Understanding the Role of VpsT in *Vibrio cholerae*

by

Thomas Guest

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

August 2021

Abstract

Cholera remains an important human pathogen despite the availability of vaccines and effective treatments. The current seventh pandemic is the longest recorded and outbreaks in Yemen have been the most deadly. The ability of *Vibrio cholerae* to shift between a human host and the aquatic environment has a vital role in its persistence. A complex transcriptional network reprograms the bacteria for life in cooler, nutrient poor environments.

In the aquatic environment *V. cholerae* uses a number of mechanisms to enhance its survival including formation of matrix-enclosed rugose phase variants, biofilm formation and colonisation of aquatic fauna. The transcriptional regulator VpsT, in response to the intracellular signalling molecule cyclic diguanylic acid (c-di-GMP), activates the expression of genes involved in the production of vibrio polysaccharide, promoting rugosity and the biofilm lifestyle. The regulator also represses expression of *rpoS*, the RNA polymerase sigma factor that co-ordinates the bacterial response to stress. However, the full repertoire of VpsT activity remains poorly defined.

To better understand the global regulatory role of VpsT we used chromatin immunoprecipitation, coupled with DNA sequencing, to map the distribution of VpsT across the genome. Our data identifies many additional targets for VpsT, substantially expanding the VpsT regulon. These targets include genes encoding proteins involved in motility, c-di-GMP metabolism and cell adhesion. Most notably, I show that VpsT is important for maintaining the correct level of c-di-GMP in the cell.

Acknowledgements

I wish to express my gratitude to everyone who has supported me in completing this work. In particular, I thank Prof David Grainger for his support, enthusiastic supervision, and for creating an excellent place to do science. Thanks to everyone in the Grainger Lab, past and present, who have been great colleagues and pub-quiz team-mates. I would also like to thank Prof Fitnat Yildiz at the University of California, Santa Cruz for gifting the c-di-GMP reporter plasmid.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

APS	ammonium persulphate
bp	base pair
BSA	bovine serum albumin
cAMP	cyclic AMP
CRP	cAMP receptor protein
c-di-GMP	cyclic-di-guanosine monophosphate
ChIP	chromatin immunoprecipitation
ChIP-seq	ChIP and genome sequencing
CTD	C-terminal domain
CVEC	conditionally viable environmental cells
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	2’-deoxyribonucleoside 5’-triphosphate
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
Frh	flagellum-regulated haemagluttinin
HEPES	4-(2-hydroxyethyl)-1- piperazineethanesulfphonic acid
H-NS	Histone-like nucleoid structuring protein
KanR	kanamycin resistance
kbp	kilobase pair
LB	Luria-Bertani media
mRNA	messenger RNA

MuGENT	multiplex genome editing by natural transformation
nt	nucleotide
NTD	N-terminal domain
NTP	nucleoside triphosphate
C	celsius
Oligo	oligonucleotide
ONPG	Ortho-nitrophenyl-β-galactosidase
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PBS	phosphate buffered saline
QS	Quorum sensing
RFI	relative fluorescence intensity
RNA	ribonucleic acid
RNA-seq	RNA sequencing
RNAP	RNA polymerase
RNAse	ribonuclease
SDS	sodium dodecyl sulphate
SNP	single nucleotide polymorphism
Sm^R	steptomycin resistance
SOE	splicing by overlap extension
$Spec^R$	spectinomycin resistance
SST	sea surface temperature
T4 PNK	T4 polynucleotide kinase
TCP	toxin co-regulated pilus
TEMED	tetramethylethylenediamine
Tet^R	tetracycline resistance
TF	transcription factor
Tris	tris (hydroxymethyl) aminoethane
TSS	transcription start site

V	Volts
VBNC	viable but non-culturable
V. cholerae	*Vibrio cholerae*
VPS	vibrio polysaccharide
vpv	vibrio phase variation
VSP	vibrio seventh pandemic island
W	Watts
wt	wild type

1 | Introduction

1.1 The central dogma

The central dogma describes a linear flow of information that begins with genes encoded in DNA and ends with proteins via the intermediary RNA (Figure 1.1). Watson-Crick base pairing allows the transfer of information from DNA to RNA. The DNA template strand is transcribed by RNA polymerase to produce a complementary messenger RNA (mRNA) transcript. The mRNA transcript is translated by ribosomes. Ribosomes read the mRNA code in triplets called codons that dictate the assembly of amino acid chains, which then fold to form protein. According to the dogma, proteins provide the structural, catalytic and regulatory activity necessary for life. However, RNA too can have catalytic and regulatory activity (Cech and Steitz, 2014).

The genome contains all the genes required for an organism to develop, grow and thrive. Bacteria must be able to respond to change in their environment. Regulation of gene transcription ensures that genes are transcribed when the proteins they encode are required. Therefore, in order to make efficient use of the information encoded in DNA various mechanisms selectively activate and repress transcription (Browning and Busby, 2004).

1.2 The operon model

The operon model proposed by Jacob and Monod was based on their research of the *lac* operon in *Escherichia coli* and lysogeny of the λ bacteriophage (Yaniv, 2011). The model provides a framework for understanding the regulation of gene transcription. Operons are groups of genes regulated by a set of shared regulatory elements including a

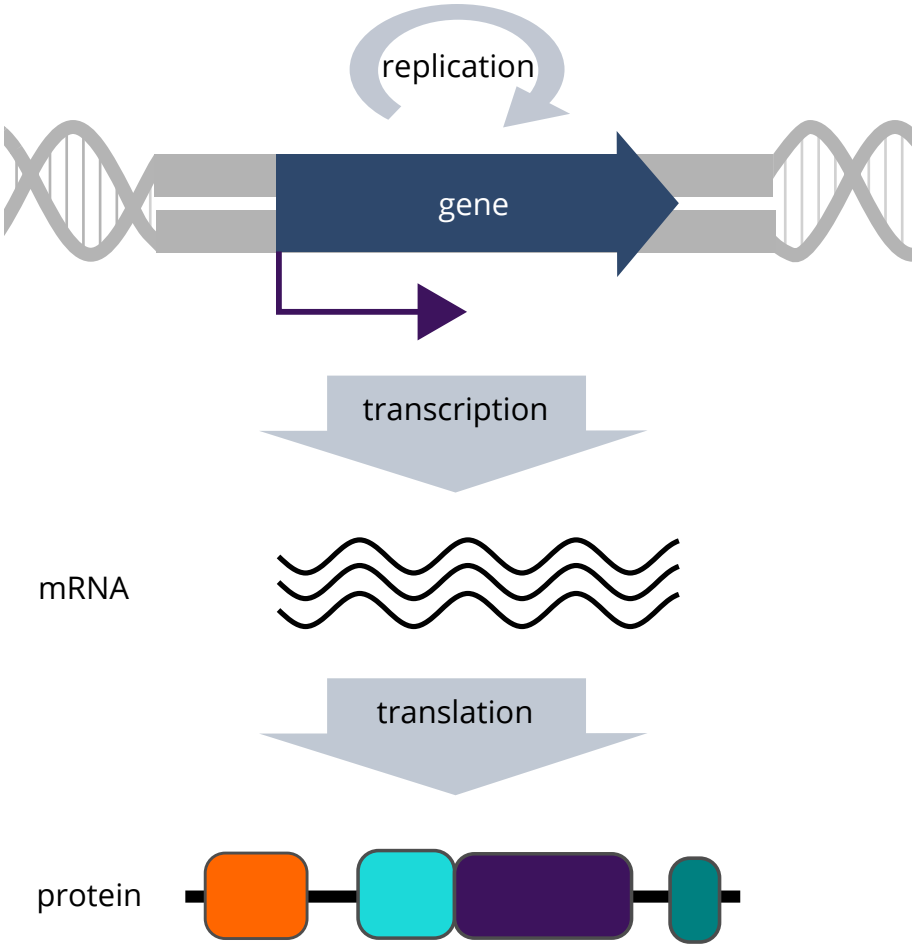

Figure 1.1: The central dogma. The double stranded DNA encoding genes is replicated by DNA polymerases, facilitated by Watson-Crick base pairing. DNA is transcribed by RNA polymerase into the single stranded complementary messenger RNA (mRNA). This is then translated by ribosomes into polypeptides, assembled from amino acids to form protein.

promoter, operator and a terminator (Browning and Busby, 2004). The promoter DNA is recognised and bound by RNA polymerase but this interaction can be influenced by proteins that bind to the operator. Theoretically, the operon model ensures that genes encoding proteins required for a particular biological function, such as lactose metabolism, are transcribed simultaneously and in the same regulatory environment.

1.3 RNA polymerase

The DNA-dependent RNA polymerase (RNAP) catalyses the transcription of genes in bacteria (Borukhov and Nudler, 2008). The RNAP core enzyme has five subunits ($\alpha^2\beta\beta'\omega$, Figure 1.2). The β and β′ form the pincers of the crab claw-shaped catalytic cleft. The catalytic site is embedded within the back of the cleft and is where the phosphodiester bonds of the elongating RNA chain are formed. The alpha subunits are composed of two independently folded domains connected by a linker. The N-terminal domain is required for enzyme assembly. The flexibility of the linker allows the C-terminal domain (□CTD) to make contact with DNA upstream of the RNAP binding site and/or transcription factors (Lee et al., 2012). The αCTD interacts with the DNA at the UP promoter element. The RNAP holoenzyme includes a sixth subunit, a sigma (σ) factor (Paget and Helmann, 2003). The σ-factor confers DNA sequence specificity by making direct contacts with promoter DNA and is required for formation of the closed complex.

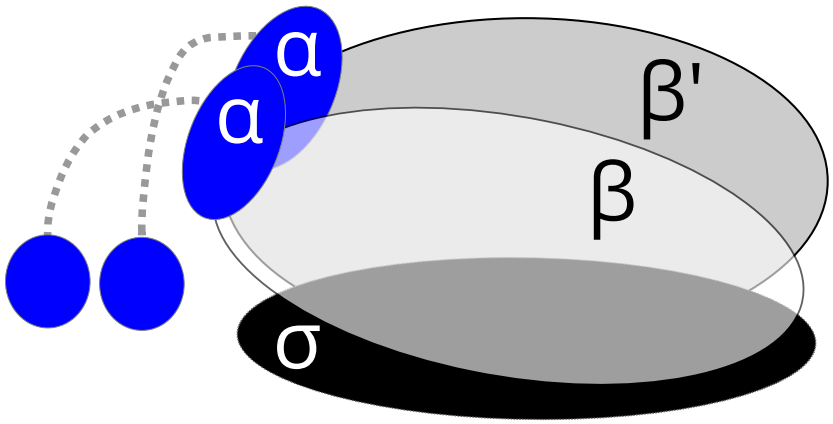

Figure 1.2: RNA polymerase subunits. A simplified diagram showing the arrangement of the major subunits of the bacterial DNA-dependent RNA polymerase holoenzyme. The β and β' subunits form a crab-claw shape, with the active site embedded within. The α subunits are composed of two domains (blue) joined by a flexible linker. The σ subunit makes specific interactions with promoter DNA, and is spread across the face of the RNAP crab claw.

Sigma factors

Many bacteria have more than one σ-factor, each with a set of target promoters that can be recognised to generate large-scale changes in gene transcription (Dennis and Bremer, 2008). The two families of σ-factors are distinguished by their homology either to σ^{70} or σ^{54} (Paget, 2015). The primary sigma factor, σ^{70} (RpoD), is responsible for directing the transcription of 'housekeeping' genes during logarithmic growth. The σ^{54} (RpoN) family factors target promoters for specific biological processes in response to environmental signals, require the hydrolysis of ATP and the activity of enhancer proteins to function (Wigneshweraraj et al., 2008).

Sigma 70 family factors are grouped by the configuration of up to four domains (Figure 1.3). The primary σ^{70} factors make up Group 1. The alternative σ-factors in Group 2 are non-essential and are structurally similar to Group 1 except for the lack of σ-1.1 (Paget, 2015). Included in Group 2 is RpoS (σ38), responsible for the general stress response and active during stationary phase. Group 3 includes many σ-factors that have roles tied to specific functions, including the regulation of motility (σ^{28}), cell envelope stress (σ^{24}) and heat shock (σ^{32}). In Group 4 are the extracytoplasmic function (ECF) factors. These are diverse, often control far smaller regulons and typically respond to extracellular signals (Mascher, 2013).

The exchange of σ-factors is mediated by mechanisms which, until required, constrain their ability to compete for the limited pool of available RNAP core enzyme (Paget, 2015). Alternative σ factors typically have lower affinity for RNAP core than σ^{70} so their concentration must usually be high for them to have an effect on transcription. While regulation at the level of transcription does occur, regulation of

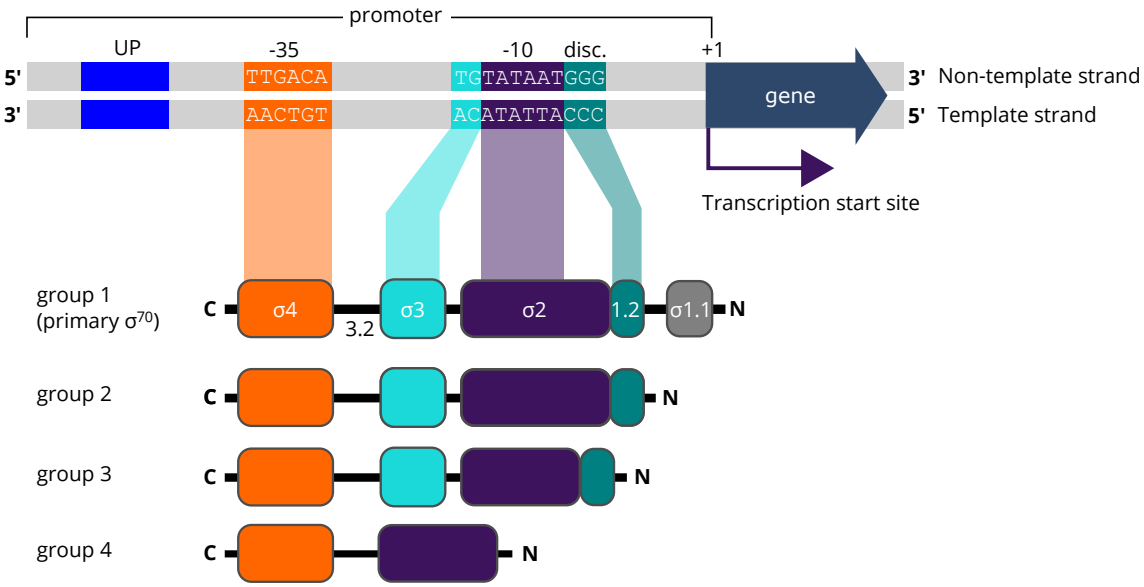

Figure 1.3: Promoter recognition by RNAPσ70 holoenzyme. Schematic shows the contacts between promoter DNA and the RNAPσ70 holoenzyme. Promoter elements are numbered from the transcription start site. The AT-rich UP element is bound by the α-CTD. The -35 element interacts with σ domain 4. The extended -10 interacts with σ domain 3. The -10 element with σ domains 2.3 and 2.4. The σ-1.2 domain interacts with the discriminator region. The β RNAP subunit also makes contact at the position $+2$ downstream of the TSS.

translation and protein turnover are often more important. For example, the *E. coli rpoS* mRNA has a long 5' untranslated region containing a hairpin structure that blocks its translation (Battesti et al., 2011). This is relieved by three regulatory RNAs (DsrA, RprA, and ArcZ) that open the hairpin and free the ribosome binding site.

Many σ-factors are sequestered or cleaved by a cognate anti σ-factor, prohibiting its binding to the RNAP core (Treviño-Quintanilla et al., 2013). Anti σ-factors are often part of their cognate σ-factors regulon and this provides a negative feedback loop. Anti-σ factors can release the σ-factor in response to environmental cues, either by directly sensing, partner-switching or proteolysis (Mascher, 2013). For example, under normal conditions σ^{24} is sequestered at the inner membrane by the anti σ-factor complex, RseA and RseB (Grigorova, 2004). However, when the membrane becomes stressed the accumulation of mis-folded outermembrane proteins triggers a proteolytic cascade (via DegS) that results in degradation of RseA and frees σ^{24}.

Promoter recognition by σ^{70} family σ-factors

The 'housekeeping' σ^{70} factor has four domains that recognise and interact with different promoter elements arranged upstream of transcription start sites (TSS) (Mascher, 2013). Other RNAP subunits, such as β and the αCTD (at the UP element) make additional contacts with the DNA. Typical σ^{70} promoters have characteristic sequence motifs: the -35 element (TTGACA), -10 (TATAAT) and the discriminator (GGG). The strength of promoters, in the absence of additional regulators, depends on the degree of conservation to the optimal recognition sequences (Browning and Busby, 2016). Some promoters have an extended -10 (TGNTATAAT) element, which can help to overcome poor sequence

conservation of the -35 promoter element (Mitchell, 2003).

Domain σ1.1 prevents the σ2,3 and 4 DNA binding domains from making contact with DNA when not in complex with RNAP core. Domain σ1.2 interacts with the discriminator region, which is located downstream of the -10 element (Figure 1.3). Domain σ2 is the most widely conserved and makes contacts with the -10 promoter element on the non-template strand. This stabilises formation of the open complex and promotes melting of the DNA duplex (Figure 1.4). The -11A and -7T bases are the most highly conserved, these are flipped out of the base stack and interact with domain σ2.3 which is essential for DNA melting. Domain σ3 interacts with the duplex DNA major groove at the extended -10 element which can help overcome instability at promoters with poorly conserved -35 elements (Mitchell, 2003). A linker domain, σ3.2, attaches to σ4 through the active site, blocking the RNA exit channel which must later be displaced. The domain σ4 interacts with the -35 promoter element via a helix-turn-helix motif (σ4.1-4.2), this induces the DNA to bend and promotes additional contacts with upstream DNA. Domain σ4 can also make contact with transcriptional activators, such as CRP, that bind to the DNA further upstream.

