changes in diverse adaptive behaviors including biofilm formation, motility, stress response, and osmotic homeostasis at the levels of transcription initiation, post-transcription regulation, and allosteric interactions with effector proteins (reviewed in [10, 11, 13]).

As described in Chapter 2, we found that an elevated intracellular concentration of cGAMP in El Tor *V. cholerae* resulted in cell toxicity, leading to the discovery of the cGAMP-activated phospholipase CapV (*vc0178*) [18]. Both *dncV* and *capV* are encoded as part of a putative four gene operon (*capV-dncV-vc0180-vc0181*) in VSP-1 and stimulation of CapV phospholipase activity by cGAMP leads to rapid degradation of the inner bacterial membrane and ultimately cell death. The function of the *dncV-capV* cGAMP signaling pathway in El Tor *V. cholerae*, along with *vc0180* and *vc0181*, has recently been suggested to coordinate an anti-phage defense system called the **c**yclic-oligonucleotide-**b**ased **a**ntiphage **s**ignaling **s**ystem (CBASS) [43]. CBASS executes an antiphage process termed abortive replication where an infected bacterium performs altruistic autolysis to prevent the virus from completing its replication cycle and contain the infection for the good of the bacteria community [43]. Such functions for *dncV* and *capV* were demonstrated by expression of these genes in the heterologous host *Escherichia coli* followed by infection with *E. coli* specific phage, but such phage protection mediated by *dncV* and *capV* in their native genome locus in *V. cholerae* has not yet been documented. It has also been shown that *dncV*-like proteins, referred to as CD-NTase [28] or SMODS [27], are capable of synthesizing a diverse array of nucleic acid compounds which are likely to specifically regulate the activity of neighboring effector proteins.

CBASS is not limited to El Tor *V. cholerae* VSP-1 as networks of homologous genes are shared in diverse mobile genetic elements found in other bacteria [43, 46]. The observation that antiphage defense systems, such as CBASS, are frequently packaged together in mobile genetic elements, called defense islands [44, 57], suggest that the seven remaining putative VSP-1 ORFs, which include three transcriptional regulators (*vc0176*, *vspR* [9], and *vc0182*), three xer–like recombinases (*vc0183*, *vc0184*, *vc0185*),

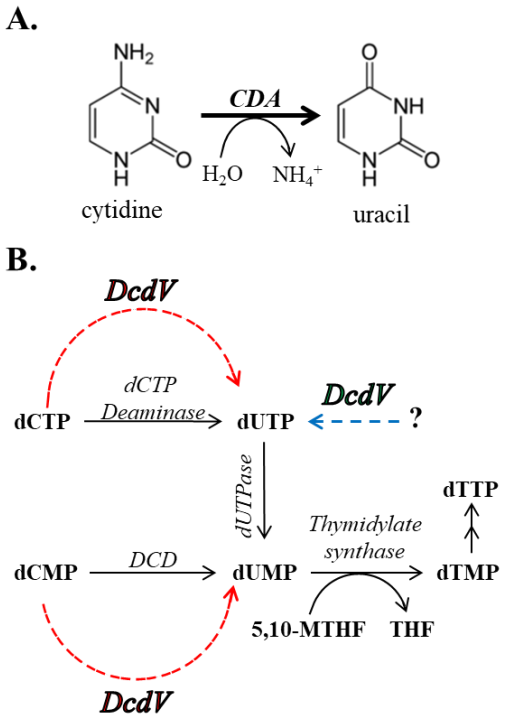

Figure 3.1: Cytidine deamination reaction and simplified de novo thymidine triphosphate biosynthesis pathway.
(A.) The activity of the cytidine deaminase family of enzymes (CDA) hydroxylate the 4'-amino group of cytidine to form uracil and free ammonium. (B.) Simplified generic de novo thymidine triphosphate biosynthesis pathway from dCTP and dCMP precursors. Enzymes involved in this process are italicized and substrates in bold. The activities of DcdV are depicted in hatched lines; red for the C-terminal DCD domain and blue for the N-terminal NK domain. 5,10-methenyltetrahydrofolate (5,10-MTHF), tetrahydrofolate (THF).

and a deoxycytidylate deaminase-like protein (*vc0175*), may also participate in anti-phage activities.

Deoxycytidylate deaminases (DCD), part of the cytidine deaminase (CDA) family of enzymes, play a vital role in pyrimidine biosynthesis in many organisms by catalyzing the deamination of deoxycytidine monophosphate (dCMP) to form deoxyuridine monophosphate (dUMP) [50] (Figure 3.1A). With the exception of the deoxycytidine triphosphate (dCTP) deaminase enzymes [58], the CDA deamination reaction is catalyzed in a conserved Zn-binding

active site that uses deprotonated water for the hydroxylation of the cytidine 4'-amino group to form uridine and free ammonium (Figure 3.1A,B) . The activity of these enzymes is critical for providing the necessary dUTP/dUMP building blocks for deoxythymidine triphosphate (dTTP) synthesis. However, accumulation of these dU intermediates poses a risk to cells as DNA polymerases poorly discriminate between dUTP and dTTP allowing for the miss incorporation of dU in place of dT during DNA replication [59]. The activity of deoxyuridine phosphatases (dUTPase) reduce the likelihood of erroneous incorporation of genomic dU by rapidly hydrolyzing available dUTP to dUMP, which is then converted into dTMP by thymidylate synthase (TS) using 5,10-methenyltetrahydrofolate (5,10-MTHF) (Figure 3.1B) [50]. The delicate balance of enzymatic activity across the pyrimidine biosynthesis pathway can be corrupted by viruses that deploy their own DCD, dUTPase, and TS enzymes to hijack host nucleotide biosynthesis to ensure the appropriate ratio of deoxyribonucleotide precursors for replicating their own A + T rich viral genomes [51, 60–63].

To predict novel biological pathways contained within El Tor *V. cholerae* VSP-1 and VSP-2 we performed a correlogy based bioinformatic analysis using a software package we call Correlogy, inspired by previously published methods [64, 65]. By mining the NCBI RefSeq non-redundant protein database for the co-occurrence of VSP island gene product homologs in bacterial genomes we identified numerous instances of co-occurring genes, called gene networks, across the bacterial phyla. Genetic components that make up a gene network are likely to function in a shared biological pathway or network. Our previous finding of the biological connection between *dncV* and *capV* [18] was identified as a gene network, validating this approach. Another identified VSP island gene network was a predicted association between the VSP-1 genes *dncV* and the putative deoxycytidylate deaminase-like protein *vc0175* [renamed

herein as **d**eoxy**c**ytidylate **d**eaminase in ***V****ibrio* (*dcdV*)]. DcdV encodes both a predicted C-terminal DCD domain and an N-terminal nucleotide kinase (NK) domain.

We found that ectopic expression of *dcdV* in El Tor strains lacking VSP-1 and a heterologous *Escherichia coli* host resulted in severely filamentous cells. This DcdV induced filamentous morphology is dependent on the catalytic activities originating from both the conserved C-terminal DCD, which deaminates both dCMP and dCTP substrates, and the less conserved structural features present in the predicted N-terminal NK domain, whose activity has yet to be elucidated.

We further show that DcdV induced cell filamentation is abrogated in WT El Tor *V. cholera*e and *E. coli* complemented with a cosmid containing VSP-1, and this negative DcdV regulatory activity was mapped to a 172 NT segment of DNA encoded 5' of the *dcdV* locus [renamed herein as **D**cdV **i**nsensitivity **f**actor in *Vibrio* (*difV*)]. While the biochemical make up of DifV has yet to be determined, DcdV represents the first CDA family enzyme, to our knowledge, to be allosterically regulated by a sRNA or small peptide. We discuss the possibilities for *dcdV*'s association with *dncV* in bacterial genomes as well as the biological functions behind its activities and regulation as they relate to phage defense.

3.4: Materials and Methods

3.4.1: Bacterial Strains, Plasmids, and Growth Conditions.

The *V. cholerae* strains constructed in this study were derived from the El Tor biotype strain C6706str2 using the pKAS32 suicide vector, as described previously [66]. All vectors were constructed by Gibson Assembly (NEB). The vectors used for chromosomal deletions were generated by Gibson Assembly using three fragments: 500 bp of sequence upstream of the gene

of interest, 500 bp of sequence downstream of the gene of interest, and cloned into the KpnI and SacI restriction sites of pKAS32. The ectopic expression vectors were constructed by Gibson Assembly using PCR-amplified inserts, and pEVS143 and pBMM67EH linearized by EcoRI and BamHI, and pET28b digested with NcoI and XhoI. Site-directed mutagenesis used for the construction of DcdV variants were performed using the SPRINP method previously described [67]. Plasmids were introduced into *V. cholerae* through biparental conjugation using *E. coli* BW29427. For expression in *E. coli,* plasmids were transformed into DH10b by electroporation. *E. coli* BL21DE3 was used for the expression of pET vectors utilized in the in-vitro cell lysate deamination assay. Unless otherwise stated, all *E. coli* and *V. cholerae* were grown in Luria-Bertani (LB) at 35^{o} C. Medium was supplemented with the following as needed: ampicillin (100 µg/mL), kanamycin (100 µg/mL), tetracycline (10 µg/mL), and isopropyl-β-D-thiogalactoside (IPTG) (100 µg/mL). *E. coli* BW29427, a diaminopimelic acid (DAP) auxotroph, was additionally supplemented with 300 µg/mL DAP.

Table 3.1: Strains and descriptions

Strains	Name in this Manuscript	Relevant characteristics	Source or reference
E. coli			
DH10b		*F-mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139Δ(ara, leu)7697 galU galK λrpsL nupG*	ThermoFisher Scientific
BW29427		*RP4-2(TetSkan1360::FRT)*, *thrB1004*, *lacZ58*(M15), *ΔdapA1341::[erm pir^{+}]*, *rpsL*(strR), *thi-*, *hsdS-*, *pro-*	Lab Stock
BL21(DE3)		*F- ompT hsdSB(rB -mB +) gal dcm (DE3)*	Lab Stock
NR8052		*Δ(pro-lac) thi ara trpE9777 ung-1*	[68]
V. cholerae			
C6706 Str2	WT	Wild type O1 El Tor; Sm^{R}	[69]
CR01	ΔVSP-1	O1 El Tor ΔVSP-1	This study
CR02	ΔVSP-2	O1 El Tor ΔVSP-2	This study
CR03	ΔVSP-1/2	O1 El Tor ΔVSP-1/2	This study
BYH0206	ΔIG	O1 El Tor Δ*intergenic vc0175-vc0176*	This study
BYH0207	Δ*vc0176*	O1 El Tor Δ*vc0176*	This study
BYH0255	Δ*vc0175-176*	O1 El Tor Δ*vc0175-176*	This study
BYH0256	Δ*vc0177-181*	O1 El Tor Δ*vc0177-181*	This study
BYH0257	Δ*vc0182-185*	O1 El Tor Δ*vc0182-185*	This study
GS05	Δ*vc0175*	O1 El Tor Δ*vc0175 (dcdV)*	This study

Table 3.2: Plasmids and descriptions

Plasmids	Name in this Manuscript	Relevant characteristics	Source
pEVS141	pVECTOR	pEVS143 without pTac; Km^r	[70]
pEVS143		Broad-host range pTac overexpression vector; Km^r	[71]
pMMB67EH		Broad-host range pTac overexpression vector; Amp^r	[72]
pET28b		T7 promoter; Km^r	Novagen
pLAFR	pLAFR	pLAFR; Tet^r	[73]
pCCD7	pCCD7	pLAFR::*VSP-1*; Tet^r	[18]
pGBS98	$pDcdV^{6xHis}$	pEVS143::*dcdV-6xHis C-term;* Km^r	This study
pGBS81	$pDcdV^{C411A,C414A}$	pEVS143::*dcdV-C411A+C414A*; Km^r	This study
pCMW204	pDcdV	pEVS143::*dcdV;* Km^r (“**pDcdV**” **for all experiments *except*** complementation of ΔIG by pOrf1 in Fig. 3.6C and in vitro lysate experiment in Fig. 3.11)	This study
pGBS87	pDcdV	pMMB67EH::*dcdV*; Amp^r (“**pDcdV**” ***only*** for complementation of ΔIG by pOrf1 in Fig. 3.6C)	This study
pGBS65	pDcdV	pET28b*::dcdV-6xHis C-term*; Km^r (“**pDcdV**” ***only*** for in vitro lysate experiment in Fig. 3.11)	This study
pGBS71	$pDcdV^{E384A}$	pEVS143::*dcdV-E384A;* Km^r (“**$pDcdV^{E384A}$” for all experiments *except* in vitro lysate experiment in Fig. 3.11**)	This study
pGBS82	$pDcdV^{E384A}$	pET28b::*dcdV-E384A-6xHis C-term*; Km^r (“**$pDcdV^{E384A}$” *only* for in vitro lysate experiment in Fig. 3.11**)	This study
pGBS80	pOrf1	pEVS143::*orf1, (position in N16961 chromosome I [177,230-177,008])*	This study
pGBS104	$pDcdV^{S52P}$	pEVS143::*dcdV-S52P;* Km^r	This study
pGBS97	$pDcdV^{S52W}$	pEVS143::*dcdV-S52W;* Km^r	This study
pGBS103	$pDcdV^{S52K}$	pEVS143::*dcdV-S52K;* Km^r	This study
pGBS92	$pDcdV^{K55A}$	pEVS143::*dcdV-K55A;* Km^r	This study
pGBS106	$pDcdV^{D162A,Q163A}$	pEVS143::*dcdV-D162A, Q163A;* Km^r	This study

3.4.2: Growth Curve Assays.

Overnight cultures were diluted 1:1000 into LB supplemented with antibiotics and IPTG in a 96-well microplate (Costar®). The cultures were grown for ~15 hours in a BioTek Synergy HTX Plate Reader, with OD_{600} measurements every 15 min.

3.4.3: Fluorescence Microscopy and Analysis.

Cells were imaged as previously described [74]. Briefly, overnight cultures were diluted 1:1000 into LB supplemented with antibiotics and IPTG. Cultures were grown and induced for 7-8 hour, at which point cells were diluted to an OD_{600} of 0.5 in 1X PBS, then membrane stain N-(3-Triethylammoniumpropyl)-4-(6-(4-(Diethylamino) Phenyl) Hexatrienyl) Pyridinium Dibromide (FM4-64) (Sigma) was added to a final concentration of 20 µg/mL. 1% agarose pads in deionized water were cut into squares of approximately 20 x 20 mm and placed on microscope slides. 2 µl of diluted cultures were spotted onto a glass coverslip and then gently placed onto the agarose pad. FM4-64 signal was visualized using a Leica DM5000b epifluorescence microscope with a 100X-brightfield objective under RFP fluorescence channel. Images were captured using a Spot Pursuit CCD camera and an X-cite 120 Illumination system. Each slide was imaged with at least 20 fields of view for each biological replicate. Cell lengths were processed using the Fiji plugin MicrobeJ, and data were visualized and analyzed using R by quantifying the length of the curvilinear (medial) axis of detected cells.

3.4.4: In-vitro Nucleic Acid Deamination Assay

Cell Lysate Prep: Overnight cultures were sub-cultured 1:333 and grown at 35° C and 210 rpm shaking to an OD_{600} of ~0.5 - 1.0. Cultures were induced with 1mM IPTG, supplemented with 100 µM $ZnSO_4$, and grown for an additional 3 hr. Cell pellets from 100 mL

of induced cultures were harvested in two successive 15 min centrifugation steps at 4k x g and 4° C. Supernatants were decanted and pellets were snap frozen in an ethanol and dry ice bath and stored at -80° C. Pellets were thawed on ice and suspended in 2 mL of lysis buffer (50 mM $NaPO_4$, pH 7.3, 300 mM NaCl, 2 mM β-mercaptoethanol, 20% glycerol). 1 mL of cell suspension was transferred to a microcentrifuge tube and sonicated on ice (20% amplitude, 20 sec total, 2.5 sec on, 2.5 sec off). Crude lysates were centrifuged at 15k x g for 8 min at 4° C and clarified lysates were transferred to fresh microcentrifuge tubes on ice. Clarified lysates were normalized for total protein to 1.9 mg/mL using Bradford reagents and a BSA standard. 26.5 µL reactions composed of lysis buffer, nucleic acid substrates, and 3.5 µL of normalized clarified lysates were assembled in PCR strip tubes, mixed by gentle pipetting, and incubated at 21° C for 60 minutes. NH_3Cl solutions at the indicated concentration were dissolved in lysis buffer and substituted for nucleic acid substrates as positive ammonium controls.

Ammonium Detection: The evolution of NH_4^+ by deamination of the nucleic acid substrates was observed using a phenol-hypochlorite reaction to produce indophenol in a clear 96-well microtiter plate and modified from Dong et al. 2015 [75] and the work of Ngo et al. [76] was considered when designing the lysis buffer. 50 µL of Reagent A (composition below) was added to each well followed by 20 µL of the completed in vitro deamination reaction described above. The phenol-hypochlorite reaction was initiated by the addition and gentle mixing of 50 µL Reagent B (composition below) to the wells. The reaction was incubated at 35° C for 30 min and the ABS_{630} was measured using a plate reader.

Reagent A = 1:1:0.04 (v/v/v), water: 0.5% (w/v) sodium nitroprusside (Sigma) in water: phenol solution (Sigma, P4557)

Reagent B = 1:1 (v/v), 6% (w/v) sodium hydroxide (Sigma) in water: 1.5% (v/v) sodium hypochlorite solution (Sigma, reagent grade) in water.

3.4.5: Protein Stability Assay

Overnight cultures were sub cultured 1:500 into fresh media supplemented with antibiotics and grown at 35 °C with 220 RPM shaking. After 1 h cultures were induced with 100 µM IPTG and allowed to continuing growing for another 1 h 45 m. 10 mL aliquots of each culture was removed prior to the addition of 50 µg/mL chloramphenicol (stop translation) and 250 µg/mL rifampicin (stop transcription). 10 mL culture aliquots were also removed 8, 16, and 32 min after addition of chloramphenicol and rifampicin. Each culture aliquot was immediately centrifuged at 4700 x g for 4 m at 22 °C, the supernatant was quickly removed by aspiration, the pellet was plunged into an EtOH + dry ice bath, and stored at -80 °C. Cell pellets were thawed on ice and suspended in 1 mL lysis buffer (50 mM NaPO4, pH 7.3, 300 mM NaCl, 2 mM β-mercaptoethanol, 20% glycerol) supplemented with Roche cOmplete™ protease inhibitor and DNAse I.. Cell suspensions were transferred to a microcentrifuge tube and sonicated on ice (20% amplitude, 20 sec total, 2.5 sec on, 2.5 sec off). Crude lysates were normalized for total protein using a Bradford assay. 34 µg of total protein per sample was loaded onto a single 10% acrylamide gel and proteins were transferred to Optitran reinforced nitrocellulose blot. The blot was blocked using skim milk and incubated with 1:5000 THETM His Tag Antibody, mAb, Mouse (GenScript) followed by 1:4000 Goat Anti-Mouse IgG Antibody (H&L) [HRP], pAb (GenScript), treated with Pierce™ ECL Western Blotting Substrate, and imaged using a Amersham 600 Imager. Protein signal intensity was quantified using ImageJ.

3.4.6: Determination of Uracil in DNA

The genomic uracil incorporation in vitro assay was carried out as described previously with slight modifications [77]. Briefly, genomic DNA was purified from cells grown in LB supplemented with antibiotics and IPTG for 7-8 hours using Wizard ® Genomic DNA Purification Kit (Promega). The purified genomic DNA (3 µg) was digested overnight at 37° C with 25 U uracil DNA glycosylase from *E. coli* (UDG) and 50 U human AP endonuclease 1 (APE 1) (New England Biolabs) in 1X NEB-buffer 4. Approximately 2.5 µg of each reaction was loaded on a 0.8% agarose gel stained with EZ-Vision® Dye (VWR), and images were taken using GelDoc system (Bio-rad).

3.5: Results

3.5.1: VSP-1 Island Gene Correlogy Study Indicates *dncV* and *dcdV* may share a functional connection

The two unique genomic regions, VSP-1 and VSP-2, are present in the seventh pandemic *Vibrio cholera* El Tor strains but not previous pandemic isolates [4]. The acquisition of these islands appears to be the final defining genomic event which allowed the El Tor lineage to emerge as the current pandemic strain [3]. However, the source of these genomic islands is unknown. BLAST searches for the entire VSP-1 region using nucleotide or amino acid sequence does not return any hits in the current NCBI database, suggesting this order of genes is not present in any other sequenced genomes.

A novel approach informed by previous work [64, 65] was developed to help define the biological function and groups of genes that function in a common pathway for these regions that lacked informative annotations. Many biological processes are the results of a set of gene

products interacting to achieve a biological task such as metabolic pathways, cell wall synthesis, or molecular machines. The sets of gene products that accomplish a biological task can be described as a "gene network", and the proteins encoded by these genes may physically interact or function in a common pathway. Many of these gene networks have deep evolutionary history and are widespread in many diverse taxa. We hypothesized that genes in a gene network will co-occur together in the genomes of diverse taxa at a higher frequency than chance alone would predict, and we used this reasoning to identify putative gene networks in VSP-1. Kim and Price [64] previously explored genetic co-occurrence across the sequenced microbial datasets and coined the term 'correlogy' from the words 'co-occurrence' and 'orthologs'. However, their mathematical approach was never widely adopted nor was it implemented in publicly available software.

Our software package, Correlogy, builds on Kim and Price's approach. Correlogy detects genes that preferentially co-occur across the sequenced bacterial domain and assigns them to putative maximum relatedness subnetworks (MRS). These MRS can be considered as putative gene networks, with each MRS containing genes that preferentially co-occur in diverse taxa and may contribute to the same biological task. Importantly, Correlogy can only determine co-occurrence of genes across entire bacterial genomes; it does not take into account nor provide information on the spatial organization of these genes within a given genome.

To establish MRS for the VSP islands genomic regions we performed a BLASTP amino acid sequence search for each gene against the NCBI non-redundant protein database

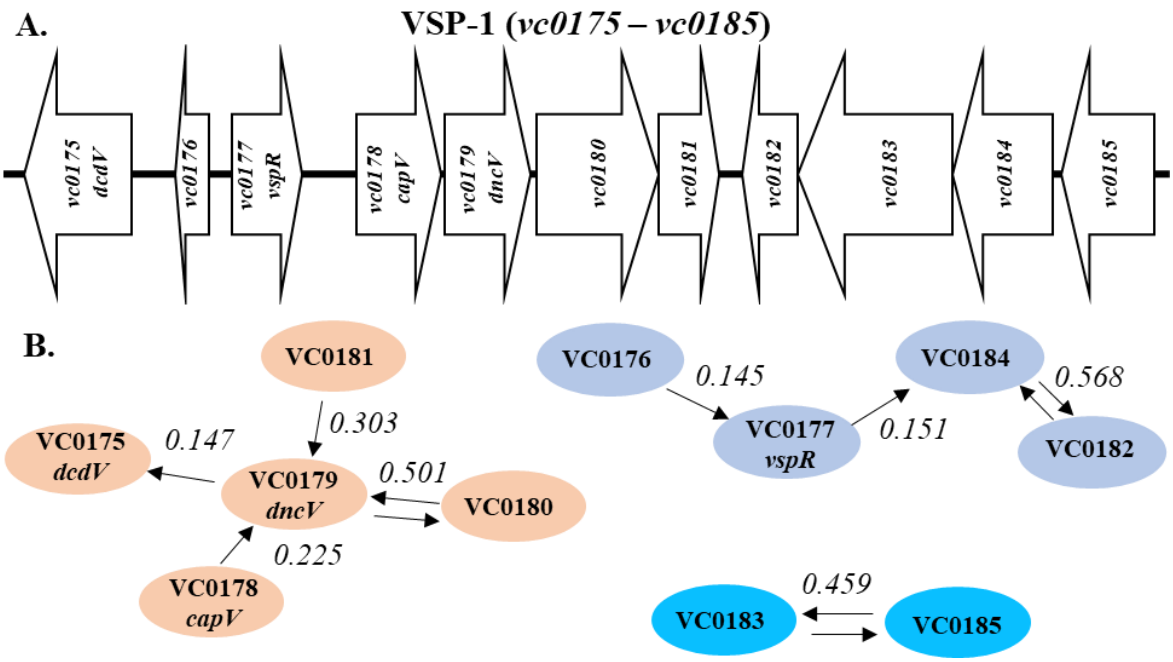

Figure 3.2: VSP-1 schematic and predicted gene networks (MRS).
(**A.**) Cartoon schematic of VSP-1 from El Tor *V. cholerae* N16961. Not to scale. (**B.**) Gene network predictions for VSP-1 where arrows show highest partial correlation W_{ij} (in italics) for each VSP-1 gene (represented by ovals). Arrows pointing in opposing directions between the same two genes have the same W_{ij}. MRS = maximum relatedness subnetworks

with an E-value cutoff of 10^{-4}. The BLAST results were limited to bacterial genomes, and all taxa belonging to the genus *Vibrio* were removed to avoid bias from closely related vertical inheritance. The BLAST results were used to generate a presence or absence matrix of VSP-1 homologues with all species along one axis and VSP genes on the other axis.

Next, a pairwise Pearson correlation value was calculated between all VSP-1 genes *i* and *j* using binary data from the above-mentioned presence/absence matrix*:*

$$r_{ij} = \frac{C_{ij}N - E_iE_j}{\sqrt{E_iE_j(N - E_i)(N - E_j)}},$$

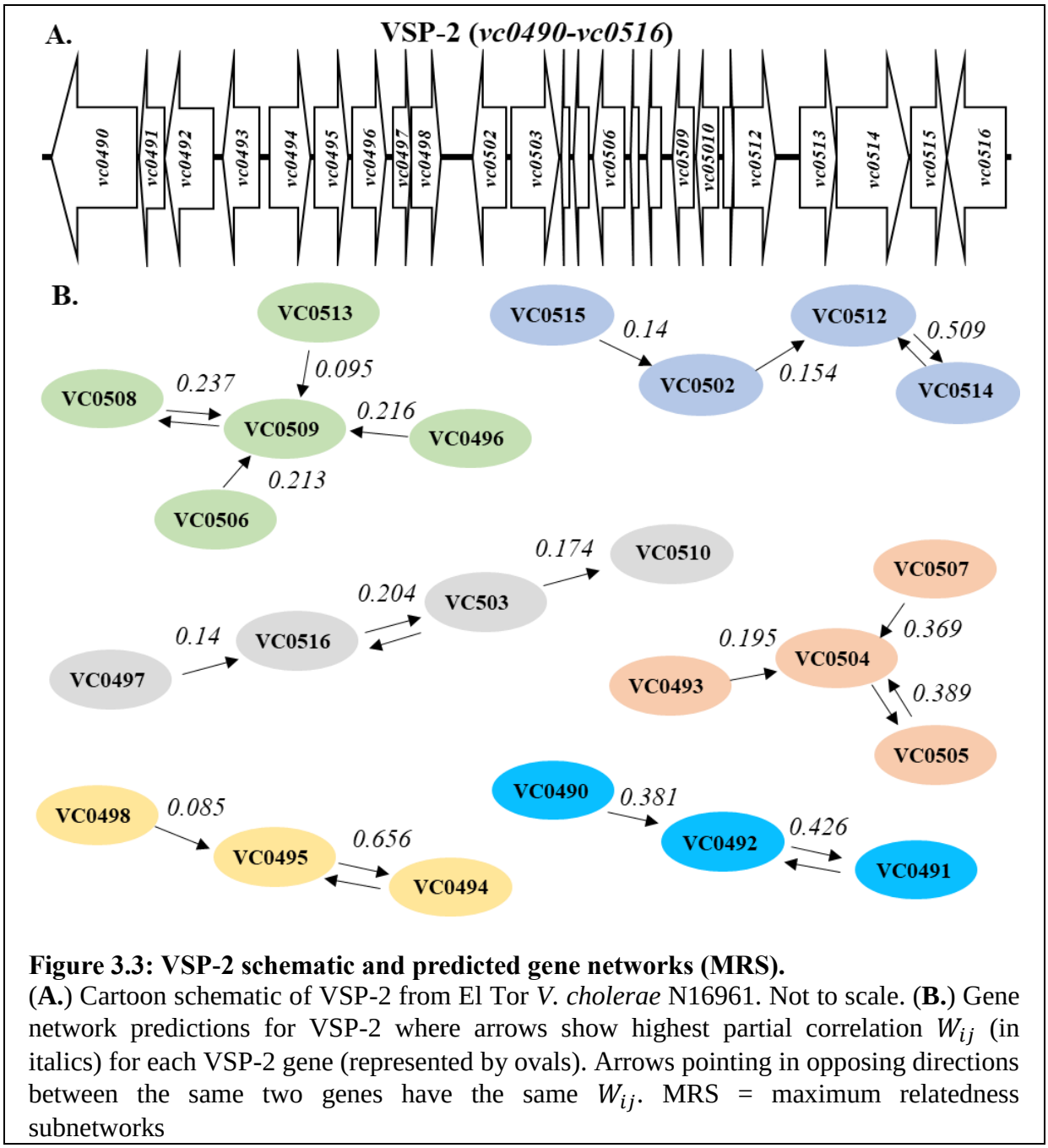

Figure 3.3: VSP-2 schematic and predicted gene networks (MRS).
(**A.**) Cartoon schematic of VSP-2 from El Tor *V. cholerae* N16961. Not to scale. (**B.**) Gene network predictions for VSP-2 where arrows show highest partial correlation W_{ij} (in italics) for each VSP-2 gene (represented by ovals). Arrows pointing in opposing directions between the same two genes have the same W_{ij}. MRS = maximum relatedness subnetworks

where N is the total number of unique species returned from the BLAST search and C_{ij} the number of species with co-occurrence of genes *i and j*. While a Pearson correlation is warranted for a normally distributed binary data set it does not account for indirect correlation. For example, if genes *i* and *j* individually associate with a third gene a Pearson correlation will

incorrectly calculate a correlation between *i* and *j*. To help correct for indirect correlation we calculate a partial correlation w_{ij} from the Pearson r_{ij}:

$$w_{ij} = \frac{P_{ij}}{\sqrt{P_{ii}P_{ij}}},$$

where the (*i, j*) element of the inverse matrix of Pearson r_{ij} is P_{ij}.

The partial correlation correction w_{ij} has the advantage of normalized output to a range of -1 to 1. For example, a w_{ij} of -1 reveals genes *i* and *j* never occur in the same species, while a value of 1 demonstrates genes *i* and *j* always co-occur in the same species. A w_{ij} of 0 is the amount of co-occurrence expected between unrelated genes *i* and *j* drawn from a normal distribution. Using the above-mentioned approach, we calculated a partial correlation value w_{ij} for all genes *i* to *j* in VSP-1 (Supplemental File) and VSP-2 (Supplemental File 4). Next, we chose to use the single highest w_{ij} value for each VSP-1 gene to represent an edge (*i.e.* line) in a visual MRS that suggests putative gene networks. This correlation-based visualization for VSP-1 is shown in Figure 3.2B and VSP-2 in Figure 3.3B. VSP-2 island MRS networks are not explored further in this manuscript.

The MRS for VSP-1 calculated *capV* as co-occurring most often with *dncV*, which reflects CapV activation by the DncV derived secondary messenger cGAMP [18]. Three other VSP-1 genes (*vc0175* [*dcdV*], *vc0180,* and *vc0181*) likewise have strongest correlogy with *dncV* (Supplemental File 3, Figure 3.2), suggesting they too may be candidates for cGAMP regulation or involved in cGAMP signaling. Furthermore, the high w_{ij} of 0.501 shared between *dncV* and *vc0180* suggests the possibility of a shared activity. Previous work using well-classified *E. coli* gene networks showed that partial correlation values $w_{ij} > 0.045$ were highly correlated with

shared biological functions [64]. The predicted gene network composed of *capV, dncV, vc0180, vc0181* is consistent with their CBASS phage defense activity [43]. Curiously, the putative deoxycytidylate deaminase *dcdV*, encoded distal to the CBASS, was found to co-occur with *dncV* (w_{ij} of 0.147) (Figure 3.2, Supplemental File 3). Recognizing this association might be indication of a novel biological function for *dncV* outside of CBASS we sought to understand the biological activity of *dcdV* and its relationship to *dncV*.