Non-canonical promoters frequently occur within AT-rich horizontally-acquired DNA due to their resemblance to the -10 promoter element, but lack other promoter recognition sequences (Wade and Grainger, 2014). Despite this lack of sequence conservation these promoters can drive the production of spurious transcripts, which can be toxic, and are typically silenced by the histone-like structuring (H-NS) protein (Lamberte et al., 2017). An additional interaction between σ^{70} R451 and an AT-rich tract upstream of the -10 element is able to compensate for mismatches in the -10 and -35 sequences (Warman et al., 2020).

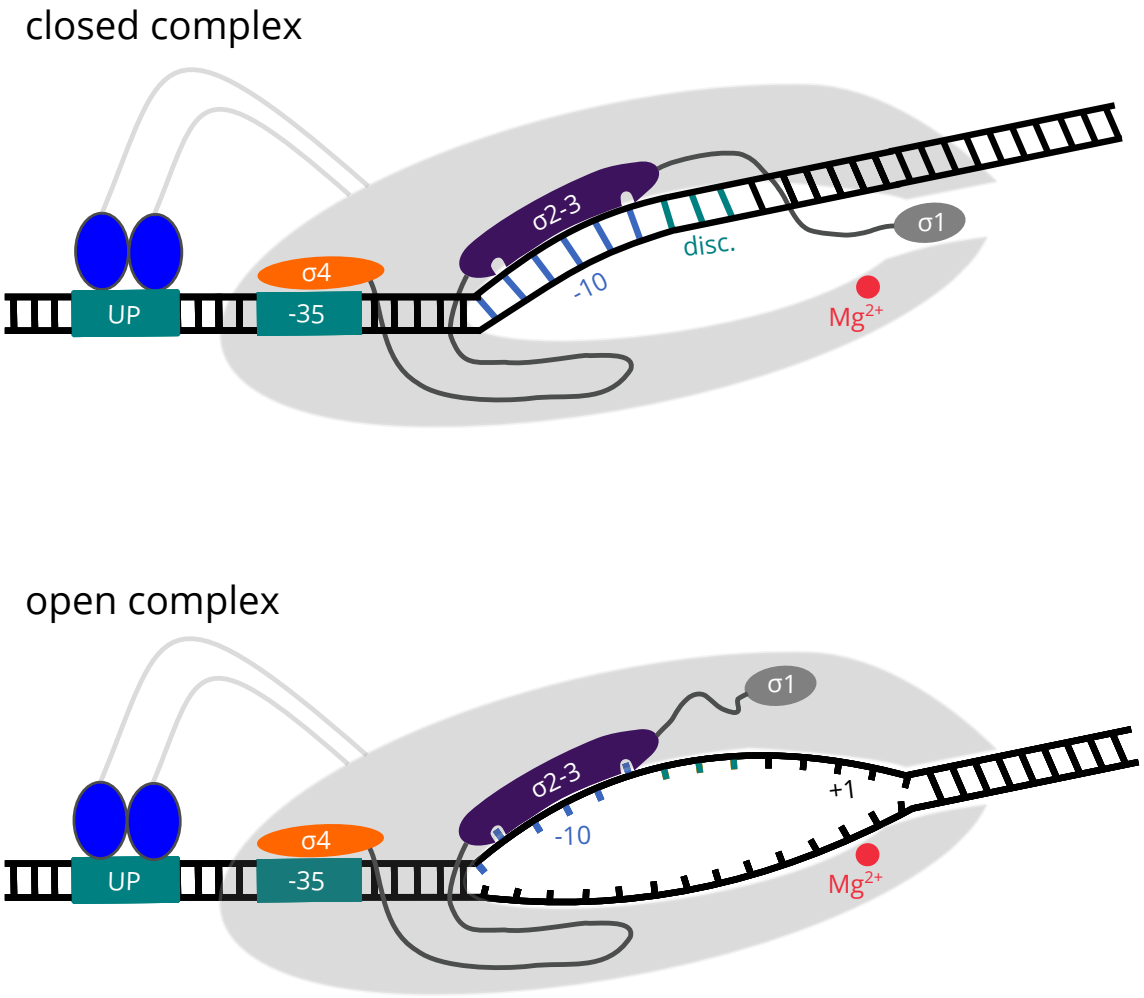

Figure 1.4: RNAP closed and open complex formation. Cartoon showing RNAP holoenzyme at a promoter. Initially, contact is made with duplex DNA. In the closed complex, the σ 3.2 protrudes into the the active site and σ 1.1 blocks the active site channel. During transition to the open complex, the DNA is melted and the template strand is positioned at the active site.

Transcription initiation

Once a promoter has been identified by the RNAP holoenzyme, RNA synthesis can begin. In the open complex the transcription bubble is melted, and DNA downstream of the +1 is clamped in place by the β and β' subunits (Figure 1.4). The active site is supplied with NTPs via a secondary channel. Initially RNA synthesis occurs while the RNAP is still bound to the promoter and the RNA exit channel is blocked by the σ-factor (Hsu et al., 2003). As the RNA grows, the complex becomes 'scrunched' and this causes a pause (Duchi, 2016). The scrunched open complex can either release the transcript (abortive initiation) and return to the open complex, or escape from the promoter. Abortive transcripts range in size from 2 to 15 nt long, and most promoters will undergo several rounds of re-initiation before productive elongation begins (Borukhov and Nudler, 2008). Unusual spacing between the -10 and -35 element, such as at rRNA promoters, promotes scrunching and helps to select the correct TSS (Winkelman and Gourse, 2017). Scrunching increases the productivity of promoters that are highly active, despite non-optimal promoter sequences. Dissociation of the σ-factor results in a conformational change that forms the elongation complex (EC). The EC can then progress along the DNA template, maintaining a constant size (~12 nt), assembling the RNA transcript until transcription is terminated (Borukhov and Nudler, 2008).

Termination

Termination is important for maintaining regulated transcription and for recycling RNAP (Roberts, 2019). In bacteria there are two principal modes; intrinsic or factor-independent and factor-dependent termination, both rely on the ability to stall the EC. The EC is highly

stable, mostly due to RNAP contacts with the hybrid formed of the elongating mRNA and DNA template (Komissarova et al., 2002).

Intrinsic terminators rely on the innate ability for RNA chains to form secondary structures that can disrupt the stability of the EC (Peters et al., 2011). Although the sequences of intrinsic terminators vary, they typically have two important features: a GC-rich repeat (that when transcribed form a hairpin) and an a T-tract (that forms a U-rich tail in the RNA:DNA hybrid). The formation of the hairpin results in a pause that halts the addition of the next nucleotide. As the hairpin extends the RNA:DNA hybrid is disrupted by either; forward translocation (where movement along the DNA continues without elongation of the RNA), slippage (that tugs the RNA transcript away) or allosteric changes (that alter the conformation of the EC) (Roberts, 2019). The effect is to favour disassociation from the RNAP. Intrinsic termination can be modulated by other factors such as NusA which enhances termination, and phage anti-terminators such as the lambda G and N proteins (Roberts et al., 1998, Schmidt and Chamberlin, 1987).

Factor-dependent termination is typically via Rho. Rho is an ATP-dependent RNA translocase that is widely conserved among bacterial species (Roberts, 2019). Rho terminates transcription of spurious (typically antisense transcripts), prophage DNA and following the occurrence of nonsense codons (Peters et al., 2011, 2012). Rho initially binds to DNA at C-rich primary sites, followed by secondary sites on another face of the protein that stimulates its ATPase activity (Richardson, 1982). Rho then moves along the DNA, passing it through its ring-like structure and blocks elongation when it meets the RNAP. Together with NusG, Rho helps to couple transcription with translation (Saxena et al., 2018). When the mRNA is being translated NusG is bound to the ribosome. If translation stops, NusG is released and instead binds

Rho, which stimulates the termination of transcription.

It is also possible for RNAP to remain bound following termination where it can 'flip' and synthesise antisense transcripts (Harden et al., 2020).

1.4 Regulation of transcription initiation

Apart from the regulation of transcription inferred by the exchange of σ-factors, the initiation of transcription can be regulated by other means. Proteins such as nucleoid associated proteins (NAPs) or transcription factors (activators and repressors) and small effector molecules can directly or indirectly regulate transcription initiation.

(p)ppGpp and the stringent response

To counter stresses such as nutrient starvation or heat shock, the stringent response diverts transcription to preserve resources (Hauryliuk et al., 2015). This is achieved via synthesis of the alarmones, pppGpp and ppGpp (collectively p(p)Gpp), by the RelA-SpoT (RSH) homologue family of proteins from GDP or GTP. RelA specifically detects amino acid depletion by monitoring tRNAs at ribosomes, accumulation of uncharged tRNAs trigger the synthesis of (p)ppGpp (Haseltine and Block, 1973). SpoT responds to multiple cues, including fatty acid, iron and carbon starvation (Hauryliuk et al., 2015). (p)ppGpp alters the activity of RNAP by binding to a site on the β' subunit that destabilises the open complex at the promoters of rRNA and ribosome proteins (Zuo et al., 2013). Simultaneously, (p)ppGpp activates transcription of amino acid biosynthesis genes via the transcription factor DksA (Paul et al., 2005).

Nucleoid associated proteins

Nucleoid associated proteins (NAPs) compact and give structure to chromosomes (Dillon and Dorman, 2010). However, they have a dual role as regulators of transcription because they alter the accessibility of DNA for RNAP. Since NAPs typically bind DNA with relatively little sequence specificity, they are thought to be an ancient form of transcriptional regulation (Visweswariah and Busby, 2015).

The histone-like structuring (H-NS) protein binds AT-rich DNA, where it forms bridges and filaments, inducing the formation of complex loops and bends in the DNA (Dame et al., 2020). The structural changes induced by H-NS can impede RNAP access and progression along the DNA. The N-terminal domain of H-NS is required for oligomerisation and the C-terminal domain is required for DNA binding (Grainger, 2016). Xenogenic silencing by H-NS minimises deleterious transcription of foreign AT-rich DNA, and also prevents the synthesis of spurious transcripts from intragenic promoter sequences (Lamberte et al., 2017).

There are other silencing proteins that are functionally analogous to H-NS but are structurally different. An example is TsrA, a silencing protein in *Vibrio cholerae* that binds to many of the same promoter sites as H-NS, including in AT-rich genomic islands that are important for pathogenesis (Caro et al., 2020). Another NAP that binds preferentially to intergenic DNA is the factor for inversion stimulation (Fis). Fis forms DNA-protein bridges and is enriched during early exponential growth, contributing to the formation of looped transcription boundaries (Dillon and Dorman, 2010, Dame et al., 2020).

In contrast to the NAPs that silence transcription, the integration host factor (IHF) targets σ^{54} promoter sequences and promotes transcription activation (Dillon and Dorman, 2010). IHF introduces bends in the DNA

that help to co-locate an enhancer binding protein (which typically bind further upstream) and the RNAP holoenzyme (Wigneshweraraj et al., 2008).

Transcription factors

Transcription factors interact at specific DNA sequences, typically located near to target promoters, and activate or repress transcription in response to environmental cues (Browning et al., 2019). Transcription factors typically have a DNA binding domain and a regulatory domain that controls its function, in response to a ligand, modification or interaction with another protein. Most transcription factors form dimers that increases their sequence specificity. The presence of a transcription factor binding site is not always indicative of a regulatory interaction (Wade et al., 2007). For example, many sites occur outside of regulatory regions, in intragenic DNA, or might instead influence chromosome structure. In the next section, repressors and activators are considered separately, however, many transcription factors act as both at different binding sites. The position of the transcription factor binding site is often a determinant of its regulatory activity.

Repressors

Repressors are transcription factors that repress transcription of their target genes. Most repressors work by steric hindrance, preventing the RNAP from binding to the promoter (Browning and Busby, 2004). Repressor binding sites often overlap the -35 and/or -10 promoter elements to block RNAP binding. Some repressors bind downstream of the promoter and trap the RNAP at the promoter. Alternatively, distal operator sites can induce DNA looping that impedes RNAP access.

Repression can be enhanced by repressors binding to secondary sites around a promoter, such as the LacI repressor which can form a tetramer, that induces DNA looping (Oehler et al., 1990). The combination of two repressors allows multiple stimuli to co-regulate transcription. This is particularly important at promoters regulating genes with lethal consequences, such as colicin K in *E. coli.* Transcription is 'double-locked' by two repressors, LexA and IscR (Butala et al., 2012). LexA is inactivated during the SOS DNA damage response but IscR delays LexA inactivation, ensuring transcription of colicin K is occurs only as a last resort.

Activators

Activator proteins are often split into different mechanistic classes (Figure 1.5, described in the legend)). Activators make contacts with the RNAP via surface residues called activating regions (AR). The cAMP receptor protein (CRP) is a well-studied model transcription activator that can activate via both the Class I and Class II mechanisms (Lee et al., 2012). However, promoters are often regulated by multiple activators utilising a combination of the two mechanisms.

Indirect activation can occur by activators that disrupt a repressor. For example, in *E. coli* transcription from the *nir* promoter is coordinated in response to both oxygen and nitrate. In the presence of nitrate the NarQ/NarX sensors activate NarL (Maris et al., 2005). NarL acts as an anti-repressor by disrupting IHF and Fis binding. This allows a second activator (triggered by oxygen deprivation) called Fnr to activate by the Class II mechanism. NarL is also able to directly activate transcription at other promoters, such as at the *yeaR-yoaG* and *ogt* operons, where it makes direct contacts with RNAP (Lin et al., 2007, Ruanto et al., 2020).

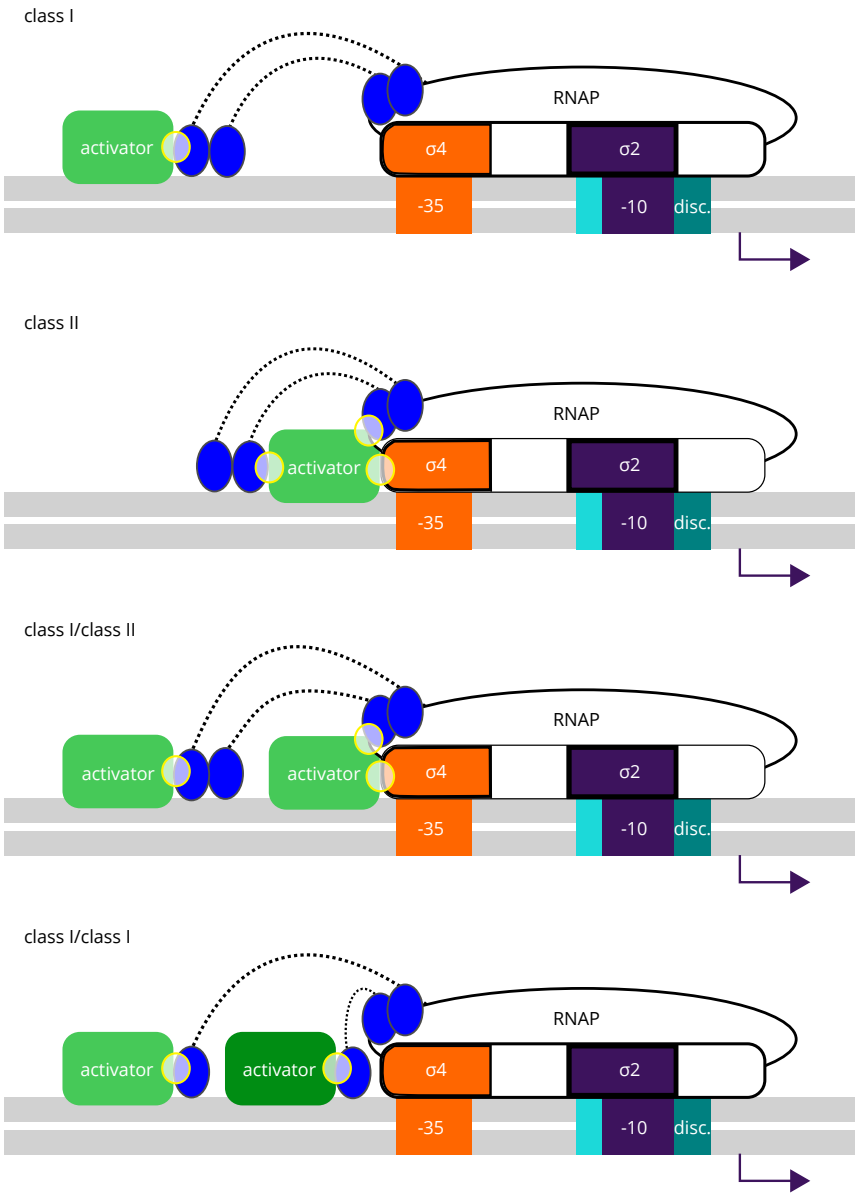

Figure 1.5: Transcription activators. Class I activators bind upstream of the promoter and interact with RNAP via the α CTD. For example, at the *E. coli lac* promoter CRP contacts the αCTD via activating region 1 (AR1). This helps to correctly align the σ-factor domain 2 with the -10 promoter element, promoting the separation of the duplex DNA strands. Class II activation requires alternative activating regions (AR2 and AR3). AR2 interacts with the α NTD and AR3 with σ domain 4. The αCTD binds upstream of CRP. At class II promoters the activator binding site is adjacent or overlaps the -35 promoter element and contact is made with the RNAP via the σ-factor.

1.5 Regulatory pathways

Two-component systems such as the NarQX, NarL system described earlier, are common. In these systems a sensor kinase phosphorylates the response regulator (the transcription factor) in response to a specific environmental cue (Gao et al., 2019). Other signalling cascades use similar mechanisms, incorporating the detection of a stimulatory cue and modification of an effector.

Quorum sensing

Quorum sensing (QS) is a signalling system that allows bacteria to respond to environmental changes in a collective manner (Papenfort and Bassler, 2016). Autoinducers (AI) are small diffusible substances produced by cells in proportion to their cell density, which are then interpreted by cellular signalling cascades to alter gene transcription, translation and other cellular processes.