3.5.2: Ectopic Expression of *dcdV* Leads to Conditional Cell Filamentation

Based on our previous result that ectopic expression of *dncV* inhibited growth by activation of CapV [18], we hypothesized that ectopic expression of DcdV could have a similar effect. To test this, we performed growth curves in both wild-type El Tor *V. cholerae* C6706str2 (WT), encoding *dncV* and *capV*, and a double VSP island deletion strain (ΔVSP-1/2) lacking *dncV* and *capV*, each over-expressing *dcdV* from a multi-copy plasmid under control of the P_{tac} promoter (pDcdV) or an empty vector control (Vector). Contrary to our prediction, DcdV overexpression did not impact WT growth, but it did reduce growth yield in the ΔVSP-1/2 background compared to the Vector control (Figure 3.4A). Hypothesizing that changes in cell morphology could manifest in reduced growth yield we measured the cellular dimensions of each of the four strains in the growth curve after 8 h using fluorescence microscopy after staining the cell with membrane dye FM4-64. Strikingly, expression from pDcdV in the ΔVSP1/2 background yielded a filamentous cell morphology not found in the other three strains (Figure 3.4B). Quantification of the cell

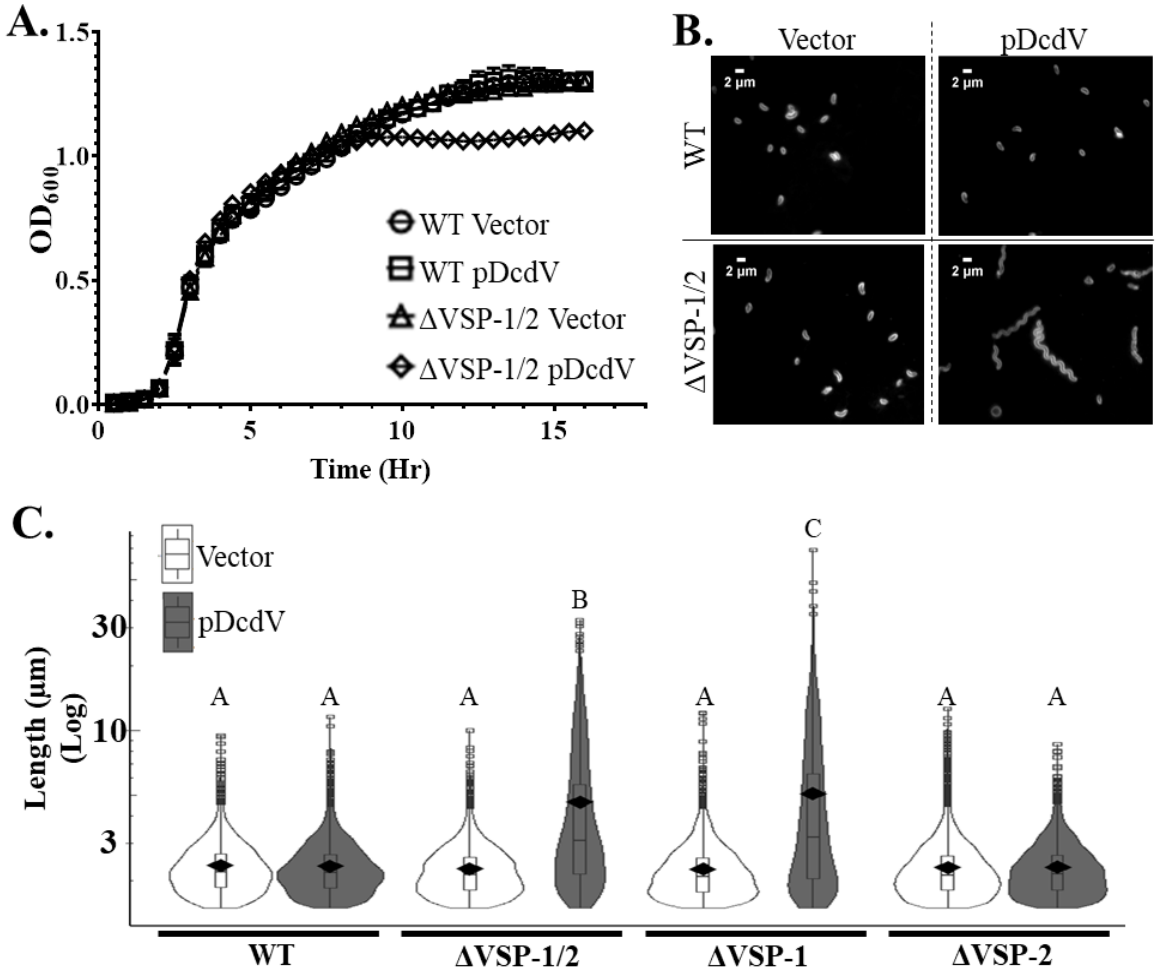

Figure 3.4: Ectopic expression of *dcdV* leads to cell filamentation in the absence of VSP-1 in El Tor *V. cholerae*.

(**A.**) Growth curves of WT El Tor *V. cholerae* and ΔVSP-1/2 strains maintaining an empty vector plasmid (Vector) or P_{tac}-inducible *dcdV* plasmid (pDcdV) grown in the presence of 100 μM IPTG. Error bars represent standard error mean from three biological replicates. (**B.**) Representative images of WT El Tor *V. cholerae* and ΔVSP-1/2 cultures maintaining an empty vector plasmid (Vector) or P_{tac}-inducible *dcdV* plasmid (pDcdV) grown in the presence of 100 μM IPTG for 8 h. Cells were stained with FM4-64 prior to imaging. Scale represents 2 μm. (**C.**) Violin plots of cell length distributions from ~2500-5000 cells pooled from three biological replicates with summary statistics: mean (diamonds), median (horizontal black line), interquartile range (box), and data below and above the interquartile range (vertical lines) Dissimilar letters represent statistically significant differences between strains ($p < 0.05$) determined by Two-Way ANOVA followed by Tukey's post-hoc multiple comparisons test.

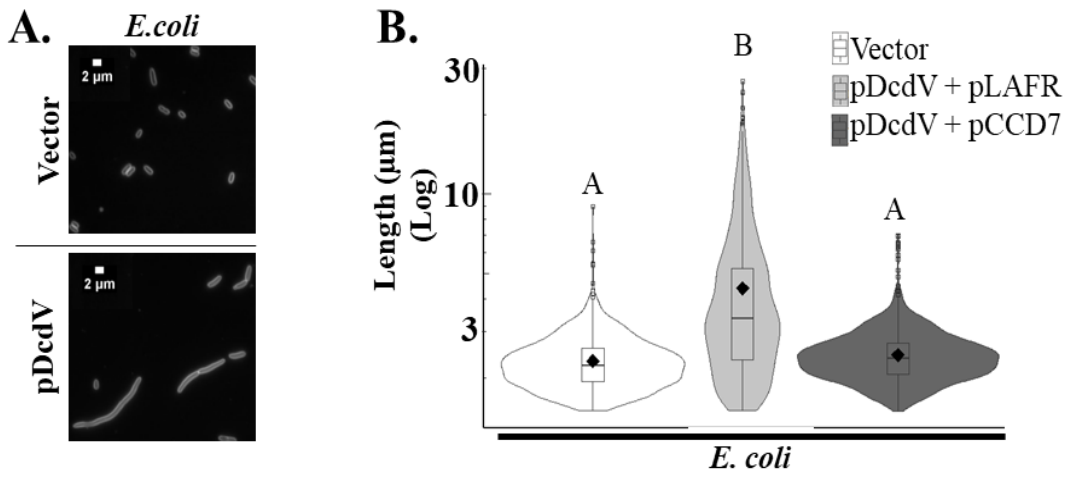

Figure 3.5: Ectopic expression of *dcdV* leads to cell filamentation in *E. coli* that is alleviated by addition of a single copy cosmid containing VSP-1.
(**A.**) Representative images of *E. coli* cultures maintaining an empty vector plasmid (Vector) or P_{tac}-inducible *dcdV* plasmid (pDcdV) grown in the presence of 100 µM IPTG for 8 h. Cells were stained with FM4-64 prior to imaging. Scale represents 2 µm. (**B.**) Distribution of cell lengths measured from three biological replicates of *E. coli* cultures carrying an empty vector (Vector) or P_{tac}-inducible *dcdV* plasmid (pDcdV) in addition to either an empty vector single copy cosmid (pLAFR) or pLAFR containing VSP-1 (pCCD7) grown in the presence of 100 µM IPTG for 8 h. Distributions represent ~1000-2000 cells measured per strain Dissimilar letters represent statistically significant differences between strains ($p < 0.05$) determined by One-Way ANOVA followed by Tukey's post-hoc multiple comparisons test.

lengths measured in these four populations revealed the mean length of ΔVSP1/2 pDcdV had increased significantly to 4.6 µm relative to 2.3 µm measured in the other three strains; ΔVSP1/2 Vector, WT pDcdV, and WT Vector (Figure 3.4C).

Because ΔVSP1/2 lacks both VSP-1 and VSP-2, we performed the same growth curve and image analysis in single island knock-outs (ΔVSP1 and ΔVSP2) to determine the genomic origin of filamentation resistance. The ΔVSP1 pDcdV strain phenocopied the ΔVSP1/2 pDcdV filamentation while ΔVSP-2 maintained a WT cell morphology when ectopically-expressing *dcdV* (Figure 3.4C). Similarly, overexpression of pDcdV in a laboratory strain of *E. coli* also induced cell filamentation which could complemented to a non-filamentous morphology by providing VSP-1 in a single copy cosmid (pCCD7) but not by an empty vector cosmid (pLAFR)

(Figure 3.5A,B.). Taken together, these results indicated that DcdV overexpression severely impacts cell physiology but only in the absence of a VSP-1 encoded resistance factor we named ***d****cdV* **i**nsensitivity **f**actor in ***V****ibrios* (*difV*).

3.5.3: DifV is encoded immediately 5' of the *dcdV* locus in VSP-1

Knowing *difV* was encoded in VSP-1, we sought to identify its precise location by screening partial VSP-1 island deletions for filamentation following pDcdV expression. Three sections of VSP-1 were deleted based on gene orientations and organization including Δ*dcvD-vc0176*, Δ*vc0177-vc0181*, and Δ*vc0182-vc0185* (Figure 3.6A). Of the three partial VSP-1 deletion strains, expression of pDcdV only induced filamentation in the Δ*dcdV-vc0176* mutant and this phenotype could be complemented back to a non-filamentous cell morphology by co-expression of *dcdV-vc0176* from a separate P_{tac} overexpression plasmid (p*dcdV-vc0176*) (Figure 3.6B). To further narrow down the location of *difV*, we constructed further deletions of *dcdV*, *vc0176*, and the 452 nucleotide intergenic region (IG) between the two loci (Figure 3.6A). When challenged with pDcdV expression both strains lacking either *dcdV* or *vc0176* maintained WT cell morphology while the ΔIG mutant became filamentous (Figure 3.6C). Interestingly, ΔIG is not filamentous in the absence of pDcdV expression indicating the laboratory conditions used throughout this study are not sufficient for the

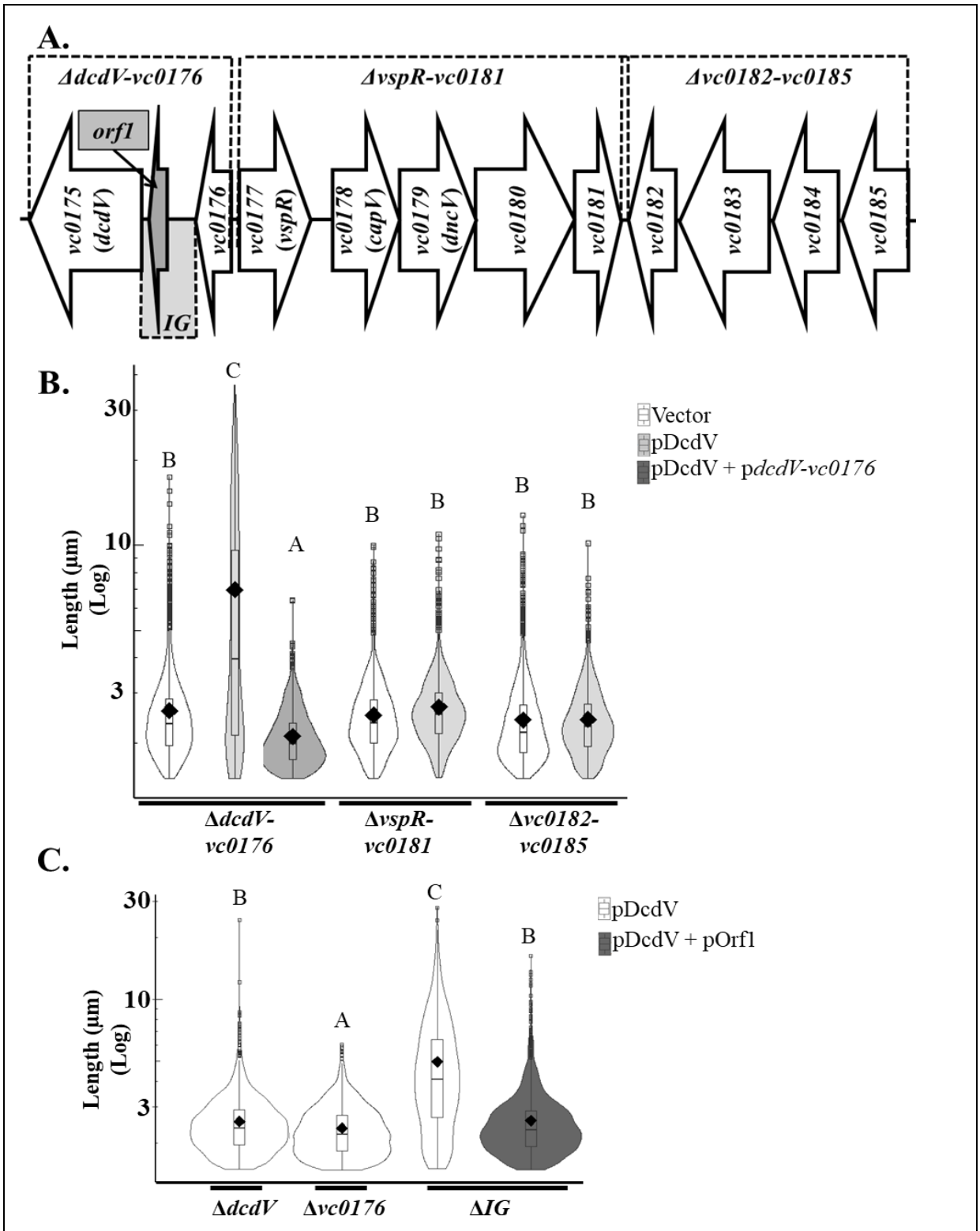

Figure 3.6: The negative regulator of DcdV induced filamentation, DifV, is encoded 5' of *dcdV* within *orf1*.

Figure 3.6 (cont'd)

(**A.**) Schematic representation of VSP-1 from El Tor *V. cholerae* N16961. Hatched lines depict deletions within VSP-1 used to find *difV*. The intergenic region between *vc0175* and *vc0176* (*IG*) is highlighted in light gray while *orf1* is highlighted in dark gray. Not to scale. (**B.**) Distribution of cell lengths measured from three biological replicates of gene deletions within VSP-1 (Δ*dcdV-vc0176,* Δ*vspR-vc0181,* and Δ*vc0182-vc0185*) maintaining an empty vector (Vector) or P_{tac}-inducible *dcdV* plasmid (pDcdV) and complemented with a P_{tac}-inducible *dcdV-vc0176* plasmid (p*dcdV-vc0176*) grown in the presence of 100 μM IPTG for 8 h. (**C.**) Distribution of cell lengths measured from three biological replicates of gene deletions within VSP-1 (Δ*dcdV,* Δ*vc0176,* and the intergenic region between *dcdV* and *vc0176* [*IG*]) maintaining a P_{tac}-inducible *dcdV* plasmid (pDcdV) and complemented with a P_{tac}-inducible *orf1* plasmid (p*orf1*) grown in the presence of 100 μM IPTG for 8 h. Complementation of ΔIG was done using pGBS80 and pGBS87. All cell length distributions represent ~700-3000 cells measured per strain. Dissimilar letters represent statistically significant differences between strains ($p < 0.05$) determined by Two-Way ANOVA followed by Tukey's post-hoc multiple comparisons test.

native *dcdV* locus to recapitulate a filamentous morphology although it is possible that *dcdV* is not expressed in the ΔIG mutant.

Several transcriptomic studies had previously identified IG as a hotspot for unannotated transcriptional activity including a putative transcriptional unit [78], a putative transcriptional start site [79, 80], a non-protein coding RNA [81], and a number of putative sRNAs [82]. Taking these studies into account, we identified a potential small 222 NT ORF (*orf1*) encoded in the same orientation immediately 5' of *dcdV* as a possible *difV* candidate (Figure 3.6A). Remarkably, ectopic expression of *orf1* was indeed sufficient to prevent pDcdV induced filamentation in the ΔIG background (Figure 3.6B). To test if translation of *orf1* was necessary for inhibition of DcdV, we constructed an *orf1* mutant where a stop codon was substituted for the native start codon, $orf1_{stop}$, but surprisingly this construct also prevented pDcdV induced filamentation (data not shown). These results indicate that the 222 NT in *orf1* contains the necessary genetic components for regulating DcdV activity and *difV*

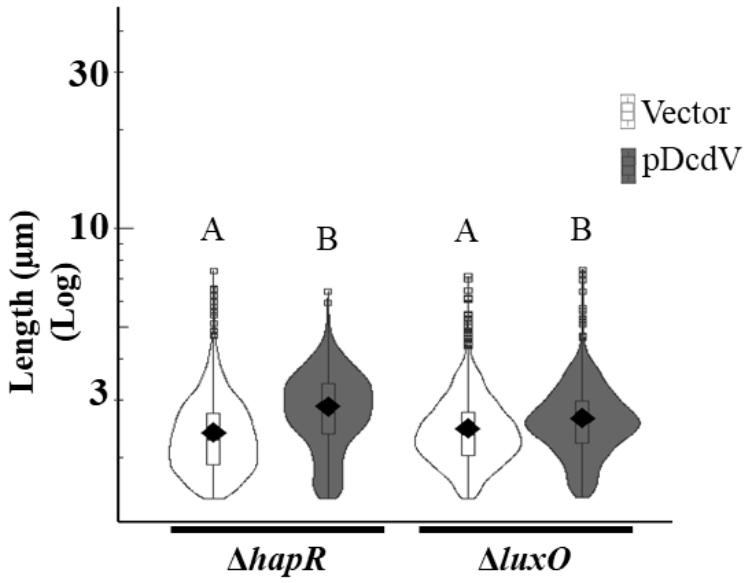

Figure 3.7: DcdV induced filamentation is not controlled by Quorum Sensing.
Distribution of cell lengths measured from three biological replicates of Δ*hapR* and Δ*luxO* maintaining an empty vector (Vector) or P_{tac}-inducible *dcdV* plasmid (pDcdV) grown in the presence of 100 µM IPTG for 8 h. Cell length distributions represent ~1300-3000 cells measured per strain. Letters represent statistical significance across strains ($p < 0.05$) determined by Two-Way ANOVA followed by Tukey's post-hoc multiple comparisons test.

is either a sRNA or a small peptide encoded in this region (< 74 AA) translated from an interior start codon.