Cyclic-di-GMP

The second messenger molecule 3’,5’-cyclic diguanylic acid (c-di-GMP) is a nucleotide second messenger that contributes to the regulation of motility, biofilm formation and virulence in many species of bacteria (Hengge, 2009). The intracellular concentration of c-di-GMP reflects the relative activity of diguanylate cyclases (DGC) and phosphodiesterases (PDE) (Conner et al., 2017). DGC proteins have characteristic GGDEF catalytic domains and synthesise c-di-GMP from two molecules of GTP. The PDE proteins have either EAL or HD-GYP catalytic domains and degrade c-di-GMP to 5’-pGpG or 2 GMP. The activity of these proteins is typically regulated via sensory domains that detect environmental cues

such as oxygen availability, redox, light and the presence of extracellular substances (Jenal et al., 2017). The regulatory effects of c-di-GMP can be mediated at multiple levels; transcription, post-transcription and by direct modification of proteins (Hengge, 2009).

1.6 Post-transcriptional regulation by non-coding RNA

Non-coding small RNAs fine-tune gene expression by modulating the stability and translation of mRNA (Cech and Steitz, 2014). However, they can also directly modulate the activity of proteins. An example of this is the *E. coli* 6S RNA that mimicks a σ70 promoter to sequester σ^{70}-RNAP holoenzyme during stationary phase (Trotochaud and Wassarman, 2005). These RNA regulators can be encoded in cis as part of the 5' untranslated region (UTR) of target mRNAs or in trans. Some sRNAs (such as RNAIII) are also coding, or may be processed from larger mRNA species (Waters and Storz, 2009). RNAs encoded in trans typically require the chaperone protein Hfq in order to effectively bind and regulate their target mRNA. Deletion of *hfq* abolishes the activity of many sRNAs, particularly in *E. coli,* while in other species such as in *B. subtilis,* Hfq is dispensable (Dambach et al., 2013).

Riboswitches are domains located on mRNAs that are able to modulate translation in response to the detection of a cognate ligand (Waters and Storz, 2009). Riboswitch domains alter conformation when bound to their ligand that results in the formation of hairpin structures, similar to those produced by terminator sequences, that can inhibit transcription or translation. Alternatively, riboswitches can activate expression of their target mRNA (Sudarsan et al., 2008).

1.7 Cholera

Cholera is an acute diarrhoeal disease caused by *Vibrio cholerae*; a halophilic, comma-shaped aquatic bacterium (Clemens et al., 2017). *V. cholerae* colonises the proximal small intestine. The cholera toxin (CT) is responsible for the characteristic profuse watery diarrhoea. Cholera is preventable; modern sanitation, improved hygiene and access to potable water has eliminated its risk to public health in developed countries. There are several oral vaccines and rehydration therapy is an effective treatment, that can be supplemented with antibiotics (Wierzba, 2019). However, despite the potential of these interventions the global burden of cholera is increasing (WHO, 2018). In 2017, according to the most recent report by the World Health Organisation (WHO) there were 1.2 million recorded cases and up to 6,000 deaths (Figure 1.6). Excluding Yemen (which accounted for 84 % of all cases) this amounted to a 45 % case increase compared with 2016 and a 33 % increase in fatalities (WHO, 2018). However, the true burden of disease is unknown because many endemic countries do not report seasonal disease.

1.8 Epidemiology of cholera

Cholera is endemic to 50 countries, concentrated around the Indian subcontinent and sub-Saharan Africa (Alam et al., 2006, Charles and Ryan, 2011). The *Vibrios* are a globally distributed genus of aquatic bacteria and like its relatives *V. cholerae* naturally inhabits tropical and temperate marine, brackish and freshwater ecosystems (Lutz et al., 2013). One-third of the planet and 1.3 billion people are at-risk of cholera (Wierzba, 2019). Modelling suggests the disease burden could increase in future due to rising sea surface temperature (SST) and

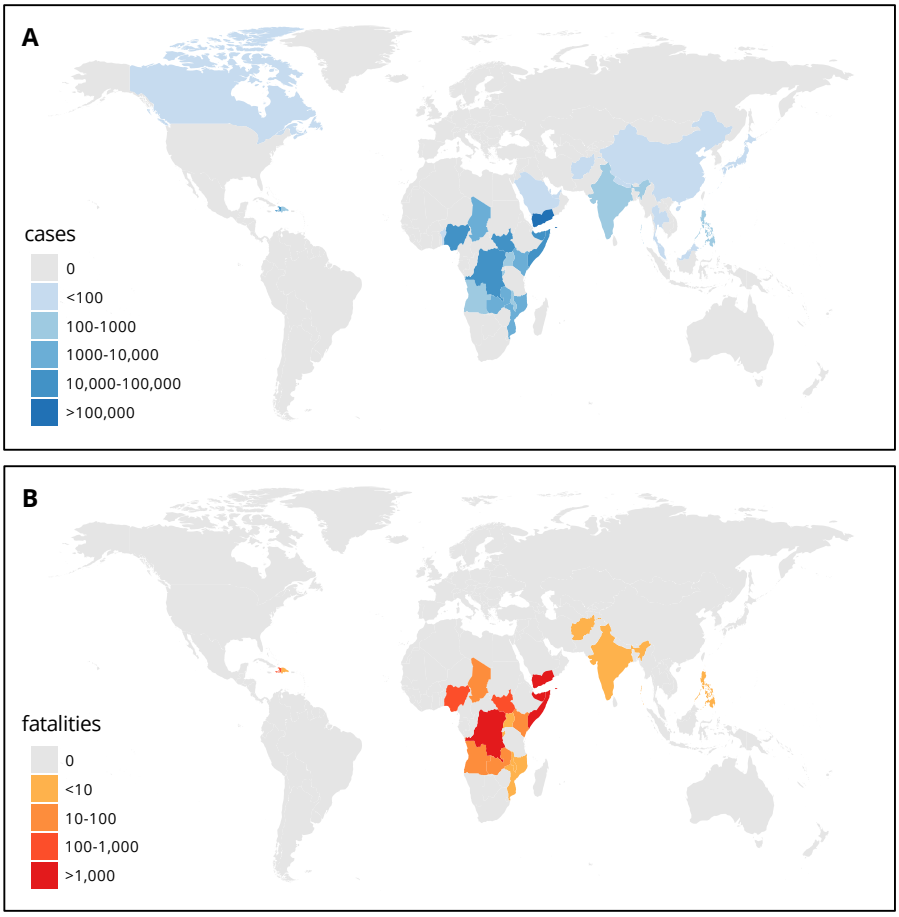

Figure 1.6: The global burden of cholera. In 2017 the WHO recorded 5,654 deaths due to cholera and a total of 1, 227, 391 cases. A) Cholera cases reported to WHO, B) Cholera fatalities reported to WHO.

changes to pH (Escobar et al., 2015). Climate change threatens not only to increase its range due to SST warming but also the number and frequency of extreme weather events, which are associated with non-endemic outbreaks. In other cases, conflict or natural disasters have made populations more vulnerable to cholera.

Diversity of *V. cholerae*

There are more than 200 serogroups of *V. cholerae* but only two (O1 and O139) cause cholera (Hsueh and Waters, 2019). Other strains are known to cause sporadic or less severe diarrhoea. *V. cholerae* serogroup and serotype classifications are determined by the composition and methylation state of the lipopolysaccharide (LPS) O antigen (Sack et al., 2004). Biotypes are assigned using other biochemical and phenotypic characteristics such as sensitivity to polymixin B antibiotics and acetoin production (Clemens et al., 2017).

Toxigenic *V. cholerae* strains carry the genes encoding the toxin co-regulated pilus (TCP) and cholera toxin (CT) on mobile genetic elements. The CT genes are located on the filamentous CTXΦ phage, which is able to integrate into the chromosome or replicate as a plasmid (Sack et al., 2004). Novel strains can acquire the CTXΦ phage only if they carry the TCP genes, located on the vibrio pathogenecity island 1 (VPI-1), since the TCP is the receptor required for the acquisition of the phage (Murphy and Boyd, 2008). This means that non-toxigenic *V. cholerae*, particularly in regions such as the Bay of Bengal where they are most diverse, are potential reservoirs for the emergence of novel pathogenic types.

The first six cholera pandemics were caused by strains of the *V. cholerae* O1 Classical biotype (Hsueh and Waters, 2019). The ongoing seventh

pandemic is caused by strains of the *V. cholerae* O1 El Tor biotype which was first identified in 1976 in the Egyptian Sinai peninsula (Cvjetanovic and Barua, 1972). El Tor strains are better adapted for long-term environmental survival but cause a less-severe disease than the Classical biotype (Cvjetanovic and Barua, 1972). El Tor strains have characteristic genomic islands (GI) such as the vibrio seventh pandemic island II which are thought to contribute to its improved environmental fitness (Nguyen et al., 2018). O1 hybrid variants that combine the environmental fitness of the El Tor biotype with the pathogenecity of the Classical have been described (Safa et al., 2010).

A new encapsulated serogroup (O139) emerged in Indonesia in 1992 (Comstock et al., 1996). O139 strains have caused series of outbreaks in India and Bangladesh (Alam et al., 2006). However, O1 El Tor strains remain the dominant organism.

A changing paradigm

The presence of *V. cholerae* within aquatic environmental reservoirs is well-established. Endemic outbreaks are seasonal and correlate with zooplankton blooms, chlorophyll concentration and increased sea surface temperature (Colwell, 1996). These factors provide an opportunity of *V. cholerae* to multiply and increase the risk of human infection via contaminated water. Transmission during outbreaks of cholera can be human-human or associated with repeat introductions from the environment (Garrine et al., 2017). Studies of the global dissemination of *V. cholerae* have implicated copepod and other invertebrate species, fish and birds as vectors (Rawlings et al., 2007, Senderovich et al., 2010, Hounmanou et al., 2019, Laviad-Shitrit et al., 2019).

A genomic analysis done in 2011 of 136 clinical isolates reported a novel

pattern for the global dissemination of cholera (Mutreja et al., 2011). Sequencing showed that seventh pandemic El Tor strains spread globally in three waves. In each wave a new strain replaced the previous. The El Tor isolates sequenced only differed from the reference N16961 (isolated in Bangladesh, 1975) by up to 250 single nucleotide polymorphisms (SNPs). This demonstrated that the seventh pandemic originated from a single source and suggested that humans were largely responsible for its dissemination. Wave 1 spread globally, waves 2 and 3 have been mostly restricted to Asia and Africa (Domman et al., 2017, Weill et al. (2017)).

1.9 Life cycle of *V. cholerae*

V. cholerae has a complex life cycle that involves survival in relatively cooler, nutrient deplete aquatic environments and within hosts (Figure 1.7). *V. cholerae* can colonise alternative host organisms such as fish and amoeba (Senderovich et al., 2010, Zago et al., 2017, Van der Henst et al., 2016, Hounmanou et al., 2019). The ability to switch between a motile planktonic state and sessile surface-attached biofilm communities is vital, and contributes to disease transmission and environmental survival (Faruque et al., 2006). In the environment some bacteria enter a quiescent 'viable non-culturable' state that allow them to persist in unfavourable conditions and makes them difficult to detect (Alam et al., 2007). Alternatively, phase variation produces biofilm-like rugose variants (Yildiz and Schoolnik, 1999). These mechanisms have in common the production of an extracellular matrix, largely formed of vibrio polysaccharide (VPS), which enhances protection from predatory protozoa, resistance to chlorinated water, osmotic and oxidative stress (Matz et al., 2005).

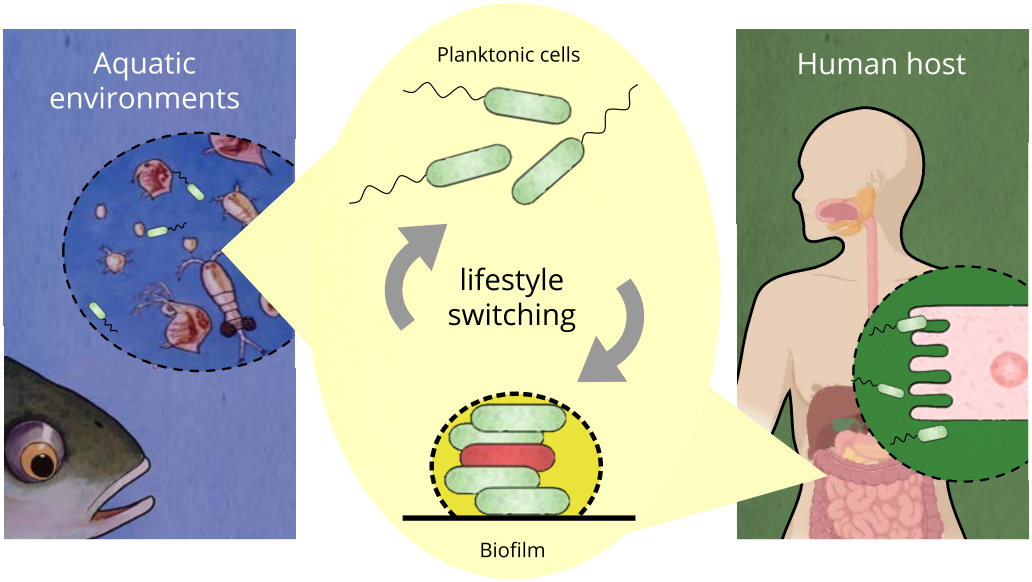

Figure 1.7: *Vibrio cholerae* life cycle. The shedding of motile and aggregate cells from a human host contaminates aquatic environments, where they can persist. Cells can remain free-swimming, form matrix-enclosed biofilms on chitinous surfaces (plankton, detritus) or colonise aquatic hosts (fish). Some cells enter a quiescent, viable non-culturable state (in red). The ingestion of contaminated sea-food or water infects the epithelial cells lining the human small intestine.

Viable but non-culturable (VBNC)

VBNC are more challenging to identify because they do not grow on laboratory media, instead they must be detected using molecular techniques such as PCR (Alam et al., 2007). VBNC *V. cholaerae* are characterised by their spherical shape, condensation of the chromosomes, an extended periplasmic space and an invaginated cellular membrane (Brenzinger et al., 2019). VBNC remain infectious with elevated transcription of virulence genes and can infect mice (Vora et al., 2005, Xu et al., 2018, Krebs and Taylor, 2011). The VBNC state requires the biofilm matrix genes and cells are typically encased a biofilm-like matrix (Faruque et al., 2006, Kamruzzaman et al., 2010).

In studies using artificial seawater the quorum sensing regulator HapR was shown to delay progression to the VBNC state (Wu et al., 2020). Naturally occurring mutations to QS regulators HapR and LuxO are common and may contribute to variation in the formation of quiescent cells (Joelsson et al., 2006). Furthermore, the QS autoinducer CAI-1 can resuscitate VBNC cells and provides a mechanism for resuscitation in environmental reservoirs (Naser et al., 2019).

Phase variation

Phase variation of *V. cholerae* is a type of population heterogeneity that results in the growth of rugose variants (Beyhan and Yildiz, 2007). In contrast to smooth cells, rugose variants form floating aggregates and pellicles in broth culture and on agar plates have a corrugated colony morphology (Yildiz and Schoolnik, 1999). The transition can occur spontaneously under carbon limitation, when grown on nutrient-poor media, by exposure to antibacterials, and during biofilm formation (Yildiz and Schoolnik, 1999, Ali et al., 2002). Rugose variants have a reduced

motility profile and an increased propensity to form biofilm which is thicker but more irregularly structured compared to biofilm produced by smooth variants (Casper-Lindley and Yildiz, 2004). Microarray analysis comparing gene expression in the two variants found 124 differentially regulated genes (Yildiz et al., 2004). Mutation to the flagellin *flaA* in the O139 biotype also results in cells locked into the rugose state (Watnick et al., 2001).

Phase variation occurs more frequently in toxigenic strains of *V. cholerae* (Ali et al., 2002). There are several molecular mechanisms that result in phase variation, including altered quorum sensing, flagellin production and nucleoside metabolism but the effect of these varies between strains (Beyhan et al., 2008). Rugose variants have been isolated from the environment, but their prevalence may be underestimated given that culture on TCBS (the routine agar used to identify *V. cholerae)* masks the rugose morphology (Ali et al., 2002).

Biofilm

Biofilms are surface-attached microbial communities enclosed within a self-produced matrix of extracellular polymeric substances; a mixture of polysaccharide, protein and nucleic acids (Flemming et al., 2016). *V. cholerae* biofilms are involved in disease transmission, host-colonisation and environmental survival (Teschler et al., 2015). The biofilm lifestyle provides fitness advantages by enhancing protection from chemicals, stomach acid and antibiotics, predation, as well as access to additional sources of nutrients (Bartlett and Azam, 2005).

The first step in the formation of biofilm is the establishment of microcolonies (Figure 1.8). Motility facilitates the search for a suitable surface, by switching between orbiting and roaming manoeuvres

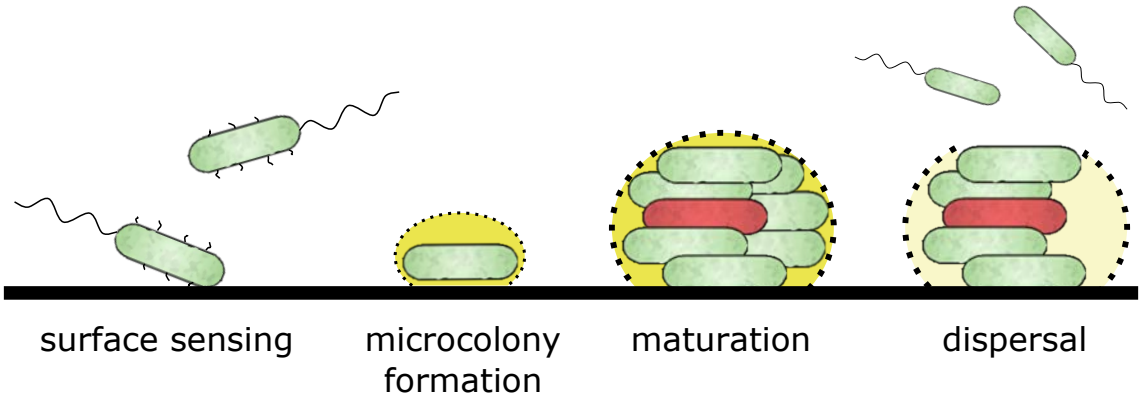

Figure 1.8: Cartoon showing development of biofilms. Motile *V. cholerae* test surfaces for attachment with the mannose-sensitive haemagglutinin pili (MSHA). Following surface attachment, a microcolony forms as the extracellular matrix is secreted. As the biofilm matures matrix production continues. Some cells may enter the quiescent VBNC state, (shown in red). During dispersal, the extracellular matrix is digested and cells disperse.