To determine if DifV is a sRNA we expressed pDcdV in an El Tor *V. cholerae* mutant with a deletion of *hfq*, a bacterial RNA binding protein that facilitates many sRNA-mediated posttranscriptional gene regulations **(**reviewed in [83]**)**. Hfq plays a critical role in the stability of the four Qrr sRNAs that operate at the fulcrum of quorum sensing, helping to determine whether pathogenic *Vibrios* adopt a low-cell or high-cell density lifestyle [84]. We hypothesized that if DifV were an Hfq dependent sRNA the loss of *hfq* would negatively impact DifV's capacity to limit pDcdV induced filamentation. Interestingly, induction of

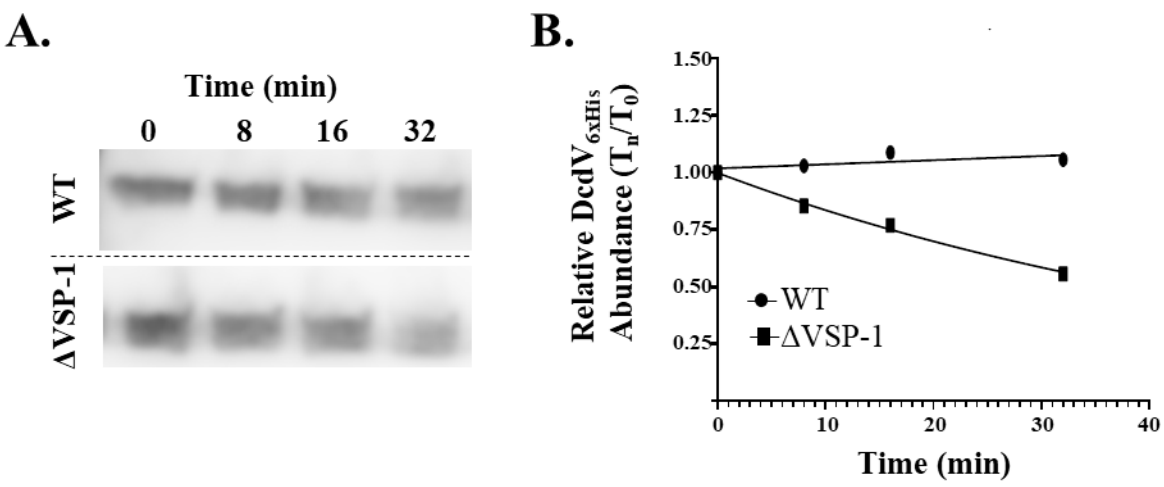

Figure 3.8: DifV inhibition of DcdV induced filamentation is unlikely to occur at the level of translation or protein stability.
(**A.**) Western blot analysis of $DcdV^{6xHis}$ abundance in cell lysate from WT *V. cholerae* and ΔVSP-1 strains maintaining a P_{tac}-inducible *dcdV* plasmid collected 0 – 32 min following chemical inhibition of transcription and translation that was preceded by 2.75 h of growth in the presence of 100 µM IPTG. (**B.**) Relative abundance of $DcdV^{6xHis}$ measured from **A.** and calculated by dividing the protein signal detected at each time point by the initial protein signal (0 min) for each strain. Data represent a single biological replicate.

pDcdV did not cause Δ*hfq* to filament indicating DifV is not an Hfq-dependent sRNA (data not shown). Strains lacking either of the quorum sensing master transcriptional regulators *luxO* or *hapR*, which respectively coordinate low and high cell density gene expression, each presented mild pDcdV induced filamentation (Figure 3.7). While these results do not rule out the possibility that DifV is a sRNA, it is clear that Hfq is not required for its activity and this activity is unlikely to be regulated by either quorum sensing state.

3.5.4: Elucidating the Mechanism of DifV Mediated DcdV Inhibition

The regulatory mechanism DifV imparts on DcdV activity appears to be negative in nature but the molecular action utilized for this control is not clear. Three mechanisms are

equally likely to occur; allosteric regulation of DcdV catalytic activity, inhibition of DcdV translation including transcript destabilization, or enhancing DcdV degradation. In regard to the second mechanism, most sRNAs that regulate translation interact with the ribosome binding site (RBS) of the target mRNA to prevent translation initiation. Translation of DcdV from the pDcdV overexpression plasmid is driven by a non-native RBS, yet DcdV activity from this plasmid can be inhibited by DifV. This finding suggests that the RBS of *dcdV* mRNA is not the target of DifV, although it remains possible that if DifV is a sRNA it could interact with other parts of the *dcdV* mRNA.

To understand if protein production or stability of DcdV was impacted by DifV, we over-expressed a 6 x His C-terminal tagged construct of DcdV ($pDcdV_{6xHis}$) in both WT El Tor and ΔVSP-1 during exponential growth for an hour, chemically halted transcription and translation, and collected protein samples at regular intervals for 32 minutes. Addition of the affinity tag in $DcdV_{6xHis}$ does not affect its activity, as ΔVSP-1 predictably filaments while WT does not when challenged with ectopic expression of this construct (data not shown). Western blot analysis of soluble protein extracts reveal $DcdV_{6xHis}$ is initially abundant in both strains and detectable throughout the time course of the experiment (Figure 3.8A). Because DcdV levels are not decreased at time 0, the function of DifV is not to reduce DcdV production in the cell via a post-transcriptional regulatory mechanism. Surprisingly, the rate of $DcdV_{6xHis}$ degradation is enhanced in the ΔVSP-1 background, while its abundance is relatively unchanged in the WT strain over the course of the experiment (Figure 3.8B). While this experiment was only performed a single time, these results suggest that DifV does not regulate DcdV activity by destabilizing the protein. Additionally, the limited degradation of $DcdV_{6xHis}$ in the WT strain suggests the activity of DifV holds DcdV inactive while simultaneously protecting it from

proteolysis. These experiments rule out two potential mechanism for DifV regulation of DcdV activity, that of decreasing DcdV production or increasing its degradation, and suggest that DifV allosterically controls the enzymatic activity of DcdV.

3.5.5: DcdV induced filamentation requires both conserved NK-family features and the CDA active site

We next sought to determine how DcdV activity leads to filament formation. DcdV is a 532 AA polypeptide composed of two putative domains; an unannotated N-terminal domain and a putative DCD-like C-terminus. We utilized the Phyre2 protein modeling suite [85] and psiBLAST to aid in the identification of conserved structural and primary sequence features present in DcdV that would give clues to its activity. Phyre2 confidently matched the DcdV termini with two independent protein families; the N-terminus contained features of Nucleoside/Nucleotide Kinases (NK) super family enzymes, while the C-terminus closely aligned to structural homologs of DCD enzymes which are part of the zinc-dependent CDA family (Figure 3.9A).

NK enzymes catalyze the reversible phosphotransfer of the γ-phosphate from a nucleotide triphosphate donor to a diverse group of substrates, depending on the enzyme class, including deoxynucleotide mono-phosphates. Three structural features commonly found in these enzymes include a P-loop/Walker A motif {GxxxxGK[ST]} and a two helical LID module that together stabilize the donor nucleotide triphosphates, and a Walker B motif {hhhh[D/E], where h represents a hydrophobic residue} is partly involved in coordinating Mg^{2+} [86, 87]. Interrogation of the Phyre2 DcdV model and psiBLAST primary sequence alignments revealed these three features are likely present in the N-terminal domain (Figure 3.9B). This observation

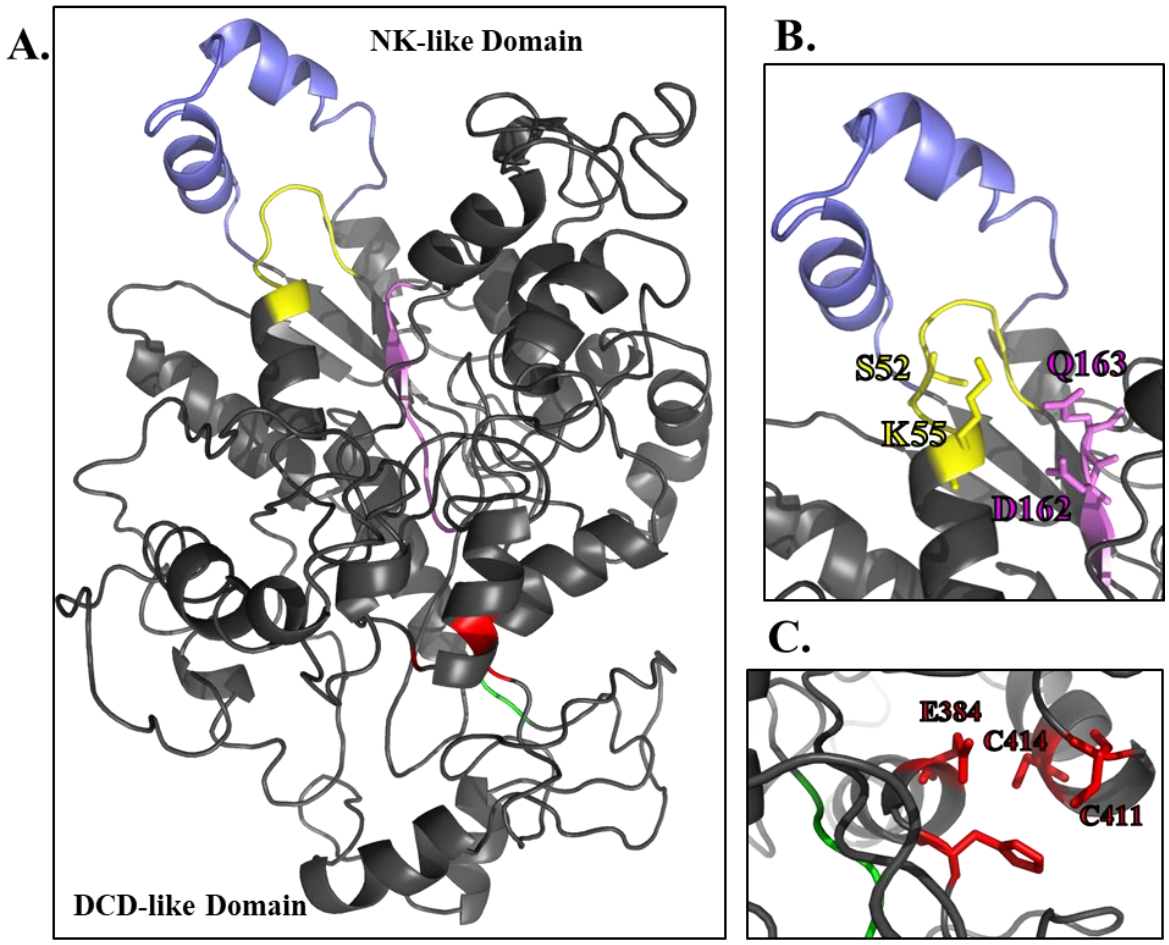

Figure 3.9: Phyre2 model of DcdV and conserved features of the N-terminal NK domain and C-Terminal DCD domain.
(**A.**) Cartoon model of DcdV from Phyre2. NK-like domain features: Walker A = yellow, Walker B = violet, and the LID module = blue. DCD-like domain features: Zn^{2+} active site = red, and the non-conserved dNTP regulator motif: green. (**B.**) Close-up view of the residues identified for mutation within conserved features of the NK-like domain. (**C.**) Close-up view of the residues identified for mutation within conserved features of the DCD-like domain.

suggests the N-terminus of DcdV is an NK super family domain involved in binding nucleotide substrates and performing a phosphotransfer reaction.

The zinc-dependent CDA active site motif, [CH]-A-E-$X_{(21-37)}$-P-C-$X_{(2-8)}$-C [88], is highly conserved in the C-terminal domain of DcdV (Figure 3.9C). The constellation of residues that make up the Zn^{2+} binding pocket is composed of three critical amino acids; H382, C411, and

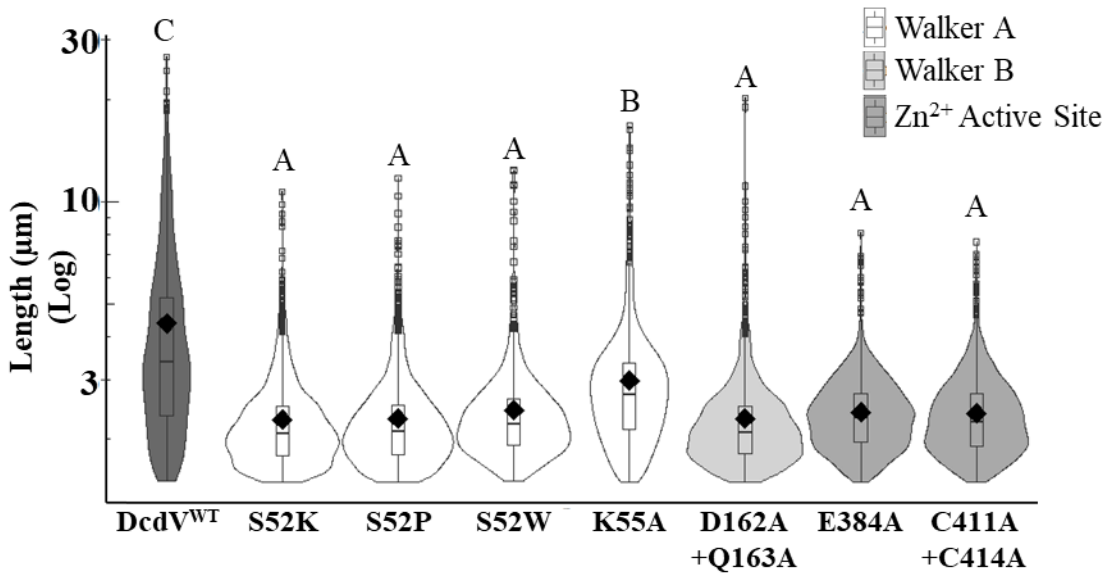

Figure 3.10: Conserved features of the N-terminal NK domain and C-Terminal DCD domain are required for DcdV induced filamentation in *E. coli*.
Distribution of cell lengths measured from three biological replicates of *E. coli* maintaining a P_{tac}-inducible *dcdV* plasmid ($DcdV^{WT}$) or a variety of DcdV variants grown in the presence of 100 µM IPTG for 8 h. *E. coli* cell length distribution for WT DcdV is depicted in dark gray while variants in the Walker A motif are in white, Walker B in light gray, and the DCD Zn^{2+} active site are in medium gray. Cell length distributions represent ~1700-3000 cells measured per strain. Letters represent statistical significance across strains ($p < 0.05$) determined by Two-Way ANOVA followed by Tukey's post-hoc multiple comparisons test.

C414. This Zn^{2+} is required for the catalytic deprotonation of water by E384 for the hydrolytic deamination of a cytosine base to uridine. Members of the CDA enzyme family specifically catalyze the deamination of a diverse assortment of cytosine containing substrates including free cytidine, deoxycytidine mono and trisphosphate, RNA and ssDNA polynucleotides[51, 52, 89, 90]. DCD enzymes are unique among the CDAs for their allosteric regulation by dCTP and dTTP which activate and repress the deamination of dCMP, respectively, through a G[Y/W]NG allosteric site motif [91, 92]. Interestingly, DCD allosteric regulation by dNTPs may not be preserved in DcdV as this motif is composed of a divergent GCND.

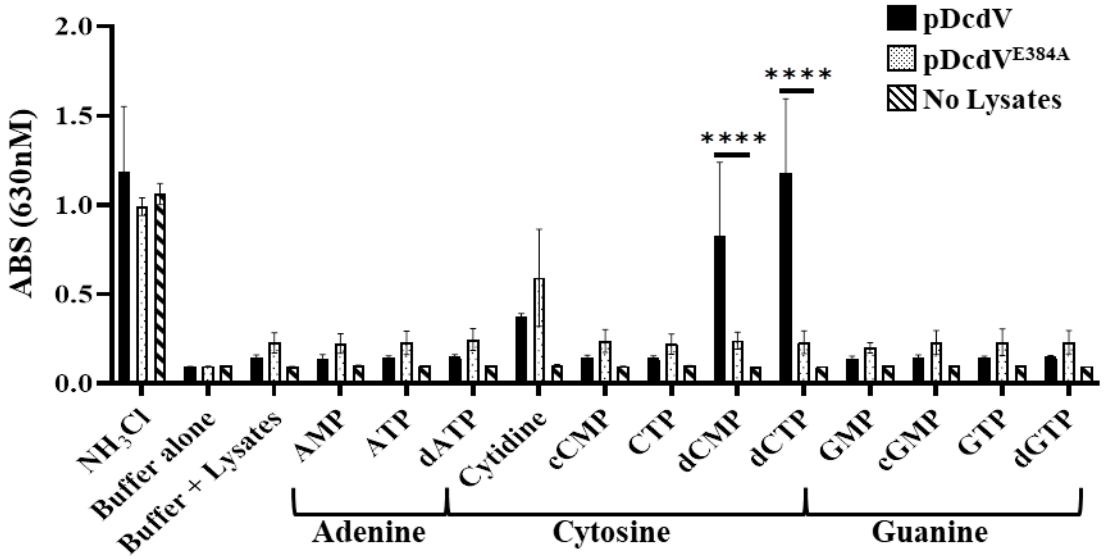

Figure 3.11: *E. coli* lysates containing DcdV preferentially deaminate dCMP and dCTP substrates.
Lysates collected from *E. coli* expressing DcdV or DcdV^{E384A} and a "no lysate" control incubated with 12 nucleic acid substrates (1.9 mM NH_3Cl as a positive control, 37.7 mM cytidine, and 7.5 mM for all other substrates). The evolution of NH_4^+ resulting from substrate deamination was detected by measuring the solution ABS_{630} after a Berthelot's reaction in microtiter plates. Data represent the mean and standard deviation of three biological replicate lysates and duplicate "no lysate" controls. Brackets and asterisks indicate statistical significance ($p<0.05$) determined by Two-Way ANOVA followed by Tukey's post-hoc multiple comparisons test. Error bars represent standard deviation.

Hypothesizing that one of the two domains present in DcdV was responsible for cell filamentation in the absence of *difV* we made site-specific mutations in the conserved residues in both the NK and CDA domains. Six variant constructs were generated from the NK domain; four in the Walker A motif (pDcdV^{S52P}, pDcdV^{S52W}, pDcdV^{S52K}, and pDcdV^{K55A}) to introduce steric bulk and interfere with hydrogen bonding and a single variant with a double substitution in the Walker B motif (pDcdV^{D162A,Q163A}). Two variants were

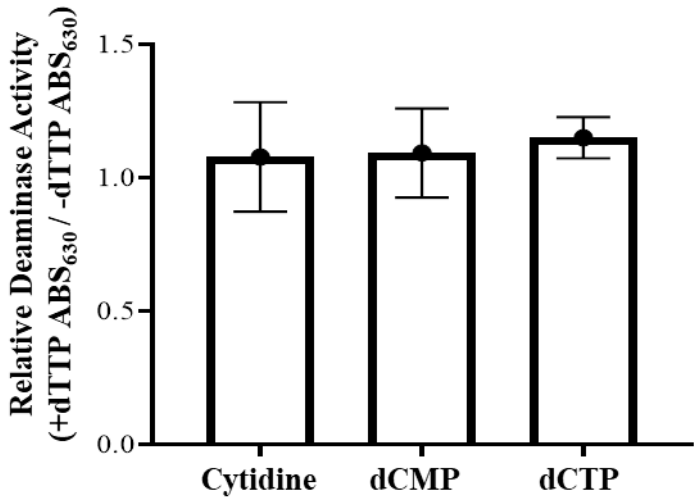

Figure 3.12: dTTP does not inhibit DcdV deamination activity in *E. coli* lysates.
Lysates collected from *E. coli* expressing DcdV incubated with or without 7.5 mM dTTP and either 75 mM cytidine, 7.5 mM dCMP, or 7.5 mM dCTP. The evolution of NH_4^+ resulting from substrate deamination was detected by measuring the solution ABS_{630} after a Berthelot's reaction in microtiter plates. The relative deaminase activity was calculated by dividing the ABS_{630} of the +dTTP reaction by the no dTTP control reaction for each lysate. Data represent the mean and standard deviation of three biological replicate lysates.

constructed in the CDA active site; a double substitution of both C411A and C414A ($pDcdV^{C411A,C414A}$) to abrogate Zn^{2+} binding, and a E384A substitution ($pDcdV^{E384A}$) to inhibit the deprotonation of water required for the hydrolytic deamination of cytosine. Surprisingly, only ectopic expression of $pDcdV^{K55A}$ was capable of inducing mild filamentation while the remaining seven variants, irrespective of domain or feature, failed to induce filamentation in *E. coli* (Figure 3.10). Global destabilization of DcdV resulting from these disparate substitutions is unlikely to explain the loss of DcdV induced filamentation in all cases as 6 x His-tagged constructs of the DCD variants maintain solubility and similar abundance to WT 6 x His-tagged DcdV (data not shown). Further investigation of the Walker A variant, K55A, will help determine whether the NK domain activity has been inhibited entirely or if it remains mildly

active. Taken together, these results indicate the functions performed by both the NK and CDA domain are required for DcdV induced filamentation.

3.5.6: The CDA domain of DcdV utilizes both dCMP and dCTP substrates and may not be allosterically regulated by dTTP

Enzymes belonging to the CDA family have a high degree of specificity for their substrates [92]. To determine the substrate preference of DcdV we expressed affinity-tagged *dcdV* and the CDA active site variant *dcdV*^{E384A} cloned under the T7 promoter on a high-copy plasmid in BL21(DE3) *E. coli*. Soluble lysates from each strain were incubated with a battery of amine containing nucleic acid substrates and monitored for the evolution of NH_4^+ using a colorimetric assay in microtiter-plates. Of the 12 substrates tested, lysates containing DcdV produced ammonium when incubated with dCMP and dCTP, which was not detected in lysates containing DcdV^{E384A} (Figure 3.11). Both strains showed similar levels of cytosine deamination indicating the presence of endogenous *E. coli* CDA activity (Figure 3.11). Interestingly, we found that deamination of dCMP and dCTP by DcdV lysates was not inhibited by the addition of dTTP, a common negative allosteric regulator of DCD enzymes (Figure 3.12). Together, these results demonstrate that DcdV is a DCD that specifically catalyzes the deamination of both dCMP and dCTP to produce dUMP and dUTP, respectively, and this activity was resilient in the presence of excess dTTP.

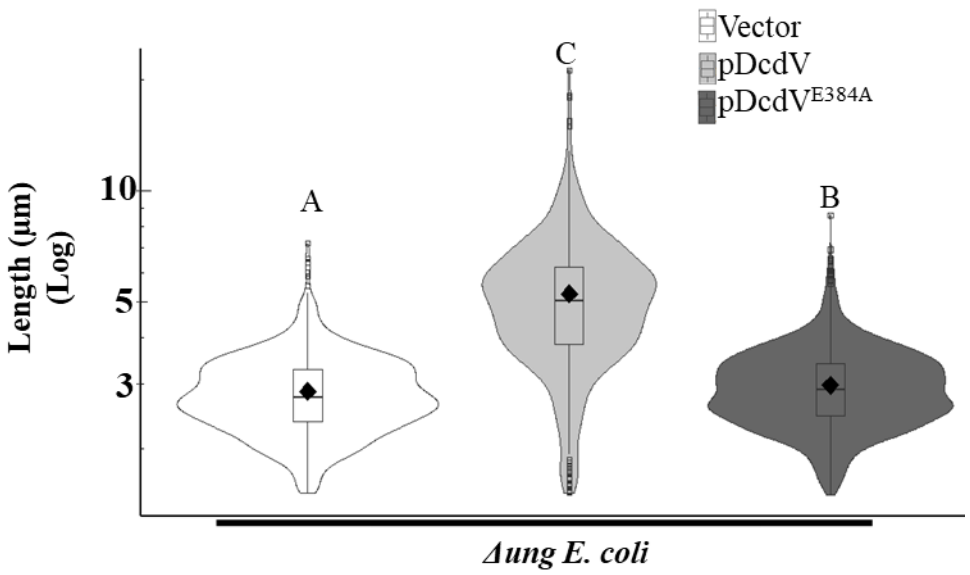

Figure 3.13: Ung activity is not required for DcdV induced filamentation in *E. coli*. Distribution of cell lengths measured from three biological replicates of Δ*ung E. coli* maintaining an empty vector plasmid (Vector) or P_{tac}-inducible *dcdV* plasmid (pDcdV) or P_{tac}-inducible *dcdV* inactive variant plasmid ($DcdV^{E384A}$) grown in the presence of 100 µM IPTG for 8 h. Cell length distribution for Vector are filled in white, pDcdV in light gray, and $pDcdV^{E384A}$ in dark gray. Cell length distributions represent ~1700-3000 cells measured per strain. Letters represent statistical significance across strains ($p < 0.05$) determined by Two-Way ANOVA followed by Tukey's post-hoc multiple comparisons test.

3.5.7: The DCD domain increases the genomic incorporation of dUMP

The accumulation of dUTP is problematic for most organisms as DNA polymerases poorly discriminate between dUTP and dTTP [59]. This leads to the erroneous incorporation of dU into replicating strands of DNA, which is normally recognized and removed by the base excision repair (BER) pathway [93]. Excessive incorporation of dU in DNA can lead to over stimulation of BER resulting in double strand breaks and ultimately cell death. In the absence of uracil-DNA glycosylase (Ung or UDG), which initiates BER by cleaving uracil bases from

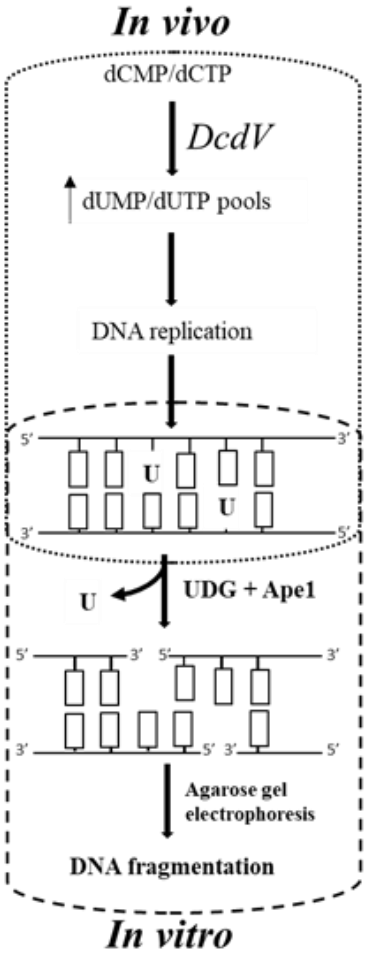

Figure 3.14: Flow chart for detecting incorporation of dU in Δ*ung E. coli* genomic DNA. From top to bottom: The *in vivo* activity of DcdV increases the cellular abundance of dUMP and dUTP leading to the miss incorporation of dU into genomic DNA following replication. In the absence of *ung*, genomic dU cannot be efficiently removed. The *in vitro* addition of purified UDG and ApeI to isolated genomic DNA leads to the hydrolytic cleavage of uracil and the phosphodiester backbone at the resulting abasic site. The level of DNA fragmentation following agarose gel electrophoresis is indicative of the relative abundance of genomic dU. U = uracil, squares represent generic non-uracil bases, dotted lines represent *in vivo* activity, hatched line depict *in vitro* processes.

DNA, a cell is unable to recognize and efficiently remove dU from its genome [94]. Therefore, the accumulation of genomic dU in a *Δung* background can be used to infer the relative abundance of intracellular dUTP. To test if DcdV increases the pool of available dUTP and its

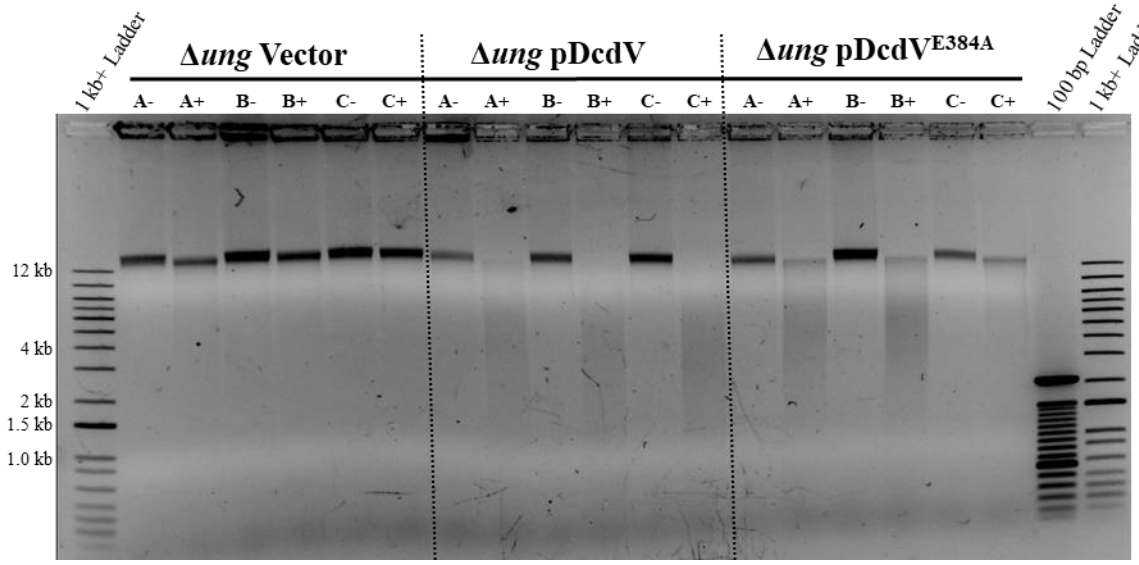

Figure 3.15: DcdV activity increases genomic dU incorporation in Δ*ung E. coli*.
Agarose gel electrophoresis of 2.5 µg genomic DNA isolated from Δ*ung E. coli* maintaining an empty vector plasmid (Vector), P_{tac}-inducible *dcdV* plasmid (pDcdV), or P_{tac}-inducible *dcdV* DCD active site mutant ($pDcdV^{E384A}$) grown in the presence of 100 µM IPTG for 8 h. A, B, and C represent genomic DNA isolated from three biological replicates for each strain. - and + indicate genomic DNA was incubated in the absence or presence of purified UDG and ApeI, respectively.

incorporation into the genome, we expressed pDcdV, $pDcdV^{E384A}$, and Vector in a Δ*ung* strain of *E. coli*, collected genomic DNA, and measured cell morphologies after 8 h. Interestingly, the *Δung E. coli* mutant expressing pDcdV was elongated, showing that Ung dependent removal of uracils in the DNA was not necessary for the filamentation phenotype (Figure 3.13). As observed before, the $DcdV^{E384A}$ variant was unable to induce filamentation (Figure 3.13). The isolated DNA was incubated with purified UDG to remove incorporated uracil bases and the endonuclease ApeI to cleave the DNA at abasic sites followed by agarose gel electrophoresis to investigate DNA fragmentation (Figure 3.14). Isolated DNA from *Δung* expressing pDcdV treated with UDG and ApeI was highly fragmented relative to the Vector control strain, indicating a dramatic increase in the abundance of genomic dU resulting from DcdV activity

(Figure 3.15). Surprisingly, despite being unable to catalyze the formation of dUTP (Figure 3.11), the pDcdV^{E384A} Δ*ung* genomic DNA also had increased incorporation of dU in its genome relative to the Vector strain, although it inserted less uracil than expression of WT DcdV (Figure 3.15). Taken together these results demonstrate that DcdV activity increases the abundance of dUTP in the cell by two mechanisms; DCD activity and a second pathway originating from the NK-domain yet to be elucidated. While the relative comparison of genomic dU through digestions and gel electrophoresis is an indirect method for analyzing nucleotide abundance it is reasonable to expect more quantitative methods will reveal an imbalance in nucleotide pools, promoting dUTP formation, that coincide with contributions from each domain.