V. cholerae tests surfaces with the mannose-sensitive haemagglutinin (MSHA) pili (Utada et al., 2014). Microcolony formation occurs when cells become surface attached by the MSHA or the toxin-co regulated pilus. Biofilm formation proceeds with the production of the biofilm matrix. The matrix is composed of extracellular DNA, VPS and 3 proteins; RbmA, RbmC and Bap-1 (Teschler et al., 2015). VPS makes up half of the total mass and is essential for formation of three dimensional structures (Yildiz and Schoolnik, 1999). The matrix proteins RbmA and RbmC interact with VPS to produce higher-order structures (Hollenbeck et al., 2014).

During infection, intestinal biofilms provide protection and the simultaneous production of virulence factors is responsible for pathogenesis (Gallego-Hernandez et al., 2020). Abolishing the ability to form biofilm, by deletion of the *vps* and *rbmA* genes reduces intestinal colonisation in mice (Fong et al., 2010). The biofilm itself exerts mechanical stress that can further disrupt the epithelial monolayer (Cont et al., 2020). Production of the cholera toxin by bacteria in the biofilm results in the secretary diarrhoea characteristic of cholera, which provides the means of rapid dissemination from the host. Biofilm-like aggregates in stool are more virulent than dispersed cells (Gallego-Hernandez et al., 2020). Matrix-encased cells remain hyper-infective state for up to five hours (Merrell et al., 2002). Removing these aggregates from water is an effective means to reduce disease transmission (Huq et al., 2010).

In aquatic environments *V. cholerae* preferentially forms biofilm on chitin which can be used as a carbon and nitrogen source and is an abundantly available surface for attachment (Rawlings et al., 2007, Meibom et al., 2004). Furthermore, chitin induces natural competence, a process which promotes the exchange of DNA, and the emergence of new strains

(Meibom, 2005).

The process of biofilm dispersal is relatively poorly understood. A small number of proteins involved in matrix digestion, motility and signalling have been identified (Bridges et al., 2020). Two extracellular nucleases, Dns and Xds, contribute to dispersal and deletion reduces intestinal colonisation in mice, suggesting dispersal aids intestinal colonisation (Seper et al., 2011).

1.10 Transcriptional regulation of lifestyle switching

Biofilm formation is a costly process and transition from the planktonic state is tightly regulated. The *Vibrio cholerae* biofilm matrix cluster (VcBMC) is a major locus of transcriptional regulation. The VcBMC contains the two *vps* operons encoding the vibrio polysaccharide production proteins, separated by the *rbm* operon encoding two of the major biofilm matrix proteins: RbmA and RbmC (Yildiz and Schoolnik, 1999). Transcriptional regulation of the VcBMC is complex as there is convergence of multiple sensory stimuli. The histone-like nucleoid structuring protein (H-NS) is a negative regulator of multiple biofilm-related genes, including *vpsT* and the two *vps* operons (Wang et al., 2012). The cyclic AMP receptor protein (CRP) is another negative regulator of biofilm formation (Fong and Yildiz, 2008). Two positive regulators, VpsR and VpsT, are the master regulators of phase variation and biofilm formation (Beyhan et al., 2007). Both VpsR and VpsT are sensitive to the concentration of the c-di-GMP (Krasteva et al., 2010, Hsieh et al., 2020).

Cyclic-di-GMP in *V. cholerae*

The *V. cholerae* genome encodes 31 proteins with predicted DGC activity, 21 with PDE and 10 with both DGC and PDE activity (Beyhan et al., 2008). However, a screen for c-di-GMP reactive proteins in *V. cholerae* identified a total of 28, including DGCs, PDEs, and c-di-GMP responsive proteins (Roelofs et al., 2015). C-di-GMP in *V. cholerae* promotes rugose phase variation, biofilm formation, virulence gene expression, and maintenance of cell shape (Conner et al., 2017, Fernandez et al., 2020).

DGC and PDEs are multi domain proteins that generally include a sensory domain that detects change in the cells environment. Several host signals trigger increased c-di-GMP production. Bile acids stimulate the production of c-di-GMP via multiple DGCs, but this is counter-acted by bicarbonate released by epithelial cells (Hung et al., 2006). Polyamines present in the gastrointestinal tract are sensed by NspS, that inhibits the activity of the PDE MbaA (Karatan et al., 2005). Temperature is another important regulator of c-di-GMP levels in *V. cholerae.* The combined activity of six DGCs allows the bacteria adapt to low temperature shifts by increasing c-di-GMP production (Townsley and Yildiz, 2015). The activity of DGCs can have large phenotypic consequences, for example, a single point mutation to the *vpvC (VC2454)* results in cells that are locked in the rugose state (Beyhan and Yildiz, 2007). However, even though rugose variants have a higher concentration of c-di-GMP, deletion of DGCs does not abolish rugosity (Lim et al., 2006). Furthermore, there is not a strong correlation between c-di-GMP concentration and biofilm formation (Massie et al., 2012). However, selective activation of some individual DGCs does correlate strongly with induction of biofilm formation, suggesting that localised pools of c-di-GMP might be more important for regulating biofilm formation.

Among different isolates and biotypes of *V. cholerae* the concentration and regulation of c-di-GMP is varied (Satchell et al., 2016). For example, the PDE VieA is effective only in the Classical biotype (Tischler and Camilli, 2004). The extent to which the differences in response to c-di-GMP influence the enhanced environmental fitness of El Tor compared to Classical strains is unknown.

C-di-GMP can exert transcriptional regulation via the transcription factors: VpsR, VpsT and FlrA, and post-transcriptional regulation via riboswitches (Conner et al., 2017). FlrA is a regulator of motility genes. There are two classes of c-di-GMP sensitive riboswitches in *V. cholerae.* The Vc1 riboswitch positively regulates gene expression of the chitin adhesin GbpA (Kariisa et al., 2016). In contrast, the Vc2 riboswitch negatively regulates the expression of *tfoY*, a transcription factor that regulates motility and type VI secretion (Pursley et al., 2018).

VpsR

VpsR was the first of the master regulators of biofilm formation in *V. cholerae* to be identified, it is homologous to sigma^{54}-dependent activators and the NtrC subclass of 2-component response regulators (Yildiz et al., 2001). VpsR directly activates transcription of the two *vps* operons and *rbm* matrix proteins (Hsieh et al., 2018, 2020). VpsR is sensitive to c-di-GMP but it can function in its absence, however, its affinity for some promoters is increased in the presence of c-di-GMP (Tischler and Camilli, 2004, Hsieh et al., 2020). Deletion of *vpsR* abolishes the ability to form biofilm (Beyhan et al., 2007).

VpsT

The second major activator of biofilm formation is VpsT, a LuxR-type, 25.8 kDa member of the FixJ, CsgD response regulator family of proteins (Casper-Lindley and Yildiz, 2004). No cognate kinase has been identified for VpsT and its activity is independent of phosphorylation (Krasteva et al., 2010). VpsT has a C-terminal helix-turn-helix domain and an atypical N-terminal receiver (REC) domain that binds a dimer of c-di-GMP (Krasteva et al., 2010). VpsT requires c-di-GMP to form the DNA-binding competent dimer. C-di-GMP independent dimerisation inhibits DNA binding. The conserved helix-turn-helix domain interacts with the DNA major groove and is thought to be capable of introducing bends and DNA looping. The sub-cellular localisation of VpsT is dependent on both DNA binding and c-di-GMP (Shikuma et al., 2012). Despite its role in promoting biofilm formation deletion of *vpsT* does not abolish biofilm formation (Casper-Lindley and Yildiz, 2004).

Regulation of VpsT

Production of VpsT is subject to multiple layers of regulation (Figure 1.9). The alternative sigma factor σ^{38} (RpoS) activates *vpsT* transcription as part of the stringent response to stress. VpsR activates *vpsT* transcription in response to the presence of c-di-GMP (Srivastava et al., 2011).

Quorum sensing is intricately linked to the regulation of *vpsT*, as summarised in Figure 1.10. The AI receptors, CqsS and LuxPQ act as kinases until the accumulation of AIs trigger the switch to phosphatase activity (Eickhoff and Bassler, 2018). At low cell density (LCD) AphA activates transcription of virulence and biofilm genes, such as *vpsT* (Waters et al., 2008). However, at high cell density (HCD) LuxO is

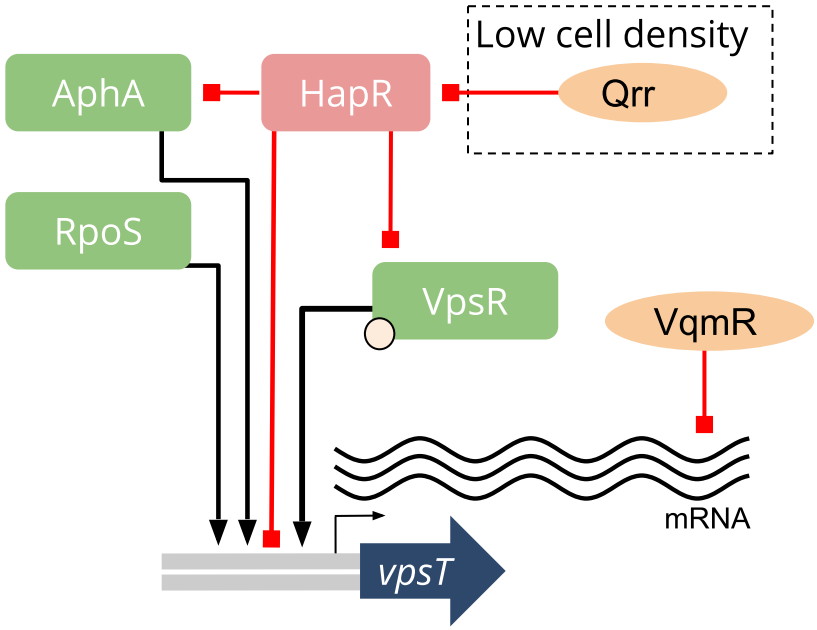

Figure 1.9: Regulation of *vpsT* transcription. Transcription of *vpsT* is activated by RpoS, AphA and VpsR. HapR represses *vpsT* transcription. AphA and HapR are regulated by quorum sensing. VqmR is a small regulatory sRNA that inhibits translation of the *vpsT* mRNA.

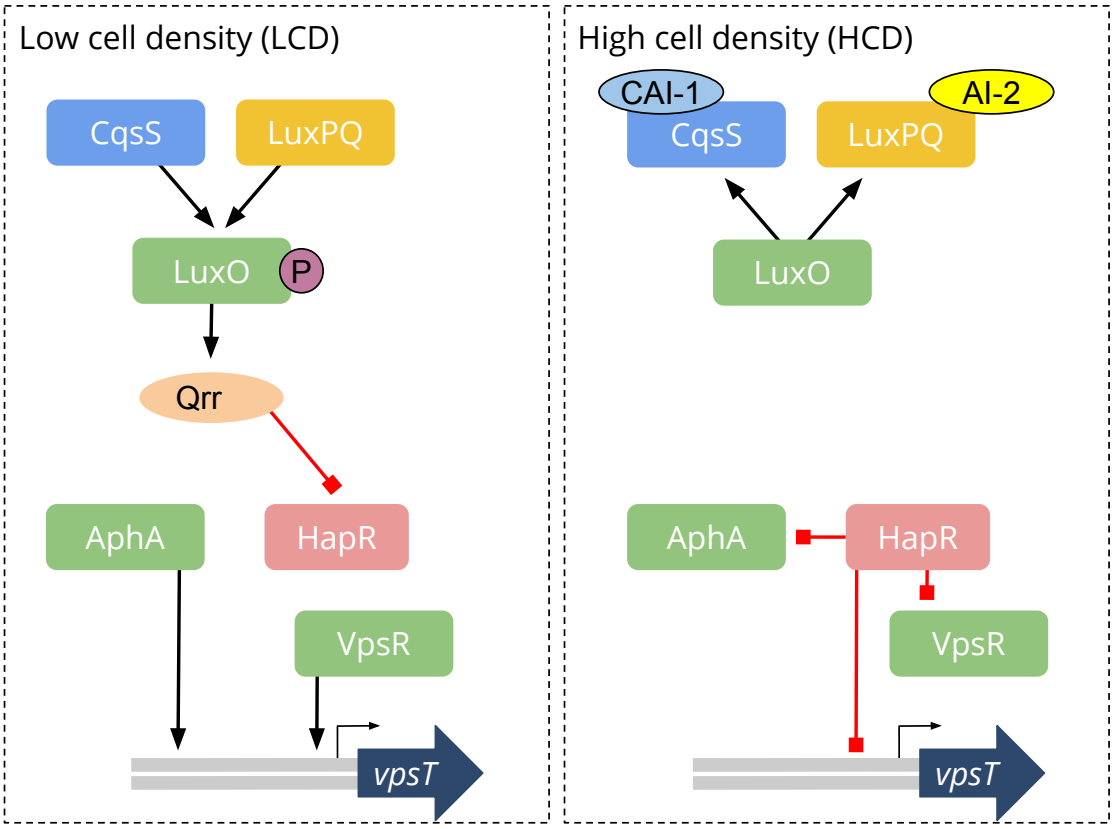

Figure 1.10: Quorum sensing and VpsT. The Quorum sensing system in *V. cholerae* has three parallel pathways that each respond to different Autoinducer molecules (only two are shown: CAI-1 and AI-2). QS behaviours are distinct at low (LCD) and high cell density (HCD) states, in response to the presence of AI molecules which are produced in proportion to cell density. CqsS and LuxPQ acts as kinases at LCD, phosphorlyating LuxO, which leads to production of the 4 Qrr sRNAs and inhibition of HapR. AphA is active and promotes biofilm formation. AphA and VpsR are activators of *vpsT*. At HCD CqsS and LuxPQ desphosphorylate LuxO and repression of HapR is relieved. HapR negatively regulates AphA, VpsR and transcription of *vpsT*.

dephosphorylated, the Qrr sRNAs are not transcribed and repression of HapR is relieved. HapR inhibits the formation of biofilm by directly repressing the expression of *vpsT, vpsR* and *aphA* (Yildiz et al., 2004).

HapR also represses the expression of multiple diguanylate cyclase genes that reduces the availability of c-di-GMP required for VpsT-DNA binding (Waters et al., 2008, Srivastava and Waters, 2012). Post-transcriptional regulation by the sRNA VqmR targets and degrades *vpsT* mRNA (Papenfort et al., 2015). VqmR transcription is activated by VqmA in response to the presence of a third autoinducer molecule 3,5-diemthylpyrazine-2-ol (DPO) (Papenfort et al., 2017).

VpsT binding targets

A small number of direct VpsT binding targets have been characterised (Figure 1.11). These include promoters that are both activated and repressed by VpsT. At the VcBMC VpsT indirectly activates transcription at the *vpsU, vpsA* and *rbmA* promoters by displacing and relieving H-NS repression (Ayala et al., 2015). VpsT activates transcription of the DNA repair gene *tag* and the catalase *katB* (Fernandez et al., 2018, Fernandez and Waters, 2019). This regulation is c-di-GMP dependent but the mechanism of activation at these promoters has not been elucidated. There are conflicting reports of auto-regulation of *vpsT* (Srivastava et al., 2011, Zamorano-Sánchez et al., 2015, Ayala et al., 2015).

The only characterised target subject to repression by VpsT is *rpoS* (Wang et al., 2014). However, transcriptional profiling and phenotypic assays suggest that VpsT also represses motility (Krasteva et al., 2010). The molecular mechanisms involved in regulation of motility are unknown.

Using the *vpsA* and *vpsL* promoters, binding motifs for both VpsR and VpsT have been described (Zamorano-Sánchez et al., 2015). However,

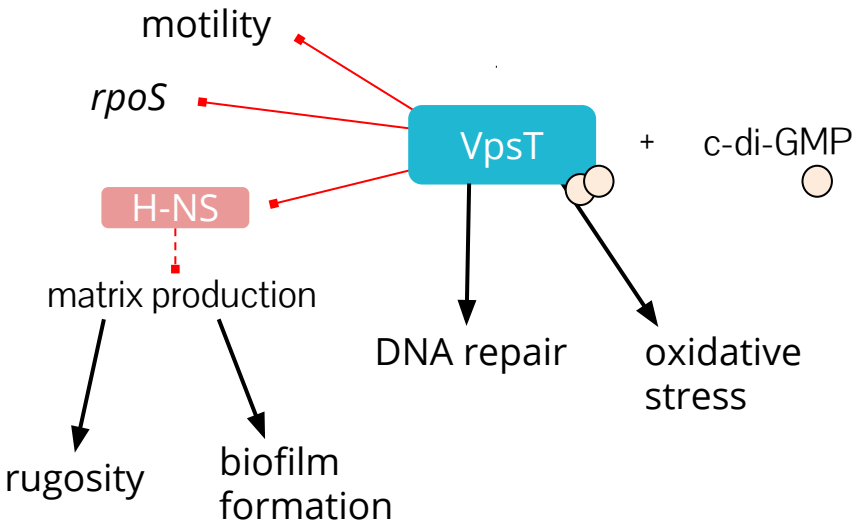

Figure 1.11: Currently characterised regulatory targets of VpsT. VpsT directly regulates matrix production by displacing, and relieving H-NS repression (dotted red line) at the *Vibrio cholerae* biofilm matrix cluster (VcBMC). VpsT can repress the transcription of *rpoS* (shown by red line). Microarray analysis of *vpsT* mutants suggests it also has a role regulating motility genes, but this may be indirect. VpsT can also activate transcription of a DNA repair gene (*tag*) and a catalase gene (*katB*) that reduces oxidative stress.

the identified VpsT 'T box' (TAAACTAAAGTTTA) differs from the VpsT binding site at the *rpoS* promoter (Wang et al., 2014). This suggested, contrary to the high degree of conservation in the published binding motif, that VpsT can tolerate degenerate binding sequences.