3.6: Discussion

Characterization of the ~36 genes encoded within the El Tor *V. cholerae* VSP-1 and 2 genomic islands is still in its infancy and uncovering their contributions to bacterial fitness may help elucidate the origin and persistence of the seventh cholera pandemic. While the general architecture and composition of VSP-1 and 2 are mostly preserved in *Vibrio sp.*, it is clear from our correlogy study that gene clusters from these islands are widely distributed across the bacterial phyla. In support of this method, our study accurately identified a correlog composed of the VSP-1 antiphage CBASS system (*capV-dncV-vc0180-vc0181*) (Figure 3.2). Interestingly, this bioinformatic survey also revealed *dncV*, whose production of cGAMP is critical for CBASS induced abortive replication, is frequently found in genomes with the previously uncharacterized gene *vc0175* (*dcdV*) (Figure 3.2). Here, we showed that DcdV contains a functional DCD domain that catalyzes the deamination of deoxycytidine nucleotides and a putative NK-like domain. In total, DcdV activity appears to distort deoxynucleotide pool homeostasis which manifests in a filamentous cell morphology. This abnormal morphology is abrogated by a novel

post-translational mechanism by *difV*, a genetic component within a 222 NT region immediately 5' of the *dcdV* locus in VSP-1 (Figure 3.6A), though the details of this process remain to be fully understood.

The deamination of dCTP is canonically performed by non-zinc dependent enzymes [58] making the dual substrate repertoire of dCMP and dCTP in DcdV a rare trait. The only other DCD reported to work on both dCMP and dCTP is biDCD from chlorovirus PBCV-1 [51]. PBCV-1 has a double stranded DNA genome with a ~40% G + C content while its microalgae host, *Chlorella* NC64A, has a genomic content of ~67% G + C. During infection, biDCD, along with other viral peptides, rapidly converts host dC nucleic acids pools into dTTP thus facilitating replication of the A + T rich viral genome. Like many DCD enzymes, biDCD is allosterically regulated by dCTP and dTTP that activate and inactivate the deaminase, respectively, providing a means to fine-tune the pool of available dNTPs [91, 92]. Interestingly, DcdV does not appear to have maintained the allosteric nucleotide binding site, suggesting it is unlikely to modulate steady state nucleotide concentrations. In support of this, addition of excess dTTP to cell lysates did not alter the catalytic activity of DcdV towards dCMP or dCTP in trial experiments (Figure 3.11). Though the biochemical complexity of cell lysates may obscure potential regulation, our results suggest DcdV deamination of dC nucleic acids proceeds unencumbered by a feedback mechanism meant to preserve dC substrates for DNA replication.

In lieu of a conserved deoxynucleotide allosteric site, DcdV is regulated post-translationally by DifV. Regardless of whether DifV is a peptide or a regulatory sRNA, to our knowledge this type of regulation is unique among the CDA-family. The spacing, orientation, and relationship of *difV* and *dcdV* are vaguely reminiscent of Type 2 and Type 3 [95] Toxin-Antitoxin (TA) systems found across the bacterial phyla. These TA types utilize a small peptide

or sRNA antitoxin to allosterically inactivate a co-transcribed toxin peptide and degradation of the antitoxin results in the liberation of the toxin (reviewed in [96]). Utilization of a Type 2 or 3 TA system-like regulation would allow for the targeted deployment of DcdV catalytic activity under specific conditions by the degradation of DifV. Additionally, the utility of deaminase domain containing proteins as potent toxins has been illustrated by their inclusion in bacterial polymorphic toxin systems used for extracellular kin selection [97, 98]. However, unlike these systems the genomic architecture and lack of a clear translocation mechanism suggest that *dcdV* and *difV* are not utilized extracellularly. Interestingly, under our experimental conditions the presence of genomic *difV* is sufficient to inhibit ectopically expressed *dcdV*, while loss of *difV* does not result in filamentation from genomic *dcdV* activity. This observation implores further experimentation to understand the transcription and biological conditions under which *difV* and *dcdV* are expressed and active. Furthermore, outstanding questions remain as to the chemical makeup of DifV and the mechanism of action utilized to inhibit one or both DcdV domain activities.

Cell filamentation is a hallmark of thiamine-less death (TLD), observed in prokaryotes and eukaryotes, which arises from a sudden loss of thiamine during robust cellular growth [99]. Insufficient dTTP substrate during DNA replication leads to the disintegration of replication forks and the accumulation of unresolved ssDNA that cause genomic instability and cell death [100]. Interestingly, this phenomenon is not limited to dTTP as dGTP starvation elicits a similar response in *E.coli* and is also hypothesized to occur when other deoxynucleotide substrates become disproportionately scarce [101]. It has also been shown that even modest changes in

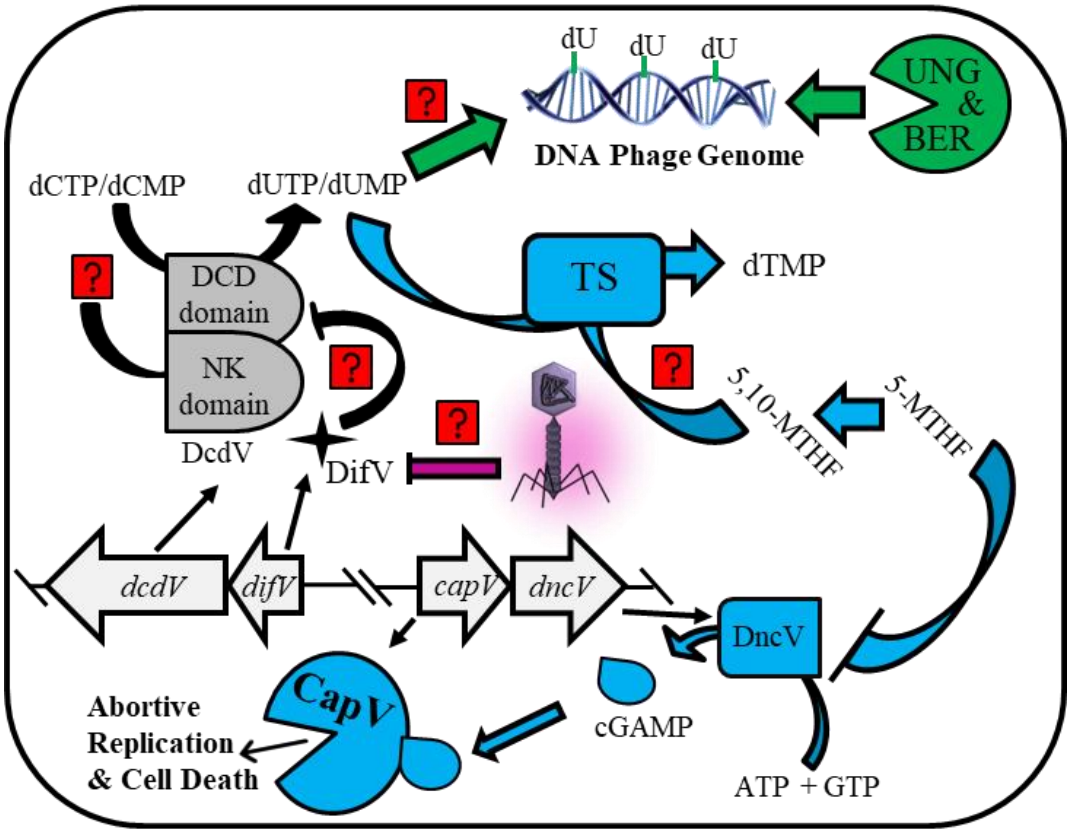

Figure 3.16: Model for hypothetical DcdV activity in phage immunity.
The catalytic activity of DcdV (dark gray) leads to an accumulation of dUTP/dUMP. Two hypothetical and complementary pathways for this activity to contribute to phage defense are; DncV dependent (blue) and DncV independent (green). Major questions/assumptions in these pathways are marked with a question mark (red). The DncV dependent pathway increases the synthesis of deoxyribonucleotides, depicted here as the biosynthesis of dTMP by thymidylate synthase (TS) which utilize 5,10-MTHF. This reduces the available 5-MTHF and de-represses DncV activity. Liberated DncV produces cGAMP to activate CapV and initiate abortive replication. The DncV independent pathway involves the corruption of replicating DNA phage genomes by misincorporation of dUTP substrates, thus targeting these genomes for scrutiny by endogenous DNA repair mechanism. A possible mechanism for the activation of DcdV is the phage dependent inhibition of DifV (purple).

dNTP pools can dramatically increase mutation rates [102] and an abundance of dUTP substrates can lead to erroneous incorporation of genomic dU [59]. In the case of DcdV, it is conceivable the observed filamentation phenotype is a consequence of a TLD-like corruption of deoxynucleotide pools. Our evidence suggests the disparate activities encoded in the DCD and

NK domains have been evolutionarily fused to enhance the formation and persistence of dUTP (Figure 3.15) to the detriment of both dCTP and dTTP pools, respectively (Figure 3.1). Inactivating mutations in conserved catalytic features of either domain are sufficient to inhibit DcdV induced cell filamentation (Figure 3.10) suggesting the native pyrimidine biosynthesis architecture is robust enough to handle a single perturbation in one deoxynucleotide pool, but not a simultaneous assault on both. We speculate that the NK domain catalyzes phosphotransfer between nucleotide substrates promoting formation of dUTP, which then increases the incorporation of dU into the genome in the absence of the DCD domain (Figure 3.15). Future experiments using more quantitative methods to measure deoxynucleotide pools will be critical to understanding the magnitude of each domain's contribution to this process.

Overall, our evidence suggests that DcdV and DifV function in concert to prepare a bacterium for rapid adaptation to a specific cellular condition that warrants a dramatic change in deoxynucleotide pools and this process is likely detrimental to the cell's own survival. Given that *dcdV* co-occurs with *dncV*, of the CBASS system, and phage defense systems often co-occur in defense islands [44, 57], we hypothesize DcdV and DifV compose a novel
phage defense mechanism. Based on our results, DcdV activity could facilitate phage defense in two ways; increasing the mutational frequency and incorporation of dU in replicating viral genomes (DncV Independent) and de-repressing DncV cGAMP synthesis (DncV Dependent) (Figure 3.16). Even moderate destabilization of dNTP pools has profound mutagenic consequences [102] and the misincorporation of dU into rapidly replicating viral genomes could flag them for scrutiny and processing by native DNA repair mechanisms. To this point the phage φ29 of *Bacillus subtilis* deploys the protein p56 to inhibit host UNG activity thus preventing BER from targeting replicating viral genomes that have incorporated dU [103].

Synthesis of cGAMP by DncV has been shown to be allosterically inhibited by certain species of folates including 5-methyltetrahydrofolic acid (5-MTHF) [23]. The two precursors of 5-MTHF, 10-formyltetrahydrofolate and 5,10-MTHF, are utilized for the de novo biosynthesis of purines and dTMP, respectively [50]. In the face of DcdV activity a native response to restore dNTP homeostasis could rapidly deplete folates leading to the de-repression of DncV cGAMP synthesis and initiate abortive replication (Figure 3.16). These pathways are not mutually exclusive and their combined effects could interfere with viral genome replication while catalyzing the DncV initiated abortive replication process. Further experiments are required to understand the relationship between *dncV* and *dcdV* and how the activity of one may influence the other, including de-repression by folates or DifV, respectively.

Chapter 4: Concluding remarks

4.1: Conclusion and Significance

Since 1817, pathogenic strains of *V. cholerae* have caused seven pandemics of the diarrheal disease cholera [3]. The seventh pandemic began in 1961 and continues to be perpetuated by strains of the El Tor biotype. Despite their genetic similarity and a shared ancestor, the El Tor biotype displaced the previous classical pandemic biotype in both environmental and clinical reservoirs. It is hypothesized that El Tor's acquisition of the genomic islands VSP-1 and 2 in the mid-20th century facilitated El Tor's pandemic evolution [3]. When considering the role of the VSP islands in El Tor *V. cholerae* three questions remain to be answered: what biological functions do they encode, what is their utility to bacterial fitness, and where did they come from? Through the work presented in this thesis I have endeavored to help answer these questions.

4.1.1: DncV – cGAMP – CapV

Following the identification of DncV in 2012, the cdN cGAMP remained an orphan second messenger for six years. Our discovery of the phospholipase CapV in 2018, described in Chapter 2, and its allosteric regulation by cGAMP represented the first cGAMP signaling network described in bacteria [18]. By connecting the activity of DncV to CapV we facilitated the identification of other discrete signaling networks across the bacterial phyla involving the *dncV*-like family of nucleotydil tranferases (CD-NTases or SMODS) [27, 28]. The work by Whiteley et al. drove a paradigm shift in the field of cdN signaling by expanding the repertoire of cdN second messengers and demonstrating the catalytic flexibility of the CD-NTase family. In short order the laboratory of Rotem Sorek connected the cGAMP activation of *V. cholerae* CapV, and an orthologous *E. coli* system, with a novel phage defense system they named

CBASS [43]. As the field of bacterial cdNs has rapidly expanded in the past two years, there remain a number of critical questions related to regulation and deployment of *capV-dncV-vc0180-vc0181* in El Tor *V. cholerae* and CBASS systems generally, and I will briefly discuss these topics.

4.1.2: DcdV and DifV

Prediction of the possible gene network between *dncV* and the putative deoxycytidylate deaminase *vc0175*, now called *dcdV*, may expand our understanding of CBASS systems and their connection to genes not encoded in a shared operon (Supplemental File 3, Figure 3.2). DCD enzymes typically play a role in the *de novo* biosynthesis of dTMP by providing the necessary dUMP building blocks (Figure 3.1). DcdV appears to be an abnormal DCD enzyme in its ability to perpetually deaminate dC substrates in spite of excess dTTP (Figure 3.12). Interestingly, the novel regulatory role of DifV, a previously unannotated ORF encoded 5' of *dcdV*, may have replaced the need for dTTP inhibition and may be important for the appropriate deployment of DcdV. Additionally, fusion of the DCD and NK domains found in DcdV is a novel configuration for CDAs, and their shared role in the likely production of dUTP is deleterious for the bacterium in which it is expressed (Figure 3.10, Figure 3.15). The activity of DcdV is not completely dependent on cGAMP, as the enzyme is active in the absence of DncV. Whether or not DcdV/DifV is influenced by cGAMP, or this system itself influences DncV activity, remains to be addressed. I will briefly address this and other outstanding experimental questions in the arena of DncV, DcdV, and DifV in a later section.

4.1.3: Correlogy and VSP island gene networks

Thus far, five VSP island genes have been experimentally analyzed (*dncV, vspR, capV, dcdV,* and *difV*) [9, 18] and the work presented in this thesis has made fundamental contributions to understanding four of them. Despite these efforts, ~31 ORFs across the VSP islands remain unstudied. Our collaborators in the laboratory of Eva Top at the University of Idaho developed a bioinformatic tool, Correlogy, to predict conserved gene pathways within the VSP islands by identifying the co-occurrence of individual genes across bacterial genomes (Figure 3.2, Figure 3.3, Supplemental File, Supplemental File 4). This analysis accurately identified two known VSP island biological networks; the VSP-1 encoded CBASS system [43] and the three gene operon *vc0490-vc0491-vc0492* encoded in VSP-2 not discussed in this thesis but under investigation in the Waters lab. Correlogy also predicted numerous additional gene networks within the VSP islands that will help to prioritize the further characterization of the function of these genes. Validation of these predicted gene networks may help to uncover the origins of the VSP islands and illuminate the evolutionary pathways that converged to generate them in the El Tor biotype. Additionally, Correlogy will be publicly available and this tool can be used to explore potential gene networks in any collection of bacterial genes.

4.1.4: Origins and exclusivity of the VSP islands

Arguably, one of the most interesting questions that still remains to be answered is; why the El Tor biotype? For reasons yet to be elucidated, VSP-1 and 2 were exclusively acquired by the progenitor strains of the El Tor biotype. Recognizing that the evolution of seventh pandemic El Tor *V. cholerae* and its acquisition of the VSP islands occurred in the presence of classical strains begs the question; what was special about the progenitor El Tor strains that led to

acquisition of these islands when the classical strains did not? By combining the biological functions encoded within the VSP islands with genomic comparison between the biotypes we will likely gain insight into this question.

4.2: Future Directions

4.2.1: Further characterization of El Tor *V. cholerae* CBASS

4.2.1.1: DncV

A key feature of CBASS systems yet to be identified is the mechanism by which they are appropriately deployed. For efficient phage defense it is likely that all the necessary protein machinery (CD-NTase, effector, etc.) must be translated prior to phage insult. In the case of El Tor *V. cholerae*, *capV* and *dncV* expression is regulated by QS suggesting that at high-cell density both proteins are present in the cytoplasm at concentrations sufficient to kill the cell. In support of this hypothesis, ectopic expression of *dncV* leads to cell death in El Tor by activating endogenously produced CapV (Chapter 2). A mechanism must therefore be established that prevents endogenously produced DncV from inappropriately activating CapV. In the case of HORMA CBASS, the formation of the active HORMA:CdnC complex requires the presence of an unidentified short peptide sequence of unknown origin that is abundant during infection [46]. While DncV activity is regulated by a variety of folates species, including 5-MTHF, it is unclear why DncV has evolved to be regulated by them [23]. Hypotheses regarding the regulation of DncV by post translational-modification by VC0180 and a connection to DcdV activity are briefly explored later. Finally, in order not to inappropriately kill their host, CBASS systems must be finely tuned to recognize an invading phage. Understanding this regulation is critical to

understanding the utility of CBASS systems and will likely reveal mechanisms phage exploit to neutralize them.