1.11 Aims of this work

The aims of this work are to further characterise the role of VpsT in *Vibrio cholerae* by:

- using ChIP-seq to identify binding sites across the genome
- refining the DNA binding sequence (T-box) that is recognised by VpsT
- studying the molecular mechanisms of repression and activation at VpsT targets
- investigating the consequences of VpsT regulation for c-di-GMP metabolism, motility and biofilm formation

Note on the use of 'we'

The work in this thesis is my own and the use of 'we' is not intended to imply otherwise. I am grateful for the use of a lab stock of RNA polymerase that was purified from *Vibrio cholerae* by Dr James Haycocks.

2 | Materials and Methods

2.1 Chemicals and reagents

Unless otherwise indicated all chemicals and media used were purchased from Sigma. All restriction enzymes are from NEB. The source of other reagents are indicated in the text.

2.2 Polymerase chain reaction

The commercially available polymerase Velocity (Bioline) was used routinely for polymerase chain reaction, following manufacturer instructions. MyTaq Red (Bioline) was used for colony PCR to check mutants for the presence/absence of genes or plasmids. Typically reactions were done in a total volume of 50 μl with 0.5 mM of each deoxynucleotide (NEB), 2 μM forward and reverse primers and 50-200 ng of template. PCR was done with 1 minute at 94/98 °C for polymerase activation followed by up to 35 cycles of: 30 seconds for DNA melting (94 or 98 °C), 30 seconds at the primer annealing temperature and 72 °C for elongation (30 seconds per kilobase), followed by a final extension (72 °C) of 5-10 minutes. The oligonucleotide primers used are listed in Table 2.1 and were synthesised by ThermoFisher.

PCR products were purified using the PCR Cleanup kit (Qiagen) or using the Agarose Gel Extraction kit (Qiagen) and eluted in 30 μL nuclease-free H_2O.

2.3 Nucleic acid extraction

Plasmid DNA was purified using either the Plasmid Mini and Maxi kits (both Qiagen) following the manufacturers instructions. DNA was eluted in 100 μL nuclease-free H_2O.

Genomic DNA was extracted from overnight bacterial cultures using the DNeasy Blood and Tissue Kit (Qiagen). RNA was extracted using the RNeasy Isolation kit (Qiagen), following the manufacturers instructions. DNA contaminants were removed using the TurboDNase kit (ThermoFisher).

2.4 Phenol-chloroform extraction

To extract DNA from mixed solutions an equal volume of phenol/chloroform/isoamyle alcohol (pH 8, 25:44:1) was added to the sample, vortexed to mix and then centrifuged for 3 minutes at 17,900 x g in a bench-top centrifuge. The upper aqueous layer was recovered for ethanol precipitation.

2.5 Ethanol precipitation of DNA

Ethanol precipitation was done by adding 0.1 volume of 3 M NaAc (pH 5.2) and 1 μl glycogen to the sample, followed by 3 volumes of ice-chilled 100 % (v/v) ethanol. Samples were mixed and then kept at -80 °C for at least 30 minutes. Samples were centrifuged at maximum speed in a bench-top centrifuge for 15 minutes at 4 °C. The supernatant was carefully discarded. The DNA pellet was washed twice with chilled 70 % (v/v) ethanol, dried under vacuum (Savant120 speedvac, Thermofisher) and then resuspended in nuclease-free H_2O.

2.6 DNA restriction digestion

All restriction enzymes were manufactured by NEB and restriction digests were done following the manufacturers instructions. Typically they were

done in a total volume of 50 μL, incubated at 37 °C for a minimum of 1 hour. Successful digestion was confirmed by agarose gel electrophoresis.

2.7 Agarose gel electrophoresis

Agarose gel electrophoresis was done routinely to check presence of correct DNA products (from PCR, restriction digestion etc.) Gels were 1 % (w/v) agarose, prepared in 1 x TBE (Fisher) and stained (1 % w/v) with SybrSafe (Invitrogen). Gels were ran at 100 V. DNA samples were mixed with 5 x blue loading dye (Qiagen) prior to gel loading. Either a 100 bp (NEB) or 1 kbp ladder (NEB) was ran alongside samples.

2.8 Polyacrylamide gel electrophoresis (PAGE)

Acrylamide gels (7.5 % w/v) were prepared using Protogel reagents (GeneFlow), 0.5 % TBE (Fisher) and polymerised with the addition of 0.01 volume 10 % (w/v) ammonium persulphate and 0.1 % (v/v) TEMED (N,N,N',N'-Tetramethylethylenediamine). Denaturing acrylamide gels (6 % w/v) were prepared using Urea Sequagel reagents (GeneFlow).

All gels were dried, exposed to a Fuji phosphor screen and then imaged using a BioRad Molecular Imager FX. Raw images were analysed using Quantity One (BioRad) and ImageJ.

Table 2.1: Oligonucleotide primer sequences

name	sequence	description
Oligonucleotide primers used to amplify VpsT target promoter DNA fragments by PCR		
TGO_0001	GGCTGCGAATTCCATTTACATTTGATTGGTATAATAATTCGC	Forward primer used to amplify VC0510 fragment
TGO_0002	GCCCGAAGCTTCATTTCGGGCTCCTGTCGGAATCAATGG	Reverse primer used to amplify VC0510 fragment
TGO_0003	GCCCGAAGCTTCATTTCGGGCTCCTGTCGGAATCAATGG	Forward primer used to amplify VC0512 fragment
TGO_0004	GGCTGCGAATTCCATTTCGGGCTCCTGTCGGAATCAATGG	Reverse primer used to amplify VC0512 fragment
TGO_0005	GGCTGCGAATTCTGACTTTCAGTGAGTCCATCCTGTTCTCC	Forward primer used to amplify VC1030 fragment
TGO_0006	GCCCGAAGCTTCACCGACCTGATTCCTAATGAGTGATCGC	Reverse primer used to amplify VC1030 fragment
TGO_0007	GGCTGCGAATTCCACTGTCCGTACCTCCCCATGCTTCGATATACACCCC	Forward primer used to amplify VC1031 fragment
TGO_0008	GCCCGAAGCTTCATCTCACGCTGATTGAGTTAAGTGTTGTTCAAAATGAGCC	Reverse primer used to amplify VC1031 fragment
TGO_0009	GGCTGCGAATTCCATTGCTCGCCTCGTTATTGTTATGCTTCCG	Forward primer used to amplify VC1303 fragment
TGO_0010	GCCCGAAGCTTCATTTGTGTAATTTGTGACAGAGAAACCG	Reverse primer used to amplify VC1303 fragment
TGO_0011	GGCTGCGAATTCCATTTGTGTAATTTGTGACAGAGAAACCG	Forward primer used to amplify VC1304 fragment
TGO_0012	GCCCGAAGCTTCATTGCTCGCCTCGTTATTGTTATGCTTCCG	Reverse primer used to amplify VC1304 fragment
TGO_0013	GGCTGCGAATTCCATATAAAGCCTCAGACCCTAATAC	Forward primer used to amplify VC1621 fragment
TGO_0014	GCCCGAAGCTTCATTACTCAATGAATGACCTTTAGTC	Reverse primer used to amplify VC1621 fragment
TGO_0015	GCCCGAAGCTTCATATAAAGCCTCAGACCCTAATAC	Forward primer used to amplify VC1620 fragment
TGO_0016	GGCTGCGAATTCCATTACTCAATGAATGACCTTTAGTC	Reverse primer used to amplify VC1620 fragment
TGO_0017	GCCCGAAGCTTCATCCGGTATCCCTAAACTGATGATG	Forward primer used to amplify VC1710 fragment
TGO_0018	GGCTGCGAATTCCATATTTTGGCCTCAAGATTAAGAGGGC	Reverse primer used to amplify VC1710 fragment
TGO_0019	GGCTGCGAATTCCATCCGGTATCCCTAAACTGATGATG	Forward primer used to amplify VC1711 fragment
TGO_0020	GCCCGAAGCTTCATATTTTGGCCTCAAGATTAAGAGGGC	Reverse primer used to amplify VC1711 fragment
TGO_0021	GCCCGAAGCTTCAAAGATTTACCTATTTTCGGGTTGTTG	Forward primer used to amplify VC1851 fragment
TGO_0022	GGCTGCGAATTCACACCGCCCTGCAATTCACTCCATTC	Reverse primer used to amplify VC1851 fragment
TGO_0023	GGCTGCGAATTCCAGTTAGTTGAAGAATTTCATACCGCCTTTATCCATCTC	Forward primer used to amplify VC2188 fragment
TGO_0024	GCCCGAAGCTTCATAGTTTGCTCTCCTATCGAGTTCGC	Reverse primer used to amplify VC2188 fragment

Table 2.1: Oligonucleotide primer sequences *(continued)*

name	sequence	description
TGO_0025	GGCTGCGAATTCTGCGGCGTTCATGCTTTGGTTCATTTC	Forward primer used to amplify VC2456 fragment
TGO_0026	GCCCGAAGCTTCAAGTGAAATATTCTCCATTTCGATCGTC	Reverse primer used to amplify VC2456 fragment
TGO_0074	GCATTTATCAGGGTTATTGTCTC	Forward primer to amplify insert in pSR
TGO_0075	CATCACCGAAACGCGCGAGG	Reverse primer to amplify insert in pSR
TGO_0076	GTTCTCGCAAGGACGAGAATTTC	Forward primer for sequencing pRW50 inserts
TGO_0077	AATCTTCACGCTTGAGATAC	Reverse primer for sequencing pRW50 inserts
Oligonucleotide primers used to amplify vpsT into the pAMNF vector		
TGO_0027	GGCTGCGGTACCAAAGATGAAAACAAACTAAACGTTAGAATGCTTTCTG	Forward primer to amplify VpsT for insertion into pAMNF vector
TGO_0028	GCCCGAAGCTTTTAAGAATTGACTTCCTCAATTCCAATATTATTTTTCGC	Reverse primer to amplify VpsT for insertion into pAMNF vector
TGO_106	TAACACCCCTTGTTTGCCTGG	Reverse primer for checking pAMNF insert
TGO_107	CGCCCTTCGCTGAAATTTATTC	Forward primer for checking pAMNF insert
TGO_108	TTGAAATTGCAGCTAAACTCTTTGTGAGTGAAAATACAGTAAAAA	Forward primer to quick change pAMNF::vpsT (D180A)
TGO_109	GAGTTTAGCTGCAATTTCAATATTAGAAGCACCACTACCAAGTAA	Reverse primer to quick change pAMNF::vpsT (D180A)
TGO_110	TTACTTGGTAGTGGTGCTTCTAATATTGAAATTGCAGATGCACTC	Forward primer to quick change pAMNF::vpsT (K181A)
TGO_111	TTTTTACTGTATTTTCACTCACAAAGAGTGCATCTGCAATTTCAA	Reverse primer to quick change pAMNF::vpsT (K181A)
TGO_112	AACTCAGTCGTGACCTCACAA	Forward primer for sequencing pAMNF::vpsT to check for mutation
Oligonucleotide primers used to amplify vpsT into pET28 for protein purification		
TGO_0029	CCCAATTCCATATGAAAGATGAAAACAAACTAAACGTTAGAATGCTTTC	Forward primer to amplify VpsT for insertion into pET28 vector
TGO_0030	GGCGGATCCTTAAGAATTGACTTCCTCAATTCCAATATTATTTTTCGC	Reverse primer to amplify VpsT for insertion into pET28 vector
Oligonucleotide primers used to insert a N-terminal 3XFLAG tag on vpsT		
TGO_0098	TAGAGGTACCGGTTGTTAACTACTAAACGAAATTACCGTATC	Forward primer for amplifying the upstrem flanking region of vpsT for making N-flagged version in pKAS
TGO_0099	TGGAATAAATCATTTCACCCCTCCTAAC	Reverse primer for amplifying the upstream flanking region of vpsT for making N-flagged version in pKAS

Table 2.1: Oligonucleotide primer sequences *(continued)*

name	sequence	description
TGO_0100	CCAAGGGTACCAAAGATGAAAACAAACTAAACG	Forward primer for amplifying the downstream flanking region of vpsT for making N-flagged version in pKAS
TGO_101	CGCCAGCTGCAGGCGGCCGCTAACATTTTTAAAGATCTGTTTGC	Reverse primer for amplifying the downstream flanking region of vpsT for making N-flagged version in pKAS
TGO_102	GGGTGAAATGATTTATTCCAATGTCACACAC	Forward primer to amplify FLAG tag from pAMNF to make N-flagged version in pKAS
TGO_103	TTTCATCTTTGGTACCCTTGTCATCGTC	Reverse primer to amplify FLAG tag from pAMNF to make N-flagged version in pKAS
TGO_0072	TTTGTCTTATTTTCTGAACATTG	Forward primer to sequence vpsT
TGO_0073	CCGACTGATGTTTCAGGTACTCAG	Reverse primer to sequence vpsT
Oligonucleotide primers used to delete vpsT		
TGO_117	ACTCGAAGAGACATTACACAACGAATTCACCCCTCCTAACACATCAAGGC	MuGENT primer to delete vpsT
TGO_118	GCCTTGATGTGTTAGGAGGGGTGAATTCGTTGTGTAATGTCTCTTCGAGT	MuGENT primer to delete vpsT
Oligonucleotide primers used to amplify VpsT targets with disrupted VpsT binding motifs		
TGO_0031	GGCTGCGAATTCGTGCTTAATTTTCC	Forward primer to amplify truncated (150nt) VC2456 fragment
TGO_0032	GGCTGCGAATTCCTTTGCACCCAACG	Forward primer to amplify truncated (110nt) VC2456 fragment
TGO_0033	GGGTGCAAAGAGTCATAAAATTCATTCTTCCGTTGG	Reverse primer to disrupt VpsT binding site in VC2456 fragment
TGO_0034	GGCTGCGAATTCGTTCTGTTTTTCCTTTC	Forward primer to amplify wildtype vpsL promoter
TGO_0035	GCCCGAAGCTTACTAGACGCTCCTAACC	Reverse primer to amplify wildtype vpsL promoter
TGO_0036	TGAAAATAAACTTTAGTTTAATTTTATGATGGTTAATAGG	Reverse primer to amplify mutant (g1) vpsL promoter
TGO_0037	ATTAACCACCATGAAAATAAAATTTAGTTTACTTTTATGATGG	Reverse primer to amplify mutant (g2) vpsL promoter
TGO_0038	CCAGTCATAACGATTTAAAGTAAAAGTATTAAATC	Reverse primer to disrupt VpsT binding site in VC1621 fragment
TGO_113	TAACCTATTAACCATCATACAAGTAAACTAAAGTTTATTT	Forward primer to amplify mutant vpsL promoter (Tbox position 2)
TGO_114	TTAACCATCATAAAAGTAAATTAAAGTTTATTTTCATGGTG	Forward primer to amplify mutant vpsL promoter (Tbox position 9)
TGO_115	ATAAAAGTAAACTAAAGTTGATTTTCATGGTGGTTAATAA	Forward primer to amplify mutant vpsL promoter (Tbox position 17)
TGO_116	AAAGTAAACTAAAGTTTATTGTCATGGTGGTTAATAAGTGA	Forward primer to amplify mutant vpsL promoter (Tbox position 21)
Oligonucleotide primers used to delete and complement the vpvABC operon		

Table 2.1: Oligonucleotide primer sequences *(continued)*

name	sequence	description
TGO_104	GCGGCCGCCTGCAGCTGG	Forward primer for making pKAS vector for vpsT-3xFLAG insert
TGO_105	GTTAACAACCGGTACCTCTAGAACTATAGCTAGCATGCGC	Reverse primer for making pKAS vector for vpsT-3xFLAG insert
TGO_0039	GGCTGCGAATTCCTATCTGAACTGATCCTGC	Forward primer to amplify vpvABC and promoter for complementation
TGO_0040	GCCCGGGATCCGTGCTTAATTTTCCCCAACGG	Reverse primer to amplify vpvABC and promoter for complementation
TGO_0041	GCCCGGGATCCGTGCTTAATTTTCCCCAACGGAAGAATGATTTTATGACTC	Reverse primer to amplify vpvABC and mutated VpsT site at promoter for complementation
TGO_0042	GTTTGCGGAGCACGAACTTCTCACAACC	Primer to check the vpvABC (A)
TGO_0043	AATTACCATCTACGGGAGAATCTAATCGG	Primer to check the vpvABC (B)
TGO_0044	CTTGTTCACTCACCACAGATTCCGGCGC	Primer to check the vpvABC (C)
TGO_0045	CTGCTGATCGATGTAAGTGGAGTTAATGCC	Primer to check the vpvABC (D)
TGO_0046	GTGAAAACACCAAAGCTAATGTTTGATCGG	Primer to check the vpvABC (E)
TGO_0047	CATGCTAGCTATAGTTCTAGA	Forward primer to check insertion into pKAS32
TGO_0048	TATCGATGGCGCCAGCTGCAG	Reverse primer to check insertion into pKAS32
TGO_0049	GAAACTGTGGTCAATACTGCGC	Forward primer to check vpvABC for deletion (200bp up/downstream)
TGO_0050	CGCCAAAGAGGCGGCTTCGA	Reverse primer to check vpvABC for deletion (200bp up/downstream)
TGO_0051	TCTGCGATTCCGGAAGAGCAACACGT	Primer to check insertion of pKAS32 at the vpvABC locus
TGO_0052	GGAACACTTAACGGCTGACATGGGA	Primer to check insertion of pKAS32 at the vpvABC locus
TGO_0053	GCTTGGCTAAAAAGCGCCCTTGTTGC	Primer to check insertion of pKAS32 at the vpvABC locus (from downstream)
TGO_0054	ATTTCCACCACAGCGCCAAC	Forward primer to sequence vpvA-P2
TGO_0055	TCCATGACTTTACCATTAGC	Reverse priemr to sequence vpvA-P2
TGO_0056	GGGGCGAGCATCGATGAGTTAGAGCCTGAGTTATTTGAGTTGCAAGAT	Forward primer to amplify megaprimer with vpvABC deletion
TGO_0057	TTAGTGATTGGTGAGACAAAAATACCCGTTGGGTGCAAAGAGTCAGAA	Forward primer to amplify megaprimer with vpvA-P2 deletion
TGO_0058	TTAGTGATTGGTGAGACAAAAATACCCGTTGGGTGCAAAGAGTCATAAAAT	Forward primer to amplify megaprimer with vpvA-P2 deletion and mutated VpsT binding site
TGO_0060	TAGAGGTACCGGTTGTTAACCCGACTATGTACTGCTCAAC	Forward primer to use with exo primers for insertion into the pKAS plasmid
TGO_0061	CGCCAGCTGCAGGCGGCCGCCGGATGGCATCACCATCCAT	Reverse primer to use with exo primers for insertion into the pKAS plasmid
TGO_0062	GTTGAGCAGTACATAGTCGGGTTAACAACCGGTACCTCTAGAACTATAGCTAGCATGCGC	Forward primer to amplify pKAS32 for deleting vpvABC