4.2.1.2: CapV

The activity of CapV and other CBASS effectors is controlled through allosteric interactions with specific cyclic oligo-nucleotides [18, 28, 46]. CBASS systems are frequently found within mobile genetic elements [43] passed through horizontal gene transfer events where they are likely to encounter hosts with preexisting nucleotide second messenger systems. Therefore, CBASS effector allosteric sites must be exquisitely insulated from interference derived from their host's endogenous biological processes. In the case of CapV, *V. cholerae* contains an extensive c-di-GMP signaling network which can range from low nM to low µM intracellular concentrations of c-di-GMP [104]. We have shown in vitro that the phospholipase activity of CapV is specifically regulated by cGAMP and not c-di-GMP or c-di-AMP (Chapter 2). In the case of *E. coli* ECOR31, two homologous CapV and DncV systems exist where each phospholipase specifically recognizes only the cdN synthesized by its partner *dncV*-like synthase (Figure 1.3) [28]. The extensive catalogue of cyclic-oligonucleotides synthesized from the common catalytic motif found in CD-NTases (Figure 1.4) [28] suggests the allosteric sites which recognize them may also share common features. Sequence alignment and structural studies of these effectors will help to uncover ligand binding motifs utilized by CBASS systems. Additionally, these studies may help to predict effectors not encoded within CBASS operons that are still able to respond to them.

4.2.1.3: VC0180 – VC0181

Connecting the catalytic activities of DncV and CapV to abortive replication represented the monumental breakthrough in our understanding of the VSP-1 [43]. The architecture of the El Tor CBASS system includes *vc0180* and *vc0181* and the intact homologous system from *E. coli* provided phage immunity to a heterologous host for 6 of the 10 phage studied [43]. VC0180 is a two domain peptide composed of a putative E1 (ubiquitin-activating enzyme) and E2 (ubiquitin-conjugating enzyme) while VC0181 is a JAB domain (deubiquitinase). Infections of *E. coli* that expressed mutations in conserved features of the E1/E2 and JAB domains while maintaining wild type *capV* and *dncV* reduced phage protection to a single phage [43]. These results suggest that this putative ubiquitination system is central in the function of *dncV* and *capV,* but its activity and role in this process have not been studied. Ubiquitination in bacteria is not recognized as a method of post translational modification and understanding its role in phage defense would greatly expand this field of study [105].

In the canonical eukaryotic ubiquitination process the E1 domain adenylates the C-terminal carboxylate of the small ~7 kDa protein ubiquitin (reviewed in [106]). Adenylated ubiquitin is then transferred to a conserved cysteine residue located on the E1. This transfer step releases AMP during the formation of a labile thioester between the C-terminus of ubiquitin and the E1 cysteine. Ubiquitin is then transferred to a conserved cysteine residue in an E2 enzyme. Finally, ubiquitin is transferred to specific lysine residues located on target proteins, creating an isopeptide bond. This process is often facilitated by a third enzyme called E3 that does not typically covalently interact with the ubiquitin, although *V. cholerae* does not encode any proteins that appear to have an E3 domain. Ubiquitination frequently marks proteins for degradation, but in other cases ubiquitination serves as a post-translational modification to

regulate the activity of the target protein. The isopeptidase activity of deubiquinating enzymes, like JAB domains, cleaves ubiquitin from ubiquitinated proteins, thus removing the post-translational modification.

Ubiquitin and ubiquitin-like proteins are often small ~ 7 kDa proteins, and canonically they contain a C-terminal GG motif. Analysis of the region within and surrounding the El Tor CBASS does not reveal any such small peptide. The dichotomous putative activities of VC0180 and VC0181 are reminiscent of the HORMA CBASS system where the activity of the Trp-13 prevents formation of the HORMA:CdnD complex in the absence of a unknown peptide, hypothesized to be brought by invading phage or produced in response to infection [46]. Therefore, we hypothesize the ubiquitin-like protein that is utilized by VC0180 and VC0181 is likely to be available under similar circumstances. Ser and Lys amino acid substitution of conserved Cys residues in E2 enzymes have been previously used to capture ubiquitin modification covalently linked to the ligase [107, 108]. We have constructed tagged variants of conserved E1/E2 Cys residues found in VC0180, which can be used in future experiments to capture the ubiquitin-like protein modifier. By expressing these variants in a variety of cellular conditions, including phage infection, it may be possible to identify the ubiquitin-like peptide and determine its origin. Additionally, characterization of this protein modifier will facilitate the identification of proteins targeted for modification by VC0180 to understand how this activity contributes to phage immunity.

Interestingly, our Correlogy analysis found that DncV and VC0180 have an MRS of 0.501 (Figure 3.2, Supplemental File), which is much greater than the 0.045 threshold suggested for a gene network determined by Price & Kim [64]. cGAS, the eukaryotic homolog of DncV, is regulated by ubiquitination at a number of specific Lys residues (reviewed [109]). A complicated

hierarchy of ubiquitination and SUMOylation, a related post-translation modification, is used to coordinate cGAS stability as well as regulate the synthesis of cGAMP. These lines of evidence suggest DncV is a likely target of VC0180 directed post-translation modification and structural comparisons of DncV and cGAS may reveal candidate residues involved in this process.

4.2.2: Characterization of the *dncV* - *dcdV* gene network

4.2.2.1: DCD Enzymatic activities

The in vitro activity of the DCD deaminase domain of DcdV was shown to deaminate both dCMP and dCTP substrates (Figure 3.11), making it the second DCD enzyme described with this catalytic repertoire [51]. Furthermore, addition of equimolar dTTP to dCMP or dCTP substrates in DcdV containing cell lysates did not appear to inhibit the enzyme's deaminase activity (Figure 3.12). While these data are reliably reproducible, the chemical complexity of cell lysates is likely to obscure much of the nuance to DcdV's activity. Millimolar concentrations of substrate are required in these reactions to produce a robust change in ABS_{630} with the Berthelot's reagent in order to monitor deamination through the evolution of NH_4. By using purified proteins and more sensitive techniques to measure nucleotide substrates and products, such as thin-later chromatography [51] or LC-MS/MS [110], detailed kinetic analysis and allosteric regulation can be determined in the absence of the lysate milieu. These experiments will help to clarify the precise function of the DCD domain in DcdV and aid in understanding its role in cell biology.

4.2.2.2: NK-Domain Activity

NK enzymes typically perform the reversible transfer of the γ-phosphate from an ATP donor to a deoxy/nucleotide monophosphate substrate with the help of a Mg^{2+} cofactor, although

many exceptions exist [86]. Despite showing no deaminase activity in cell lysates (Figure 3.11) the expression of the DCD domain variant DcdVE384A was still able to enhance the misincorporation of dU into the genome of *Δung E. coli* (Figure 3.15). This result suggests that the activity of the NK domain is likely to produce dUTP. *In vitro* experiments using purified DcdV and DcdV^{E384A} in the presence of deoxy/nucleotide substrates and *in vivo* analysis of intracellular nucleotide concentrations using LC/MS/MS will reveal whether the NK domain of DcdV is synthesizing dUTP.

4.2.2.3: DifV

While Correlogy predicted a shared gene network between *dncV* and *dcdV* there is currently no information about the conservation of *difV* or its association with either gene. As *difV* was previously unannotated, it was not included in the initial bioinformatic analysis and subsequent BLAST searches have revealed no homologs to sequences found within the orf1, where *difV* resides (Figure 3.6A). Further characterization of the precise location and coding sequence of *difV* will aid in the understanding of its conservation. Analysis of some *dcdV* homologs has revealed a handful of cases where a small ORF is annotated immediately 5' of the *dcdV* homolog but both nucleotide and amino acid alignments have failed to reveal any conservation between them. Future experiments involving *dcdV* homologous will help to determine if a 5' regulatory gene product is a conserved feature of these enzymes, and if so, what common features and activities do these regulators share.

4.2.2.4: DncV – DcdV – DifV

While much remains to be understood about the DcdV and DifV, a clear connection between these two gene products has been established. On the other hand, the connection

between DncV and DcdV predicted by Correlogy remains mysterious. From the data presented in Chapter 3, it is clear that DcdV activity is unlikely to be allosterically activated by cGAMP. Additionally, DcdV and DncV are not encoded in a shared operon (Figure 3.2). Both these features make DncV-DcdV unique from the CBASS CD-NTase – effector relationships described to date [43, 46], suggesting an alternative mechanism is responsible for the frequency with which they co-occur. DifV may be the key to bridging the divide between these two enzymes and further experiments will be performed to test the stability and expression of DifV in the context of phage infection and cGAMP abundance.

An alternative hypothesis for the predicted DcdV/DncV network could arise from the enzymatic activities that intersect at folate. The enzymatic activity of DncV is repressed by certain folate species, including 5-MTHF [23]. Two precursors to 5-MTHF are 5,10-MTHF and 10-MTHF which are used for the denovo biosynthesis of purines and dTMP, respectively [50]. When active, DcdV is likely to alter deoxynucleotide pools which could lead to the consumption of 5,10-MTHF and 10-MTHF, which in turn depletes available 5-MTHF. In this scenario, DcdV activity could serve as a mechanism to de-repress DncV and induce activation of CBASS. Future studies measuring intracellular concentrations of nucleotides, cGAMP, and folates, using LC/MS-MS, will reveal if DncV and DcdV are linked through nucleotide metabolism. This regulatory mechanism could be easily deployed for the activation of other CBASS CD-NTases as a wide variety of nucleotide metabolizing enzymes could achieve similar results as DcdV. These experiments will help shed light on whether or not the primary role of VSP-1 is to facilitate phage defense, and if not, what alternative benefits have been bestowed upon the El Tor biotype.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Ali M, Nelson AR, Lopez AL, Sack DA (2015) Updated Global Burden of Cholera in Endemic Countries. 1–13. doi: 10.1371/journal.pntd.0003832

2. Ryan ET (2011) The cholera pandemic, still with us after half a century: Time to rethink. PLoS Negl Trop Dis 5:1–2. doi: 10.1371/journal.pntd.0001003

3. Hu D, Liu B, Feng L, et al (2016) Origins of the current seventh cholera pandemic. Proc Natl Acad Sci U S A 113:E7730–E7739. doi: 10.1073/pnas.1608732113

4. Dziejman M, Balon E, Boyd D, et al (2002) Comparative genomic analysis of Vibrio cholerae: Genes that correlate with cholera endemic and pandemic disease. Proc Natl Acad Sci U S A 99:1556–1561. doi: 10.1073/pnas.042667999

5. O'Shea YA, Finnan S, Reen FJ, et al (2004) The Vibrio seventh pandemic island-II is a 26·9 kb genomic island present in Vibrio cholerae El Tor and O139 serogroup isolates that shows homology to a 43·4 kb genomic island in V. vulnificus. Microbiology 150:4053–4063. doi: 10.1099/mic.0.27172-0

6. Nguyen TH, Pham TD, Higa N, et al (2018) Analysis of Vibrio seventh pandemic island II and novel genomic islands in relation to attachment sequences among a wide variety of Vibrio cholerae strains. 150–157. doi: 10.1111/1348-0421.12570

7. Taviani E, Grim CJ, Choi J, et al (2010) comparative genomic analysis. 16961:130–137. doi: 10.1111/j.1574-6968.2010.02008.x

8. Nusrin S, Gil AI, Bhuiyan NA, et al (2009) Peruvian Vibrio cholerae O1 El Tor strains possess a distinct region in the Vibrio seventh pandemic island-II that differentiates them from the prototype seventh pandemic El Tor strains. J Med Microbiol 58:342–354. doi: 10.1099/jmm.0.005397-0

9. Davies BW, Bogard RW, Young TS, Mekalanos JJ (2012) Coordinated Regulation of Accessory Genetic Elements Produces Cyclic Di-Nucleotides for V. cholerae Virulence. Cell 149:358–370. doi: 10.1016/j.cell.2012.01.053

10. Sondermann H, Shikuma NJ, Yildiz FH (2012) You've come a long way: C-di-GMP signaling. Curr Opin Microbiol 15:140–146. doi: 10.1016/j.mib.2011.12.008

11. Corrigan RM, Gründling A (2013) Cyclic di-AMP: Another second messenger enters the fray. Nat Rev Microbiol 11:513–524. doi: 10.1038/nrmicro3069

12. Jenal U, Reinders A, Lori C (2017) Cyclic di-GMP: Second messenger extraordinaire. Nat Rev Microbiol 15:271–284. doi: 10.1038/nrmicro.2016.190

13. Römling U, Galperin MY, Gomelsky M (2013) Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol Mol Biol Rev 77:1–52. doi: 10.1128/MMBR.00043-12

14. Ross P, Weinhouse H, Aloni Y, et al (1987) Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. Nature 325:279–81.

15. Witte G, Hartung S, Büttner K, Hopfner KP (2008) Structural Biochemistry of a Bacterial Checkpoint Protein Reveals Diadenylate Cyclase Activity Regulated by DNA Recombination Intermediates. Mol Cell 30:167–178. doi: 10.1016/j.molcel.2008.02.020

16. Kuchta K, Knizewski L, Wyrwicz LS, et al (2009) Comprehensive classification of nucleotidyltransferase fold proteins: Identification of novel families and their representatives in human. Nucleic Acids Res 37:7701–7714. doi: 10.1093/nar/gkp854

17. Hallberg ZF, Wang XC, Wright TA, et al (2016) Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3′, 3′-cGAMP). Proc Natl Acad Sci 201515287. doi: 10.1073/pnas.1515287113

18. Severin GB, Ramliden MS, Hawver LA, et al (2018) Direct activation of a phospholipase by cyclic GMP-AMP in El Tor Vibrio cholerae. Proc Natl Acad Sci U S A 115:E6048–E6055. doi: 10.1073/pnas.1801233115

19. D'Souza M, Glass EM, Syed MH, et al (2007) Sentra: A database of signal transduction proteins for comparative genome analysis. Nucleic Acids Res 35:2006–2008. doi: 10.1093/nar/gkl949

20. Kazi MI, Conrado AR, Mey AR, et al (2016) ToxR Antagonizes H-NS Regulation of Horizontally Acquired Genes to Drive Host Colonization. PLoS Pathog 12:1–26. doi: 10.1371/journal.ppat.1005570

21. Van Kessel JC, Rutherford ST, Shao Y, et al (2013) Individual and combined roles of the master regulators apha and luxr in control of the Vibrio harveyi quorum-sensing regulon. J Bacteriol 195:436–443. doi: 10.1128/JB.01998-12

22. Li YH, Tian X (2012) Quorum sensing and bacterial social interactions in biofilms. Sensors 12:2519–2538. doi: 10.3390/s120302519

23. Zhu D, Wang L, Shang G, et al (2014) Structural Biochemistry of a Vibrio cholerae Dinucleotide Cyclase Reveals Cyclase Activity Regulation by Folates. Mol Cell 55:931–937. doi: 10.1016/j.molcel.2014.08.001

24. Nelson JW, Sudarsan N, Phillips GE, et al (2015) Control of bacterial exoelectrogenesis by c-AMP-GMP. Proc Natl Acad Sci 112:5389–94. doi: 10.1073/pnas.1419264112

25. Hallberg ZF, Chan CH, Wright TA, et al (2019) Structure and mechanism of a hypr GGDEF enzyme that activates cGAMP signaling to control extracellular metal respiration. Elife 8:1–36. doi: 10.7554/eLife.43959

26. Li F, Cimdins A, Rohde M, et al (2019) DncV Synthesizes Cyclic GMP-AMP and Regulates Biofilm Formation and Motility in Escherichia coli ECOR31. MBio 10:e02492-18. doi: 10.1128/mBio.02492-18

27. Burroughs AM, Zhang D, Schäffer DE, et al (2015) Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. Nucleic Acids Res 43:10633–10654. doi: 10.1093/nar/gkv1267

28. Whiteley AT, Eaglesham JB, de Oliveira Mann CC, et al (2019) Bacterial cGAS-like enzymes synthesize diverse nucleotide signals. Nature 567:194–199. doi: 10.1038/s41586-019-0953-5

29. Nelson JW, Breaker RR (2017) The lost language of the RNA World. Sci Signal 10:1–11. doi: 10.1126/scisignal.aam8812

30. Margolis SR, Wilson SC, Vance RE (2017) Evolutionary Origins of cGAS-STING Signaling. Trends Immunol 38:733–743. doi: 10.1016/j.it.2017.03.004

31. Chen Q, Sun L, Chen ZJ (2016) Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing. Nat Immunol 17:1142–1149. doi: 10.1038/ni.3558

32. Tan X, Sun L, Chen J, Chen ZJ (2018) Detection of Microbial Infections Through Innate Immune Sensing of Nucleic Acids. Annu Rev Microbiol 72:447–478. doi: 10.1146/annurev-micro-102215-095605

33. Sun L, Wu J, Du F, et al (2013) Cyclic GMP-AMP Synthase Is a Cytosolic DNA Sensor That Activates the Type I Interferon Pathway. Science (80-) 339:786–791. doi: 10.1126/science.1232458

34. Gao D, Wu J, Wu YT, et al (2013) Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. Science (80-) 341:903–906. doi: 10.1126/science.1240933

35. West AP, Khoury-Hanold W, Staron M, et al (2015) Mitochondrial DNA stress primes the antiviral innate immune response. Nature 520:553–557. doi: 10.1038/nature14156

36. Kranzusch PJ, Lee ASY, Wilson SC, et al (2014) Structure-guided reprogramming of human cgas dinucleotide linkage specificity. Cell 158:1011–1021. doi: 10.1016/j.cell.2014.07.028

37. Zhang X, Shi H, Wu J, et al (2013) Cyclic GMP-AMP containing mixed Phosphodiester linkages is an endogenous high-affinity ligand for STING. Mol Cell 51:226–235. doi: 10.1016/j.molcel.2013.05.022

38. Woodward JJ, Lavarone AT, Portnoy DA (2010) C-di-AMP secreted by intracellular Listeria monocytogenes activates a host type I interferon response. Science (80-) 328:1703–1705. doi: 10.1126/science.1189801

39. Burdette DL, Monroe KM, Sotelo-Troha K, et al (2011) STING is a direct innate immune sensor of cyclic di-GMP. Nature 478:515–518. doi: 10.1038/nature10429

40. McFarland AP, Luo S, Ahmed-Qadri F, et al (2017) Sensing of Bacterial Cyclic Dinucleotides by the Oxidoreductase RECON Promotes NF-κB Activation and Shapes a Proinflammatory Antibacterial State. Immunity 46:433–445. doi: 10.1016/j.immuni.2017.02.014

41. Gao P, Ascano M, Zillinger T, et al (2013) Structure-function analysis of STING activation by c[G(2′,5′) pA(3′,5′)p] and targeting by antiviral DMXAA. Cell 154:748–762. doi: 10.1016/j.cell.2013.07.023

42. McFarland AP, Burke TP, Carletti AA, et al (2018) RECON-dependent inflammation in hepatocytes enhances listeria monocytogenes cell-to-cell spread. MBio 9:1–15. doi: 10.1128/mBio.00526-18

43. Cohen D, Melamed S, Millman A, et al (2019) Cyclic GMP–AMP signalling protects bacteria against viral infection. Nature 574:691–695. doi: 10.1038/s41586-019-1605-5

44. Makarova KS, Wolf YI, Snir S, Koonin E V. (2011) Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems. J Bacteriol 193:6039–6056. doi: 10.1128/JB.05535-11

45. Schulman BA, Wade Harper J (2009) Ubiquitin-like protein activation by E1 enzymes: The apex for downstream signalling pathways. Nat Rev Mol Cell Biol 10:319–331. doi: 10.1038/nrm2673

46. Ye Q, Lau RK, Mathews IT, et al (2020) HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity. Mol Cell 77:709-722.e7. doi: 10.1016/j.molcel.2019.12.009

47. Seed KD (2015) Battling Phages: How Bacteria Defend against Viral Attack. PLoS Pathog 11:1–5. doi: 10.1371/journal.ppat.1004847

48. Hampton HG, Watson BNJ, Fineran PC (2020) The arms race between bacteria and their phage foes. Nature 577:327–336. doi: 10.1038/s41586-019-1894-8

49. Samson JE, Magadán AH, Sabri M, Moineau S (2013) Revenge of the phages: Defeating bacterial defences. Nat Rev Microbiol 11:675–687. doi: 10.1038/nrmicro3096

50. O'Donovan GA, Neuhard J (1970) Pyrimidine metabolism in microorganisms. Bacteriol Rev 34:278–343. doi: 10.1128/mmbr.34.3.278-343.1970

51. Zhang Y, Maley F, Maley GF, et al (2007) Chloroviruses Encode a Bifunctional dCMP-dCTP Deaminase That Produces Two Key Intermediates in dTTP Formation. J Virol 81:7662–7671. doi: 10.1128/jvi.00186-07

52. Stavrou S, Ross SR (2015) APOBEC3 Proteins in Viral Immunity. J Immunol 195:4565–4570. doi: 10.4049/jimmunol.1501504

53. Koning FA, Newman ENC, Kim E-Y, et al (2009) Defining APOBEC3 Expression Patterns in Human Tissues and Hematopoietic Cell Subsets. J Virol 83:9474–9485. doi: 10.1128/jvi.01089-09

54. Son MS, Taylor RK (2011) Genetic Screens and Biochemical Assays to Characterize Vibrio cholerae O1 Biotypes : Classical and El Tor. 1–17. doi: 10.1002/9780471729259.mc06a02s22

55. Beyhan S, Tischler AD, Camilli A, Yildiz FH (2006) Differences in gene expression between the classical and El Tor biotypes of Vibrio cholerae O1. Infect Immun 74:3633–3642. doi: 10.1128/IAI.01750-05

56. Taviani E, Grim CJ, Choi J, et al (2010) Discovery of novel Vibrio cholerae VSP-II genomic islands using comparative genomic analysis. FEMS Microbiol Lett 308:130–137. doi: 10.1111/j.1574-6968.2010.02008.x

57. Doron S, Melamed S, Ofir G, et al (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. Science (80-) 359:0–12. doi: 10.1126/science.aar4120

58. Johansson E, Fanø M, Bynck JH, et al (2005) Structures of dCTP deaminase from Escherichia coli with bound substrate and product: Reaction mechanism and determinants of mono- and bifunctionality for a family of enzymes. J Biol Chem 280:3051–3059. doi: 10.1074/jbc.M409534200

59. Vértessy BG, Tóth J (2009) Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. Acc Chem Res 42:97–106. doi: 10.1021/ar800114w

60. Graziani S, Xia Y, Gurnon JR, et al (2004) Functional analysis of FAD-dependent thymidylate synthase ThyX from Paramecium bursaria chlorella virus-1. J Biol Chem 279:54340–54347. doi: 10.1074/jbc.M409121200

61. Zhang Y, Moriyama H, Homma K, Van Etten JL (2005) Chlorella Virus-Encoded Deoxyuridine Triphosphatases Exhibit Different Temperature Optima. J Virol 79:9945–9953. doi: 10.1128/jvi.79.15.9945-9953.2005

62. Chiu CS, Ruettinger T, Flanegan JB, Greenburg GR (1977) Role of deoxycytidylate deaminase in deoxyribonucleotide synthesis in bacteriophage T4 DNA replication. J Biol Chem 252:8603–8608.

63. Miller ES, Kutter E, Mosig G, et al (2003) Bacteriophage T4 Genome. Microbiol Mol Biol Rev 67:86–156. doi: 10.1128/mmbr.67.1.86-156.2003

64. Kim PJ, Price ND (2011) Genetic co-occurrence network across sequenced microbes. PLoS Comput Biol. doi: 10.1371/journal.pcbi.1002340

65. Batushansky A, Toubiana D, Fait A (2016) Correlation-Based Network Generation, Visualization, and Analysis as a Powerful Tool in Biological Studies: A Case Study in Cancer Cell Metabolism. Biomed Res Int. doi: 10.1155/2016/8313272

66. Skorupski K, Taylor RK (1996) Positive selection vectors for allelic exchange. Gene 169:47–52. doi: 10.1016/0378-1119(95)00793-8

67. Edelheit O, Hanukoglu A, Hanukoglu I (2009) Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. BMC Biotechnol 9:1–8. doi: 10.1186/1472-6750-9-61

68. Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci U S A 82:488–492. doi: 10.1073/pnas.82.2.488

69. Helene Thelin K, Taylor RK (1996) Toxin-coregulated pilus, but not mannose-sensitive hemagglutinin, is required for colonization by Vibrio cholerae O1 El Tor biotype and O139 strains. Infect Immun 64:2853–2856. doi: 10.1128/iai.64.7.2853-2856.1996

70. Dunn AK, Millikan DS, Adin DM, et al (2006) and pES213-Derived Tools for Analyzing Symbiotic. Society 72:802–810. doi: 10.1128/AEM.72.1.802

71. Bose JL, Rosenberg CS, Stabb E V. (2008) Effects of luxCDABEG induction in Vibrio fischeri: Enhancement of symbiotic colonization and conditional attenuation of growth in culture. Arch Microbiol 190:169–183. doi: 10.1007/s00203-008-0387-1

72. Fürste JP, Pansegrau W, Frank R, et al (1986) Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. Gene 48:119–131. doi: 10.1016/0378-1119(86)90358-6

73. Vanbleu E, Marchal K, Vanderleyden J (2004) Genetic and physical map of the pLAFR1 vector. DNA Seq - J DNA Seq Mapp 15:225–227. doi: 10.1080/10425170410001723949

74. Fernandez NL, Nhu NT, Hsueh BY, et al (2020) VIBRIO CHOLERAE ADAPTS TO SESSILE AND MOTILE LIFESTYLES BY CYCLIC DI-GMP REGULATION OF CELL SHAPE. bioRxiv 2020.04.15.043257. doi: 10.1101/2020.04.15.043257

75. Dong H, Liu Y, Zu X, et al (2015) An enzymatic assay for high-throughput screening of cytidine-producing microbial strains. PLoS One 10:1–14. doi: 10.1371/journal.pone.0121612

76. Ngo TT, Phan APH, Yam CF, Lenhoff HM (1982) Interference in Determination of Ammonia with the Hypochlorite-Alkaline Phenol Method of Berthelot. Anal Chem 54:46–49. doi: 10.1021/ac00238a015

77. Schmidt TT, Sharma S, Reyes GX, et al (2020) Inactivation of folylpolyglutamate synthetase Met7 results in genome instability driven by an increased dUTP/dTTP ratio. Nucleic Acids Res 48:264–277. doi: 10.1093/nar/gkz1006

78. Mandlik A, Livny J, Robins WP, et al (2011) RNA-seq-based monitoring of infection-linked changes in vibrio cholerae gene expression. Cell Host Microbe 10:165–174. doi: 10.1016/j.chom.2011.07.007

79. Krin E, Pierlé SA, Sismeiro O, et al (2018) Expansion of the SOS regulon of Vibrio cholerae through extensive transcriptome analysis and experimental validation. BMC Genomics 19:1–18. doi: 10.1186/s12864-018-4716-8

80. Papenfort K, Förstner KU, Cong J-P, et al (2015) Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. Proc Natl Acad Sci 112:E766–E775. doi: 10.1073/pnas.1500203112

81. Raabe CA, Hoe CH, Randau G, et al (2011) The rocks and shallows of deep RNA sequencing: Examples in the Vibrio cholerae RNome. Rna 17:1357–1366. doi: 10.1261/rna.2682311

82. Liu JM, Livny J, Lawrence MS, et al (2009) Experimental discovery of sRNAs in Vibrio cholerae by direct cloning, 5S/ tRNA depletion and parallel sequencing. Nucleic Acids Res 37:1–10. doi: 10.1093/nar/gkp080

83. Updegrove TB, Zhang A, Storz G (2016) Hfq: The flexible RNA matchmaker. Curr Opin Microbiol 30:133–138. doi: 10.1016/j.mib.2016.02.003

84. Lenz DH, Mok KC, Lilley BN, et al (2004) The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae. Cell 118:69–82. doi: 10.1016/j.cell.2004.06.009

85. Kelley LA, Mezulis S, Yates CM, et al (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. doi: 10.1038/nprot.2015.053

86. Leipe DD, Koonin E V., Aravind L (2003) Evolution and classification of P-loop kinases and related proteins. J Mol Biol 333:781–815. doi: 10.1016/j.jmb.2003.08.040

87. Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1:945–951. doi: 10.1002/j.1460-2075.1982.tb01276.x

88. Reizer J, Buskirk S, Reizer A, et al (1994) A novel zinc-binding motif found in two ubiquitous deaminase families. Protein Sci 3:853–856. doi: 10.1002/pro.5560030515

89. Sharma S, Patnaik SK, Thomas Taggart R, et al (2015) APOBEC3A cytidine deaminase induces RNA editing in monocytes and macrophages. Nat Commun. doi: 10.1038/ncomms7881

90. Cohen R, Wolfenden R (1971) Cytidine deaminase from Escherichia coli. Purification, properties and inhibition by the potential transition state analog 3,4,5,6-tetrahydrouridine. J Biol Chem 246:7561–5.

91. Hou HF, Liang YH, Li LF, et al (2008) Crystal Structures of Streptococcus mutans 2′-Deoxycytidylate Deaminase and Its Complex with Substrate Analog and Allosteric Regulator dCTP·Mg2 +. J Mol Biol 377:220–231. doi: 10.1016/j.jmb.2007.12.064

92. Marx A, Alian A (2015) The First crystal structure of a dTTP-bound deoxycytidylate deaminase validates and details the allosteric-inhibitor binding site. J Biol Chem 290:682–690. doi: 10.1074/jbc.M114.617720

93. Krokan HE, Bjoras M (2013) Base Excision Repair. Cold Spring Harb Perspect Biol 5:a012583–a012583. doi: 10.1101/cshperspect.a012583

94. Tye BK, Chien J, Lehman IR, et al (1978) Uracil incorporation: A source of pulse-labelled DNA fragments in the replication of the Escherichia coli chromosome. Proc Natl Acad Sci U S A 75:233–237. doi: 10.1073/pnas.75.1.233

95. Blower TR, Short FL, Rao F, et al (2012) Identification and classification of bacterial Type III toxin-antitoxin systems encoded in chromosomal and plasmid genomes. Nucleic Acids Res 40:6158–6173. doi: 10.1093/nar/gks231

96. Kang SM, Kim DH, Jin C, Lee BJ (2018) A systematic overview of type II and III toxin-antitoxin systems with a focus on druggability. Toxins (Basel). doi: 10.3390/toxins10120515

97. Jamet A, Nassif X (2015) New Players in the Toxin Field : Polymorphic Toxin Systems in. MBio 6:1–8. doi: 10.1128/mBio.00285-15.Copyright

98. Iyer LM, Zhang D, Rogozin IB, Aravind L (2011) Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic Acids Res 39:9473–9497. doi: 10.1093/nar/gkr691

99. Ahmad SI, Kirk SH, Eisenstark A (1998) Thymine Metabolism and Thymineless Death in Prokaryotes and Eukaryotes. Annu Rev Microbiol 52:591–625. doi: 10.1146/annurev.micro.52.1.591

100. Guzmán EC, Martín CM (2015) Thymineless death, at the origin. Front Microbiol 6:1–7. doi: 10.3389/fmicb.2015.00499

101. Itsko M, Schaaper RM (2014) dGTP Starvation in Escherichia coli Provides New Insights into the Thymineless-Death Phenomenon. PLoS Genet. doi: 10.1371/journal.pgen.1004310

102. Schaaper RM, Mathews CK (2013) Mutational consequences of dNTP pool imbalances in E. coli. DNA Repair (Amst) 12:73–79. doi: 10.1016/j.dnarep.2012.10.011

103. Serrano-Heras G, Bravo A, Salas M (2008) Phage φ29 protein p56 prevents viral DNA replication impairment caused by uracil excision activity of uracil-DNA glycosylase. Proc Natl Acad Sci U S A 105:19044–19049. doi: 10.1073/pnas.0808797105

104. Koestler BJ, Waters CM (2013) Exploring environmental control of cyclic di-gmp signaling in vibrio cholerae by using the ex vivo lysate cyclic di-GMP assay(TELCA). Appl Environ Microbiol 79:5233–5241. doi: 10.1128/AEM.01596-13

105. Iyer LM, Burroughs AM, Aravind L (2006) The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like β-grasp domains. Genome Biol 7:1–23. doi: 10.1186/gb-2006-7-7-r60

106. Pickart CM, Eddins MJ (2004) Ubiquitin: Structures, functions, mechanisms. Biochim Biophys Acta - Mol Cell Res 1695:55–72. doi: 10.1016/j.bbamcr.2004.09.019

107. Streich FC, Lima CD (2018) Strategies to trap enzyme-substrate complexes that mimic michaelis intermediates during E3-mediated ubiquitin-like protein ligation. Methods Mol Biol 1844:169–196. doi: 10.1007/978-1-4939-8706-1_12

108. Bard JAM, Martin A (2018) The Ubiquitin Proteasome System. Methods Mol Biol. doi: 10.1007/978-1-4939-8706-1

109. Saeed AFUH, Ruan X, Guan H, et al (2020) Regulation of cGAS-Mediated Immune Responses and Immunotherapy. Adv Sci. doi: 10.1002/advs.201902599

110. Chen P, Liu Z, Liu S, et al (2009) A LC-MS/MS Method for the Analysis of Intracellular Nucleoside Triphosphate Levels. Pharm Res 26:1504–1515. doi: 10.1007/s11095-009-9863-9