Table 2.1: Oligonucleotide primer sequences *(continued)*

name	sequence	description
TGO_0063	ATGGATGGTGATGCCATCCGGCGGCCGCCTGCAGCTGGCG	Reverse primer to amplify pKAS32 for deleting vpvABC
TGO_0064	TAGAGGTACCGGTTGTTAACTTGAAGATTGATTTGCTCAACT	Forward primer with homology to pKAS32 for deleting vpvA P2
TGO_0065	CGCCAGCTGCAGGCGGCCGCGGATGGCATCACCATCCATC	Reverse primer with homology to pKAS32 for deletng vpvA P2, use with exo deletion
TGO_0066	GATGGATGGTGATGCCATCCGCGGCCGCCTGCAGCTGGCG	Forward primer to amplify pKAS32 for deleting vpvA P2
TGO_0067	TTGAGCAAATCAATCTTCAAGTTAACAACCGGTACCTCTAGA	Reverse primer to amplify pKAS32 for deleting vpvA P2
TGO_0068	ATCACGAGGCCCTTTCGTCTTCAAG	Primer for sequencing pBR322 from EcoRI site
TGO_0069	CCGGCCACGATGCGTCCGGCGTAG	Primer for sequencing pBR322 from BamHI site
TGO_0070	GAGTCAGAAAACTCATTTTTCCATTGGGGAAAATTAAGCAC	Reverse primer to disrupt VpsT binding site (M2) in VC2456 fragment
TGO_0071	TTAGTGATTGGTGAGACAAAAATACCCGTTGGGTGCAAAGAGTCATAAAAT	Forward primer for exomugent to delete vpvA P2 and mutate T box
TGO_0080	GAAACTGTGGTCAATACTGCGC	Forward primer to sequence vpvABC operon
TGO_0081	CGCCAAAGAGGCGGCTTCGA	Reverse primer to sequence vpvABC operon
TGO_0087	ATGGATGGTGATGCCATCCGGCGGCCGCCTGCAGCTGGCG	Forward primer for making pKAS vector for dvpvABC insert
TGO_0088	AGTTGAGCAGTACATAGTCGCTCTGTACATGTCCGCGGTCGCGACG	Reverse primer for making pKAS vector for dvpvABC insert
TGO_0089	CTAGAGGTACCGGTTGTTAACGACTATGTACTGCTCAACTTAGCG	Forward primer for dvpvABC arm1
TGO_0090	ACGCCGTTAATCGCGAGAGGCTTCGAGC	Reverse primer for dvpvABC arm1
TGO_0091	CCTCTCGCGATTAACGGCGTTCAAGGATC	Forward primer for dvpvABC arm2
TGO_0092	CGCCAGCTGCAGGCGGCCGCATGCAGCGGAAATTGCTG	Reverse primer for dvpvABC arm2
TGO_0093	CTAGAGGTACCGGTTGTTAATTGAAGATTGATTTGCTCAAC	Forward primer for dvpvA-p2 arm1
TGO_0094	ACGCCGTTAAGTGCTTAATTTTCCCCAAC	Reverse primer for dvpvA-p2 arm1
TGO_0095	AATTAAGCACTTAACGGCGTTCAAGGATC	Forward primer for dvpvA-p2 arm2
TGO_0096	CGCCAGCTGCAGGCGGCCGCGGATGGCATCACCATCCATC	Reverse primer for dvpvA-p2 arm2
TGO_0097	TAGAGGTACCGGTTGTTAACGCTAGAAAAAGCGCGCGC	Forward primer for dvpvA-p2 (mutated VpsT site) arm1

2.9 Strains and plasmids

All strains used are listed in Table 2.2. Strains of *Escherichia coli* were used for routine cloning and protein over-expression. Strains of *Vibrio cholerae* are derived from the parental wildtype El Tor E7946 (Miller et al., 1989). Mutants were constructed using the MuGENT method or by utilising the pKAS32 suicide plasmid for allelic exchange (Dalia et al., 2014, 2017, Skorupski and Taylor, 1996). Derivatives were constructed either by chitin induced transformation for pAMNF, pBR322 and pAT1662-derived plasmids (Panda et al., 1991) or by tri-parental conjugation for pFY4535 and pRW50T (Zamorano-Sánchez et al., 2019, Lodge et al., 1992, Manneh-Roussel et al., 2018, Goldberg and Ohman, 1984). Constructs were checked by colony PCR and/or plasmid purification followed by Sanger sequencing. Sequencing was done by Eurofins or the Functional Genomics facility at the University of Birmingham.

Naturally induced transformation

Bacterial cultures were grown overnight and then sub-cultured (100 μL) in fresh 5 mL LB broth and grown to mid-log phase OD (600 nm) of 1. Cells were pelleted by centrifugation (1,600 x g for 10 minutes) and washed twice with 500 μL 0.7 % (w/v) Instant Ocean (SwellUK) and then resuspended in 1 mL. The cell suspensions were added to tubes containing sterile chitin flakes and incubated at 30 °C for 48 hours. Up to 10 μg of plasmid or linear DNA fragment was added, gently mixed and left for a further 24 hours before the biofilms were disrupted. Transformants were recovered by adding 1 mL LB broth and incubating for 2 hours. The sample was then plated on selective media.

Table 2.2: Strains and plasmids used

Name	Description	Source
E. coli DH5	*endA1, glnV44, thi-1, recA1, relA1, gyrA96, deoR, nupG, purB20, 80dlacZΔM15, Δ(lacZYA-argF) U169, hsdR17(rK–mK+),*	NEB
E. coli T7 express		NEB
E. coli JCB387	*ΔnirB, lac*	Page et al. (1990)
E. coli S17 *pir*	*lacU169 (lacZM15), recA1, endA1, hsdR17, thi-1, gyrA96, relA1, pir* NA	
V. cholerae E7946	Wildtype SmR derivative of E7946, O1 El Tor Ogawa. Isolated in 1978 from Bahrain.	Miller et al. (1989)
Plasmids		
pET28a	High copy number protein expression plasmids. N-terminal His^6 -tag, Kan^r	
pAMNF	Expression plasmid for adding 3X N-terminal FLAG, Kan^r	Sharma et al. (2017)
pRW50T	Derivative of pRW50, a broad-host range *lacZ* expression vector, with the *oriT* from pRK, Tet^r, Tra+	Manneh-Roussel et al. (2018), Lodge et al. (1992)
pBR322	Cloning vector, Amp^r, Tet^r	Bolivar et al. (1977)
pSR	pBR322-derivative containing transcription terminator (that follows insert)	Kolb et al. (1995)
pBAD33	pBR322 orgin, expression plasmid, induced with arabinose Cm^r	
pAT1662	pBAD33::*VCA0956*-His^6, Cm^r	Tischler and Camilli (2004)
pRK2013	Helper plasmid for DH5 conjugations, *Kan^r, oriColE1, RK2-Mob+ Tra+*	
pKAS32	Suicide plasmid for mutant strain construction, Amp^r, counter-select with streptomycin	Papenfort Lab
pFY4535	pMMB67EH derivative $lacI^+$, Ptac promoter, Gm^r contains a c-di-GMP biosensor plus hok/sok cassette from pXB300	Zamorano-Sánchez et al. (2019)

Bacterial conjugation

Conjugations were performed by mixing the recipient *V. cholerae*, donor *E.coli* DH5-α and a DH5-α helper strain with the pRK2013 plasmid in a volumetric 1:1:2 ratio (Perez-Soto et al., 2017). When *E.coli* S17-λpir was used as a donor the helper strain is not required and instead the donor and recipients strains were mixed in a 1:1 ratio.

Overnight cultures (1 mL of each) were centrifuged at 1,600 x g for 2 minutes in a bench-top centrifuge. The cell pellets were washed twice with 500 µL of 0.9 % (v/v) NaCl. Then, the cell pellets were resuspended in 1 mL LB broth. The donor and recipient suspensions were mixed. The mixed cell suspensions were spotted onto LB agar plates containing no antibiotics. After overnight incubation at 30 °C, the spots were dislodged and resuspended in 100 μL LB which was then plated on selective LB or TCBS media containing the appropriate antibiotic(s) for selection. When TCBS was used successful conjugants were re-streaked on LB plates to minimise selection of false positives.

Preparation and transformation of calcium-competent cells

Overnight cultures were used to inoculate 50 mL LB broth and grown to an OD (650 nm) between 0.3-0.6. Cells were then chilled, harvested by centrifugation (1,600 x g for 10 minutes) and washed with 0.01 M chilled $CaCl_2$. Following overnight incubation on ice, cells were mixed with 50 % (v/v) glycerol and aliquots stored at -80 °C until use.

To transform the competent cells they were thawed, mixed with 1-5 μL of plasmid DNA and then kept on ice for 1 hour. Following a 2 minute heat-shock at 42 °C, cells were recovered with the addition of 700 μL

fresh LB or SOC broth (NEB) and incubated for 30 minutes at 37 °C, with shaking. The cells were harvested by centrifugation (1,600 x g for 2 minutes in a bench-top centrifuge), resuspended in 100 μL of the supernatant and plated on selective LB agar plates.

2.10 Antibiotics

Table 2.3 shows the antibiotic concentrations used in media, when required. All were filter sterilised prior to use and then stored at -20 °C.

Table 2.3: Antibiotics

antibiotic	concentration
ampicillin	100 μg/mL
chloramphenicol	10 μg/mL
gentamicin	10 μg/mL
kanamycin	50 μg/mL
polymixin B	6.25 μg/mL
spectinomycin	50 μg/mL
streptomycin	100 μg/mL
tetracycline	5 or 35 μg/mL

2.11 Growth conditions

Unless otherwise indicated, all bacteria were cultured in LB broth or on LB agar plates at 37 °C, broth cultures were aerated by shaking. TCBS agar plates were used to discriminate *V. cholerae* from mixed suspensions with *E. coli*. Biofilm and motility assays were conducted at 30 °C. Motility assays were performed on LB 0.3 % (w/v) agar plates.

2.12 VpsT purification

VpsT was purified as previously described, with some alterations (Krasteva et al., 2010). VpsT was tagged at the N-terminus (using the pET28a plasmid) and was over-expressed in T7 express cells. Cells were grown to an OD (600 nm) of 0.8 (incubated at 37 °C, with aeration). Over-expression was then induced with the addition of 1 mM IPTG and the culture was incubated at 16 °C for a further 18 hours. Cells were harvested by centrifugation and then lysed by homogenisation (EmulsiFlex-C3, Avestin).

The his-tagged VpsT was purified from the cell lysate by FPLC using an AKTA Explorer (Amersham Biosciences). First using a metal ion affinity column (HisTrap Nickel, GE), using buffers:

- HisTrap-Buffer A: 25 mM Tris-HCl pH 7.5, 550 mM NaCl, 20 mM imidazole
- HisTrap-Buffer B: 25 mM Tris-HCl pH 7.5, 550 mM NaCl, 500 mM imidazole

This was followed by a second purification step using an heparin-affinity column (GE) using buffers:

- Hep- Buffer A: 25 mM Tris-HCl pH 7.5, 100 mM NaCl, 5 % glycerol
- Hep- Buffer B: 25 mM Tris-HCl pH 7.5, 1 M NaCl, 5 % glycerol

The protein extract was dialysed overnight and concentrated using a Vivaspin column (Sartorius) before the his-tag was cleaved by thrombin digestion (BioVision). Protein concentration was assayed using the Bradford method (Alfa Aesar). The purified protein was stored in 30 % glycerol at -80 °C until use.

2.13 Radio-labelling DNA fragments

DNA fragments were excised from plasmid pSR by restriction digestion with HindIII and either EcoRI or AatII (for DNase I footprints) and treated with Calf alkaline phosphatase. DNA fragments were radio-labelled at the 5' end using [γ-^{32}P]-ATP (7000 Ci/mmol, Perkin Elmer) in a total volume of 20 μL with T4 polynucleotide kinase, following the manufacturers instructions. Labelling reactions were washed to remove unincorporated nucleotides using micro bio-spin chromatography columns (BioRad) prepared with sephadex G50 beads (GE). The reactions were passed through 2 columns by centrifugation (1,600 x g, 3 minutes) and the eluate (labelled DNA fragment) was stored at -20 °C until use.

2.14 Preparation of Maxam Gilbert GA ladders

Radio-labelled DNA was treated with formic acid to depurinate the adenine and guanine residues. The reaction was stopped with the addition of 0.3 M NaAc (pH 7). The DNA was ethanol precipitated, washed with 70 % (v/v) ethanol, dried under vacuum (Savant120 speedvac, Thermofisher) and then resuspended in 1 M piperidine solution to cleave DNA at the modified bases. The reaction was incubated at 90 °C for 30 minutes and then the DNA was ethanol precipitated, washed with 70 % (v/v) ethanol, dried under vacuum and resuspended in 20 μL loading dye (composed of 0.025 % (v/v) each bromophenol blue and zylene cyanol). GA ladders were then stored at -80 °C until use.

2.15 Transcription assays

In vitro transcription assays were done as previously described (Haycocks et al., 2015).

Composition of buffers required:

- 10x transcription buffer: 400 mM Tris acetate pH 7.9, 10 mM $MgCl_2$, 1 M KCl, 10 mM DTT
- NTP mix: 1.25 mM ATP, 1.25 mM CTP, 1.25 mM GTP, 62.5 µM UTP
- STOP solution: 97.5 % deionised formamide, 10 mM EDTA, 0.3 % (v/v) Bromophenol Blue/Xylene cyanol FF

Super-coiled pSR plasmid templates, each containing a promoter insert, were isolated using a Qiagen maxi-prep kit (Qiagen). The pSR template was mixed with 2 μM VpsT plus 50 μM c-di-GMP. The reaction was started with the addition of purified *V. cholerae* RNA polymerase and RpoD, purified by Dr James Haycocks (Manneh-Roussel et al., 2018) which incorporates [α^{32}P]-UTP (Perkin Elmer). Reactions were stopped by adding 20 µL of STOP solution. The RNA products were resolved on a denaturing polyacrylamide gel.

2.16 Electrophoretic mobility shift assays

Purified VpsT was used to perform electrophoretic mobility shift assays (EMSAs) with DNA probes synthesised by PCR. EMSAs were done as previously described (Grainger et al., 2008), using transcription buffer (described above) and Herring-sperm DNA as a competitive inhibitor. The labelled probes (10 nM) were mixed with purified VpsT (0.5 - 7.5 μM) plus or minus 50 μM cyclic-di-GMP. The reactions were incubated

at 37 °C for 20 minutes and then loaded onto a polyacrylamide gel and ran at 200 V.

2.17 M13 sanger sequencing reaction

M13mp18 phage DNA sequencing reactions (T7 sequencing kit, USB) were used to determine the length of the primer extension and in vitro transcription products. The template DNA was diluted in H_2O and mixed with 8 µl 2M NaOH. This was incubated at room temperature for 10 minutes, then mixed with 7 µl 3M sodium acetate (pH 5.1), 4 µl H_2O and 120 µl ice cold 100 % ethanol. DNA was purified by ethanol precipitation and dissolved in 10 µl nuclease-free H_2O To anneal the primer and template 2 µl of Universal Primer and 2 µl annealing buffer was added to the reaction and incubated at 65 °C for 5 minutes, then at 37 °C for a further 10 minutes. T7 DNA polymerase was prepared by diluting 1 in 5 in dilution buffer. The annealed primer-template mix was mixed with 3 µl Labelling Mix-dATP, 1 µl [α^{32}P]-dATP and 2µl T7 DNA polymerase. The reaction was incubated for 5 minutes at room temperature before 4.5 µl was removed and added to 4 pre-warmed micro-centrifuge tubes containing either A, C, G or T Mix-Short. The tubes were were incubated at 37 °C for 5 minutes. The reactions were terminated with 5 µl STOP solution. Prior to gel electrophoresis, the products were heated to 72 °C in a heat block.

2.18 Primer extension assays

A radio-labelled oligonucleotide primer which anneals to pRW50T was hybridised to the RNA samples extracted from cells with promoter constructs in the pRW50T plasmid. This was then ethanol precipitated.

The pellet was resuspended in 30 μL hybridisation buffer, vortexed and then incubated at 50 °C for 5 minutes. This was followed by 15 minutes at 75 °C and a further 3 hours incubation at 50 °C. The annealed RNA-primer was ethanol precipitated, resuspended in RNase-free H_2O and then reverse transcribed using AMV reverse transcriptase (Promega). The enzyme was inactivated by heating to 72 °C for 10 minutes. The sample was then treated with 10 mg/mL RNase (Qiagen) for 30 minutes to remove RNA. Following ethanol precipitation, the DNA pellet was resuspended in loading dye and resolved on a denaturing polyacrylamide gel.

2.19 DNase I footprinting

DNase I footprinting was done as previously described (Singh and Grainger, 2013). Labelled DNA was mixed with molar dilutions of purified VpsT plus 50 μM c-di-GMP and incubated at 37 °C for 10 minutes. DNase I (Roche) was used to digest the DNA at room temperature for 2-4 minutes. Digestion was stopped with 200 μL DNase I stop solution and the digested DNA was purified by phenol-chloroform extraction and ethanol precipitation. The DNA was dried under vacuum (Savant120 speedvac, Thermofisher) and resuspended in 4 μL DNase I blue. Foot-printing reactions were resolved on a 6 % polyacrylamide sequencing gel, and ran with a GA ladder.

2.20 β-galactosidase promoter fusion assays

Z- buffer was prepared in advance, as follows:

- Z-buffer: 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_42H_2O$, 0.75 g KCl, 0.25 g $MgSO_4$ in 1 L ddH_2O

Prior to use, Z-buffer was completed with the addition of 8 mg/ml 2-Nitrophenyl β-D-galactopyranoside (ONPG) and 0.27% (v/v) β-mercaptoethanol.

Promoter fusion assays were performed by cloning promoters upstream of *lacZ* in plasmid pRW50T, a derivative of pRW50 containing the pRK origin of transfer (Manneh-Roussel et al., 2018). These plasmids were introduced into the wild-type and *ΔvpsT* strains of *V. cholerae* El Tor E7946 by tri-parental conjugation. Promoter activity was inferred by production of β-galactosidase (Miller, 1972). Cultures were grown in LB broth or on LB agar plates (supplemented with 5 μg mL^{-1} tetracycline) and incubated at either 30 or 37 °C overnight. The assays were conducted either by suspending a single colony in 1 mL phosphate-buffered saline (PBS) or sub-culture and grown to OD (650 nm) between 1-1.2. The cultures were lysed using toluene and sodium deoxycholate, 3 drops of each were added to the cultures, vortexed and then left at 37 °C to allow the toluene to evaporate. 100 μL of each lysate was transferred into glass test tubes and mixed by vortexing with 2.5 mL of Z-buffer. The assays were incubated at 37 °C until yellowing, at which point the time was recorded and the reaction stopped with 1 mL of 1 M sodium carbonate. The absorbance of the yellow o-nitrophenol was

recorded using a spectrophotometer (420 nm) and the promoter activity (Miller units) was calculated using the equation:

$$promoter\ activity\ (Miller\ units) = \frac{1000 * 2.5 * v_{total} * Abs_{420}}{4.5 * t * v_{sample} * Abs_{650}}$$

where:

Abs_{420} is the absorbance of the yellow o-nitrophenol at 420 nm,
Abs_{650} is the optical density of the cell culture before lysis at 650 nm,
t = reaction time in minutes,
v_{total} = total volume of assay reaction in mL,
v_{sample} = volume of culture assayed in mL

2.21 Rugose colony switching

In order to assess the ability of cells to undergo phase variation, bacteria were incubated in a modified alkaline peptone water (APW) media. Overnight cultures of bacteria were diluted 1:1000 in 3 mL of APW and incubated statically for 72 hours at 30 °C (Ali et al., 2002). Aggregates were disrupted by vortexing with glass beads and then a serial dilution was plated on LB agar plates. We counted the number of smooth and rugose colonies on the plate following 18 hours incubation at 37 °C, to calculate the frequency of smooth-to-rugose switching.

2.22 Biofilm formation assays

Biofilm formation was assayed by crystal violet staining. Overnight cultures were sub-cultured and grown to and OD (600 nm) of 1. These were diluted 1:100 to an initial OD (600) of 0.01 and 100 μL was loaded

in triplicate to wells of a 96-well plate (Costar). The plate was incubated for 18 hours at 30 °C. Media was removed and then the wells were washed to remove unattached cells by submerging the plate in H_2O. The biofilms were stained with 125 μL 0.1 % (v/v) crystal violet solution for 15 minutes. Excess crystal violet was removed by submerging the plate in water. The plate was blotted dry and then stained biofilm was solubilised with 125 μL 100 % (v/v) ethanol. Absorbance (600 nm) was recorded using a FLUOstar OPTIMA (BMG) plate reader.

2.23 Motility assays

Motility assays were done on LB plates with 0.3 % (w/v) agar. Samples were inoculated on the plate by pipetting 1 μL into the surface of the agar. Plates were then incubated for 8-18 hours and the extent of the spread was measured using ImageJ.

2.24 Western blots

Buffers required:

- Grinding buffer: 50 mM Tris-HCl pH 8, 200 mM NaCl, Complete protease cocktail inhibitor
- Transfer buffer: 50 mM Tris pH 8.3, 40 mM glycine, 5 % (v/v) methanol
- Blot buffer: PBS, 0.1 % (v/v) Tween
- Blocking buffer: blot buffer, 5 % (w/v) milk powder

Sample cultures were grown to an optical density (650 nm) of 1. The cells were harvested by centrifugation (1,600 x g for 10 minutes) and then the pellet was stored at -80 °C until use. The cell pellets were resuspended

in grinding buffer and then sonicated (Diagenode BioRupter) for ten 30 second on/off cycles. The lysate was centrifuged (1,600 x g for 10 minutes) and the supernatant collected. Total protein concentration was measured by Bradford assay (Alfa Aesar) following the manufacturers instructions and using BSA (NEB) for the protein standards. For the gel 10 μg of total protein was mixed with 2 μL protein loading dye and boiled for 2 minutes prior to loading. Samples were loaded alongside 10 μL of Broad Range Protein Standard (NEB). Gels were pre-cast Novex 6-12 % bis-tris (Invitrogen) and were ran at 180 V for 1 hour. The gel, foam blot pads, filter paper and nitrocellulose membrane were soaked for 20 minutes in transfer buffer and then loaded into the blotting apparatus for the transfer. The transfer was done at 400 mA for 1 hour. The membrane washed and incubated in 10 ml blot buffer with 2 % (w/v) milk powder and 1 µl antibody (anti-FLAG) for 1 hour. The membrane was washed thoroughly and incubated in 10 ml blot buffer with 2 % (w/v) milk powder with 1 µl HRP-conjugated secondary antibody (anti-mouse) for 1 hour. The membrane was washed with blot buffer, excess buffer was removed and the membrane placed on plastic cling film. The two detection reagent components were mixed (2 mL of each) and then pipetted onto the membrane and left for 5 minutes at room temperature before imaging was done using a Amersham (GE) imaging system.

2.25 Quantification of intracellular c-di-GMP using a biosensor

The relative concentration of intracellular c-di-GMP was measured on LB agar plates with bacteria containing the c-di-GMP biosensor (pFY4535, (Zamorano-Sánchez et al., 2019)). After 18 hours incubation at 30 °C plates were imaged in three channels using a ChemiDoc MP system (Bio-

Rad) as done previously (Zamorano-Sánchez et al., 2019). In summary; Blue Epi with the 530/28 filter was used to detect fluorescence from AmCyan and green Epi with the 605/50 filter for TurboRFP. Image Lab (version 6.0.1) was used to measure the fluorescence intensity, with the same number of pixels recorded for each colony sampled. The relative fluorescence intensity (RFI) was then calculated using the ratio of mean TurboRFP and mean AmCyan fluorescence intensities.

2.26 Chromatin immunoprecipitation and DNA sequencing (ChIP-seq)

Buffers required:

- FA lysis buffer '150 mM NaCl': 50 mM HEPES-KOH pH 7, 150 mM NaCl, 1 mM EDTA, 1 % (w/v) Triton-X-100, 0.1 % (w/v) sodium deoxycholate, 0.1 % (w/v) SDS
- FA lysis buffer '500 mM NaCl': 50 mM HEPES-KOH pH 7, 500 mM NaCl, 1 mM EDTA, 1 % (w/v) Triton-X-100, 0.1 % (w/v) sodium deoxycholate, 0.1 % (w/v) SDS
- ChIP wash buffer: 10 mM Tris-HCl pH 8, 250 mM LiCl, 1 mM EDTA, 0.5 % (w/v) Nonidet P40, 0.5 % sodium deoxycholate
- ChIP elution buffer: 50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 % (w/v) SDS
- 1x TE buffer: 10 mM Tris-HCl pH 8, 1 mM EDTA

Cultures were grown to an optical density (650nm) of 1. DNA-protein complexes were cross-linked with 37 % formaldehyde solution and clenched with 2.5 M glycine. Cultures were then lysed with 4 mg/mL lysozyme and sonicated for ten 30 second cycles (Diagenode BioRuptor).

Immunoprecipitation was done using Protein A beads and anti-FLAG antibody, with rotation at room temperature for 90 minutes and transferred to Corning Spin-X columns (Sigma) for further processing. Following washes in FA lysis buffer and 0.01 M Tris-HCl (pH 7.5), the precipitate was blunted (NEB quick blunting kit), washed (2 FA lysis washes and 2 Tris washes) and then adenylated (NEB Klenow fragment (3’ -> 5’ exo-) kit). ChIP-seq bar codes (NEXTflex) were ligated using quick ligation kit (NEB), followed by three washes with 10.05 M FA lysis buffer and one 0.5 M FA lysis buffer wash. Then two further wash steps with the ChIP wash buffer and TE. The sample was eluted in ChIP elution buffer (65 °C for up to 20 minutes). Samples were de-cross-linked by boiling and cleaned using 1.1 x magnetic beads (Ampure). The beads were warmed, mixed and then placed on a magnetic rack for 2 minutes. The liquid phase was removed and then the beads were washed with 70 % (v/v) ethanol. Nuclease-free H_2O was used to elute DNA from the magnetic beads. DNA samples were quantified by qPCR, the DNA libraries were pooled and the sequencing was done using the MiSeq (Illumina) platform.

2.27 ChIP-seq data analysis

Unless indicated analyses were done with R and RStudio using scripts or packages from the Bioconductor project. The single-end reads from two biological replicates were mapped to the reference *V. cholerae* El Tor str.N16961 genome (chromosome I: NC_002505.1 and chromosome II: NC_002506.1) with Bowtie 2 using the QuasR package (Au et al., 2010). Read counts per base (pileup files) were produced using the coverage function from the IRanges package. MACS2 was used to call peaks using the standard configuration, which were then visually inspected

using the Artemis genome browser to confirm peak centres (Zhang et al., 2008). MEME was used to identify a VpsT binding motif from sequences extracted 125 nucleotides from each peak centre (Bailey and Elkan, 1994). For annotating ChIP-seq peaks wi th their nearest gene ChIP-seeker was used (Yu et al., 2015). Coverage plots for normalised average read-depth ChIP-seq data were produced using Gviz (Hahne and Ivanek, 2016).

2.28 Analysis of whole genome sequences

Genomic DNA was sequenced by a commercial supplier, Microbes NG. Sequences reads were aligned to the reference *V. cholerae* El Tor str.E7946 genome (GCA_002749635.1). BCFTools was used to process the BAM alignment files, produce read counts across the genome and identify single nucleotide polymorphisms (SNPS) (Danecek et al., 2021). SNP data was imported into R and the VariantAnnotation package was used to annotate SNPs (Obenchain et al., 2014).

3 | VpsT binding sites across the *Vibrio cholerae* genome

3.1 Introduction

Vibrio cholerae switches between motile and sessile lifestyles throughout its life cycle to promote environmental transmission, survival and pathogenesis. Vibrio polysaccharide (VPS) is a key component of the biofilm matrix and is also required for the rugose colonial morphotype (Fong et al., 2010). The VPS production proteins are encoded by two *vps* gene operons (Yildiz et al., 2004). Increased concentration of the second messenger cyclic-di-guanosine monophosphate (c-di-GMP) promotes production of VPS via two c-di-GMP sensing transcription factors: VpsR and VpsT (Tischler and Camilli, 2004, Krasteva et al., 2010, Hsieh et al., 2018). Transcriptional regulation of these two operons occurs at the promoter DNA upstream of *vpsA* and *vpsL*.

VpsR acts by directly activating gene transcription from the *vps* promoters (Hsieh et al., 2018). Deletion of *vpsR* reverts rugose morphology and abolishes the ability to form biofilm (Yildiz et al., 2001). Conversely, *vpsT* mutants lose the rugose morphology but maintain some ability to form biofilm (Casper-Lindley and Yildiz, 2004).

VpsT activates gene transcription from the *vps* promoters by displacing the histone-like nucleoid structuring (H-NS) protein (Ayala et al., 2015). Gene expression profiling using microarrays identified other potential VpsT targets and many of these are likely to be indirectly regulated (Krasteva et al., 2010). However, other direct VpsT targets have been characterised such as the catalase *katB* and DNA repair gene *tag* (Fernandez et al., 2018, Fernandez and Waters, 2019). A role for VpsT in regulating cell shape has also been proposed, via repression of *crvA*, however the mechanism for this is unclear and most likely indirect (Fernandez et al., 2020).

VpsT can also act as a repressor of gene transcription. This has been characterised for the alternative sigma factor gene *rpoS* (Wang et al., 2014). Microarray data suggests VpsT also represses the transcription of some genes involved in motility (Krasteva et al., 2010).

In this chapter we used chromatin immunoprecipation (ChIP) combined with whole genome sequencing (ChIP-seq) to identify direct binding targets for VpsT. ChIP-seq has been used previously to map genome-binding for the *V. cholerae* quorum-sensing regulator AphA (Haycocks et al., 2019).

3.2 Identifying direct targets of VpsT using chromatin immunoprecipation

To capture the complete regulon of VpsT ChIP was done with a derivative of *V. cholerae* El Tor str. E7946 which expressed an N-terminal 3XFLAG-tagged VpsT from the pAMNF plasmid (Haycocks et al., 2019). The bacteria were grown at 37 °C in LB media to mid-log phase and then DNA-protein interactions were cross-linked using formaldehyde. Immunoprecipitations were done using anti-FLAG antibody to selectively enrich for DNA bound by VpsT. A mock immunoprecipitation with no antibody was done as a control. The DNA was sequenced using the Illumina MiSeq platform. The sequence reads from duplicate experiments were aligned to the reference *V. cholerae* O1 El Tor str. N16961 genome. Reads were normalised to the same average read depth for visualisation (Figure 3.1). The average read depth is shown in light blue on the innermost track, the outer tracks represent genes in the forward and reverse orientation. To check that VpsT was produced at biologically relevant levels in these conditions, we also introduced a 3XFLAG tag at the native *vpsT* locus. We used Western blots to compare VpsT3XFLAG production from the plasmid and its native promoter in smooth and rugose phase variants (Figure 3.2). An anti-FLAG non-specific band is due to cross-reactivity with the outer membrane Porin4 (Shin et al., 2020).

Following alignment to the reference genome, we used MACS to identify VpsT binding peaks (Zhang et al., 2008). The peaks were inspected using the Artemis genome browser and one additional peak was added manually.

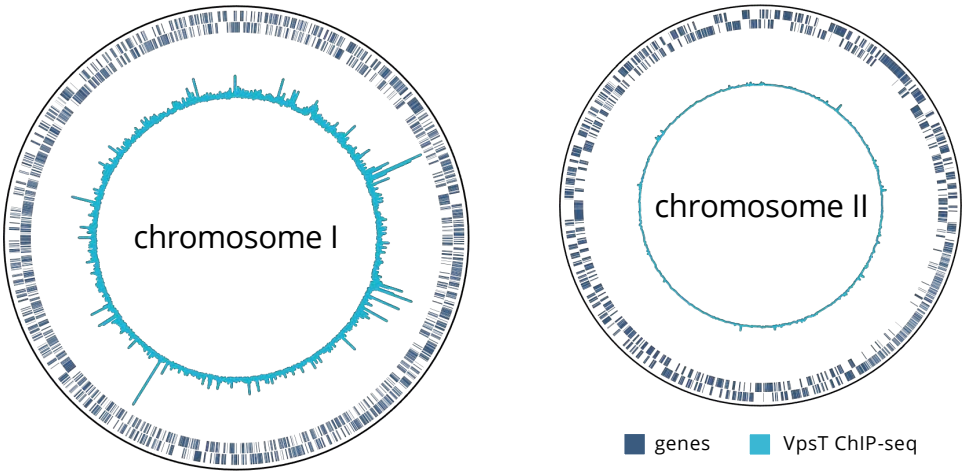

Figure 3.1: VpsT binding across the *V. cholerae* genome. ChIP-seq was done with wildtype *V. cholerae* O1 El Tor str.E7946 with an N-terminally 3XFLAG-tagged VpsT constitutively expressed from plasmid pAMNF. The aligned reads from two biological replicates were normalised for coverage and the average reads per base were mapped to the reference *V. cholerae* O1 El Tor str.N16961 genome. The two chromosomes are shown separately, genes are represented by navy blocks on the outer tracks and the average VpsT ChIP-seq reads are light blue.

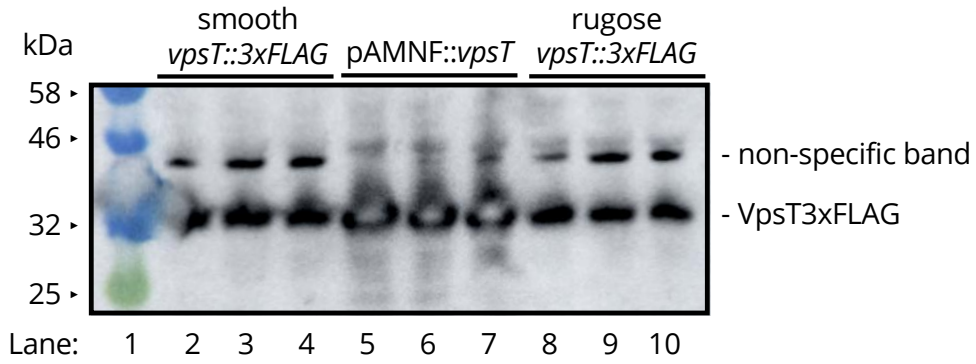

Figure 3.2: VpsT is produced at similar levels from the pAMNF plasmid. Western blot done to measure production of VpsT3XFLAG. Lane 1 is a molecular weight ladder, sizes are marked. The blot shows bacteria expressing *vpsT3XFLAG*, in Lanes 2-4 from its native promoter in smooth phase variants, Lanes 5-7 from plasmid pAMNF and Lanes 8-10 from its native promoter in rugose variants.

We identified 23 VpsT binding peaks on chromosome I and 0 peaks on chromosome II (Table 3.1). Elevated reads aligned to *vpsT (VCA0952)* on chromosome II are an artefact caused by plasmid encoded copies of *vpsT* and do not appear to be indicative of auto-regulation. Eleven of the 23 targets were known previously and include six that were identified by gene expression profiling but that have not been characterised (Krasteva et al., 2010). Two known targets *tag (VC1672)* and *katB (VC1587)* were not identified as peaks in our analysis, however weak enrichment for VpsT can be seen at these loci.

Table 3.1: VpsT binding peaks identified by ChIP-seq

peak centre[1]	adjacent gene(s)[2]	bound by H-NS[3]	known VpsT target[4]	motif sequence (5 -3)[5]
177184	*VC0175*	yes	no	TACTTTAGTG
383248	*VC0364*	no	no	AGGTTGGGTT
540341	*VC0510* <> *VC0512*	no	no	AGCCTTGGTT
544169	*VC0513*	no	no	TACTTTAGGT
564171	*VC0534*	no	yes*	AGGTTTAGCT
888647	*VC0825* <> *VC0826*	yes	no	CCCTATAGTG
891494	*VC0829*	no	no	TACATAGGTT
903390	*VC0841*	yes	no	ACCCAAACTT
979211	*VC0917*	yes	yes*	AACTTTAGTT
990915	*VC0928*	yes	yes*	AACCAAAGTT
999012	*VC0932* <> *VC0934*	yes	yes*	AACTTTAGTT
1096963	*VC1029* <> *VC1031*	no	yes	GAGTTAAGTG
1385301	*VC1303* <> *VC1304*	no	no	AAGTTTGGTT
1412875	*VC1329*	yes	no	TCCAAAAGTA
1741639	*VC1620* <> *VC1621*	no	yes	GACTTTAGTT
1845393	*VC1710* <> *VC1711*	no	yes	TACTTTAGTT
1947982	*VC1802* <> *VC1803*	no	no	ACGTTTTGTT
1990005	*VC1851*	no	no	AGCCAAAGCT
2221277	*VC2065*	no	yes	AAGTTAAGTT
2336167	*VC2188*	no	yes	CGCTTTGGTT
2376037	*VC2221*	no	no	CGCTTTAGGT
2640531	*VC2456*	no	yes	TGCTTTGGTT
2835743	*VC2667*	no	yes	CGCTTAAGCT

[1] Centre of the ChIP-seq peak was defined as the centre of the 250 bp region used to search for a VpsT binding motif

[2] Genes located close to the ChIP-seq peak, '<>' denotes a peak occurring upstream of two divergent genes.

[3] Overlapping H-NS ChIP-seq binding (Ayala et al., 2015)

[4] *: direct binding has been shown experimentally. Others identified by gene expression profiling (Krasteva et al., 2010).

[5] DNA sequence of the VpsT binding site identified using MEME.

The functions of the genes adjacent to VpsT peaks were extracted from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) protein database. Genes were classed as 'other' if their assigned function was not shared with another target gene. The number of genes in each functional class was plotted as a pie chart (Figure 3.3). Approximately a quarter (23 %) of the genes have no known function. 16 % of functions were not shared but included classes such as DNA repair and membrane components. Secretion system, biofilm formation, c-di-GMP metabolism and chemotaxis genes were overrepresented amongst target genes which have assigned functions and are consistent with the known role of VpsT.

Transcription factors typically bind to regions of DNA upstream of gene transcription start sites. Therefore, we expected VpsT peaks to occur upstream of annotated genes. We took the centre position of each peak and calculated the distance to the start codon of the adjacent gene(s). For peaks upstream of two divergent genes only the shortest distance was retained. The distance between the peak centre and its nearest gene was binned into groups of 100 bp intervals and plotted as a histogram (Figure 3.4A). We found that the majority of peaks were located upstream of genes. Most peaks were between 100 and 200 bp from the start codon. Two of the 23 peaks were downstream of their nearest gene start codon. We calculated the proportion of peaks occurring in coding and non-coding DNA which showed a bias for VpsT peaks in non-coding regions of DNA (Figure 3.4B).

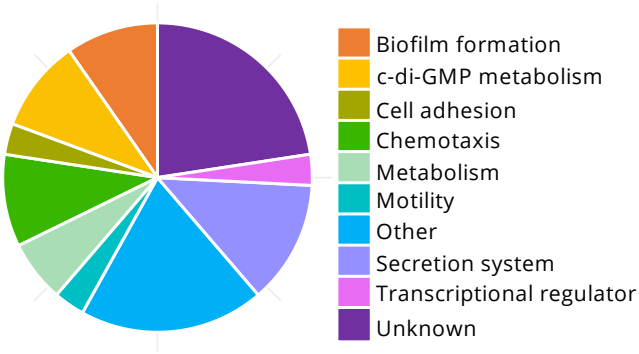

Figure 3.3: Biological functions associated with VpsT targets. Pie chart showing the functional classification of genes located adjacent to each of the ChIP-seq peaks. Protein function was sourced from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) protein database. Classes which contained only a single gene are grouped together as 'other'.

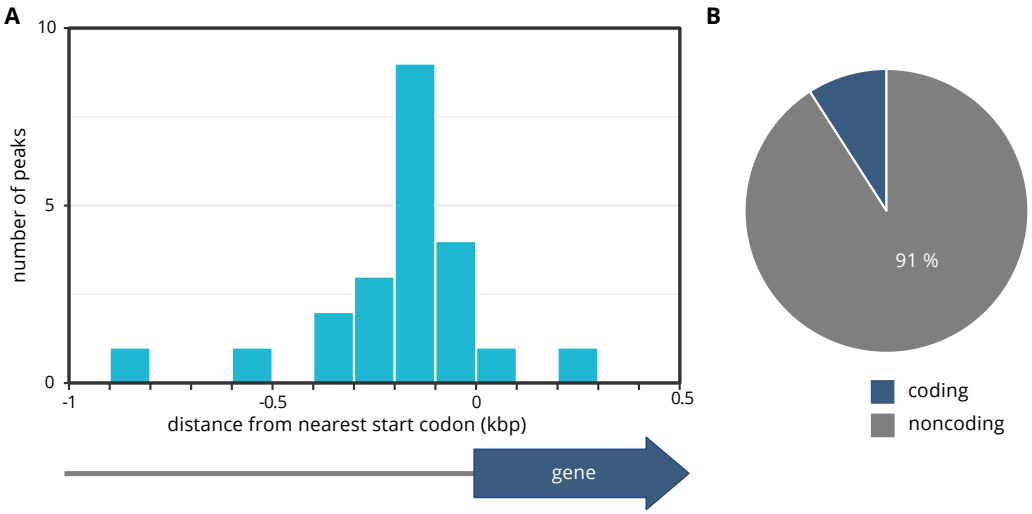

Figure 3.4: VpsT ChIP-seq peaks distance from nearest gene start codon. This plot shows the number of peaks that occur at distances of -1 kbp to +0.5 kbp from the start codon of their nearest adjacent genes. The distances were calculated from the centre of each peak to the nearest start codon, binned into groups of 100 bp intervals and plotted as a histogram. The schematic represents the DNA sequence upstream of a gene with a black line and the gene as filled-blue arrow. B) Pie chart showing the proportion of VpsT binding peaks that occur in non-coding and coding regions of DNA.

3.3 Identification and validation of the VpsT binding motif

A DNA binding motif for VpsT ('T-box') has been proposed before (Zamorano-Sánchez et al., 2015). The T-box motif is 22 nt in length and contains a highly conserved 8 nt A/T rich section at each end (Figure 3.5A). However, this consensus motif was based on DNase I footprinting of a single promoter (*vpsL*) and DNA sequence alignments to three other known targets of VpsT. The motif is not present at all VpsT targets, including the *rpoS* promoter (Wang et al., 2014). Using a larger number of target sequences should improve the resolution of motif predictions. Therefore, we used the 23 VpsT targets identified by ChIP-seq to search for a motif using MEME (Bailey and Elkan, 1994). To do this the DNA sequence 125 nucleotides up- and down-stream from the centre of each peak was extracted. MEME was used to search for a palindromic motif that occurred at least once in each 250 nt sequence. The DNA binding motif found by MEME for each peak is listed in Table 3.1 and the consensus VpsT 'T-box' motif is shown in Figure 3.5B. The consensus T-box motif is composed of an AT-rich core flanked at each end by a conserved C or G. The motif resembles the central portion (between position 7-16) of the longer 22 nucleotide binding motif (Figure 3.5A). There appears to be some flexibility in the orientation of the G-C pairs and this is reflected in the motifs identified by MEME at the the peak sites.

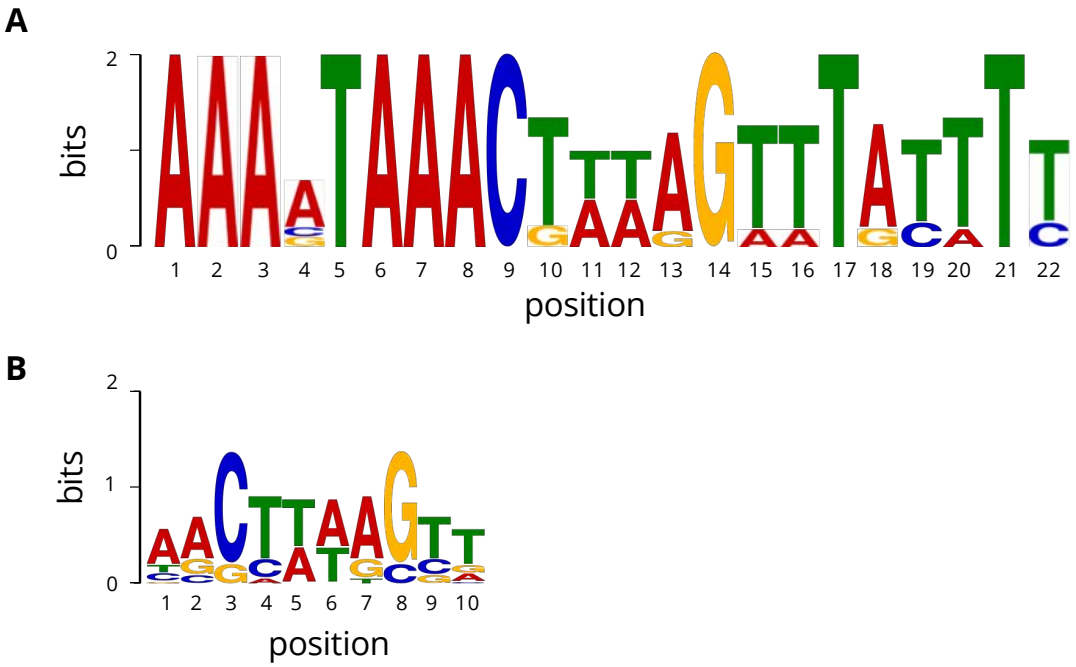

Figure 3.5: VpsT DNA binding motifs. A) Previously published DNA binding motif for VpsT that was inferred from only 5 known binding targets (Zamorano-Sánchez et al. (2015)). B) Motif inferred from the 23 VpsT binding targets identified by ChIP-seq. MEME was used to search for palindromic motifs that occurred at least once in each sequence. The DNA sequence 125 nt up- and down-stream of each peak centre was used.

We surmised that the T box DNA binding motif must occur in the genome at a higher frequency than VpsT binding peaks. To test this we used the FIMO (Find Individual Motif Occurrences) algorithm to scan the genome for occurrences of the consensus motif (Grant et al., 2011). In total FIMO found 332 binding sites, 235 on chromosome I and 97 on chromosome II. The number of sites identified using FIMO broadly correlates with the relative size of the two chromosomes but the actual number of sites (0 were identified on chromosome II) do not. We did the same analysis for FIMO sites as VpsT ChIP-seq peaks to see if they also occurred predominantly in non-coding regions of DNA (Figure 3.6). We found that in contrast to the position of VpsT ChIP-seq peaks the T-box motif typically occurred further from the nearest start codon and therefore more commonly in gene coding DNA.

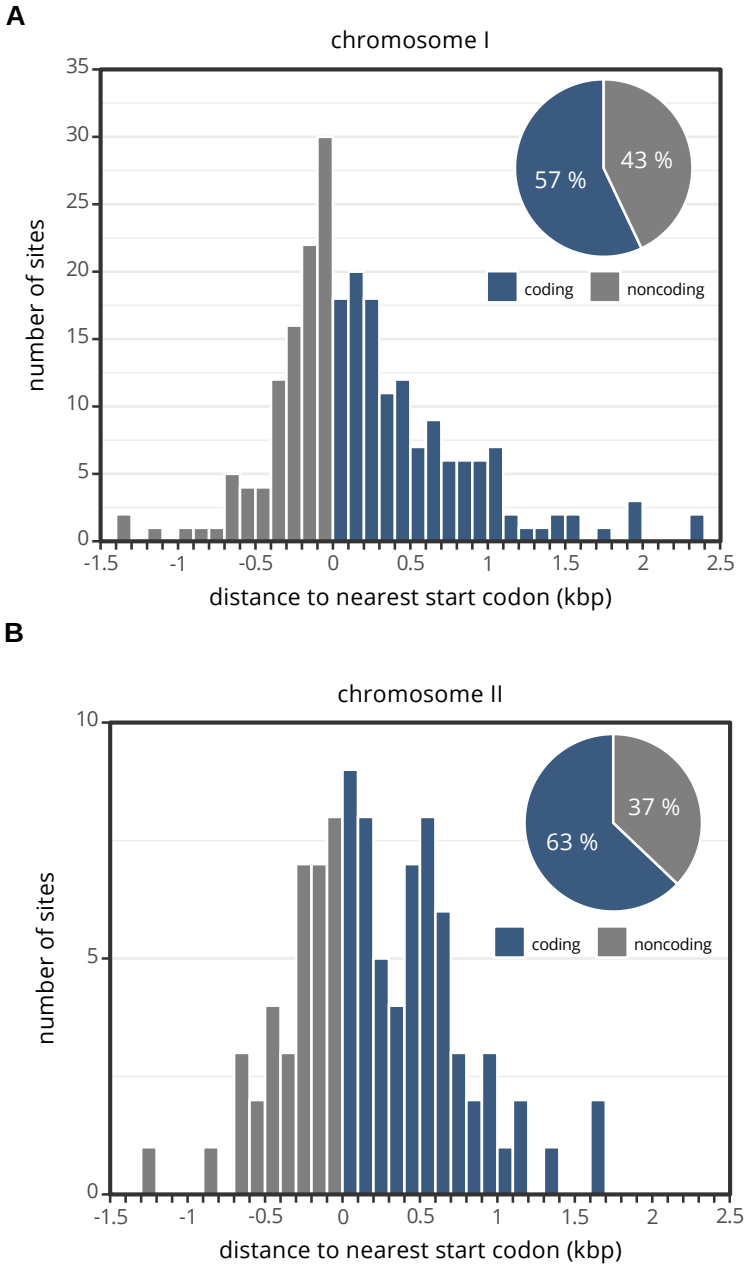

Figure 3.6: VpsT binding motif occurrences in the *V. cholerae* genome. This plot shows the number of motif sites that occur at distances between -1.5 kbp and +2.5 kbp from the start codon of their nearest adjacent genes. The distances were calculated from the centre of each motif site to the nearest start codon, binned into groups of 100 bp intervals and plotted as a histogram. A) shows results for chromosome I and B) for chromosome II. Inset pie charts show as a proportion the number of sites in gene coding (blue) and non-coding (grey) DNA.

The C at position 3 and the G at position 8 are the most highly conserved nucleotides in the T box motif (Figure 3.5B). As we noted previously, there is some flexibility at each of these positions for the order of the CG pair to be reversed, so that position 3 is a G and position 8 is a C. This is reflected in the motifs identified at the ChIP-seq targets (Table 3.1). We noticed that at the *vpsL* promoter sequence, there are two possible binding sites, a GC or CG pair. To identify which is required for VpsT binding we made two point mutations to guanine residues, 'G1' and 'G2' (marked in red, Figure 3.7A). DNase I footprinting assays were done to assess how this affected VpsT binding to the promoter of *vpsL* in two conditions, with and without c-di-GMP (Figure 3.7B and C). A Maxam Gilbert GA ladder was used to calibrate the gel and is labelled from the transcription start site (Hsieh et al., 2018). Three transcription start sites have been identified by primer extension at the *vpsL* promoter, the additional two are indicated with dotted arrows (Zamorano-Sánchez et al., 2015). In the absence of additional c-di-GMP there is no clearly defined VpsT footprint (Figure 3.7B). Mutating G1 altered the pattern of DNA cleavage, two bands between -210 and -220 are not present, due to a small change in DNase I sensitivity (Lane 4). With the addition of 50 μM c-di-GMP (Figure 3.7C) in Lane 3 a large region of DNA (-195 to -285) is protected from cleavage by VpsT, as indicated by the blue bar. There are bands, indicative of DNase I hypersensitivity, at nucleotide positions -215, -227 and -245. In Lane 4 there is a small shift of the hypersensitive band at -212 due to the mutation of G1. In Lane 5 the protection is lost as a result of the mutation made to G2 but the hypersensitive band at -227 appears unchanged.

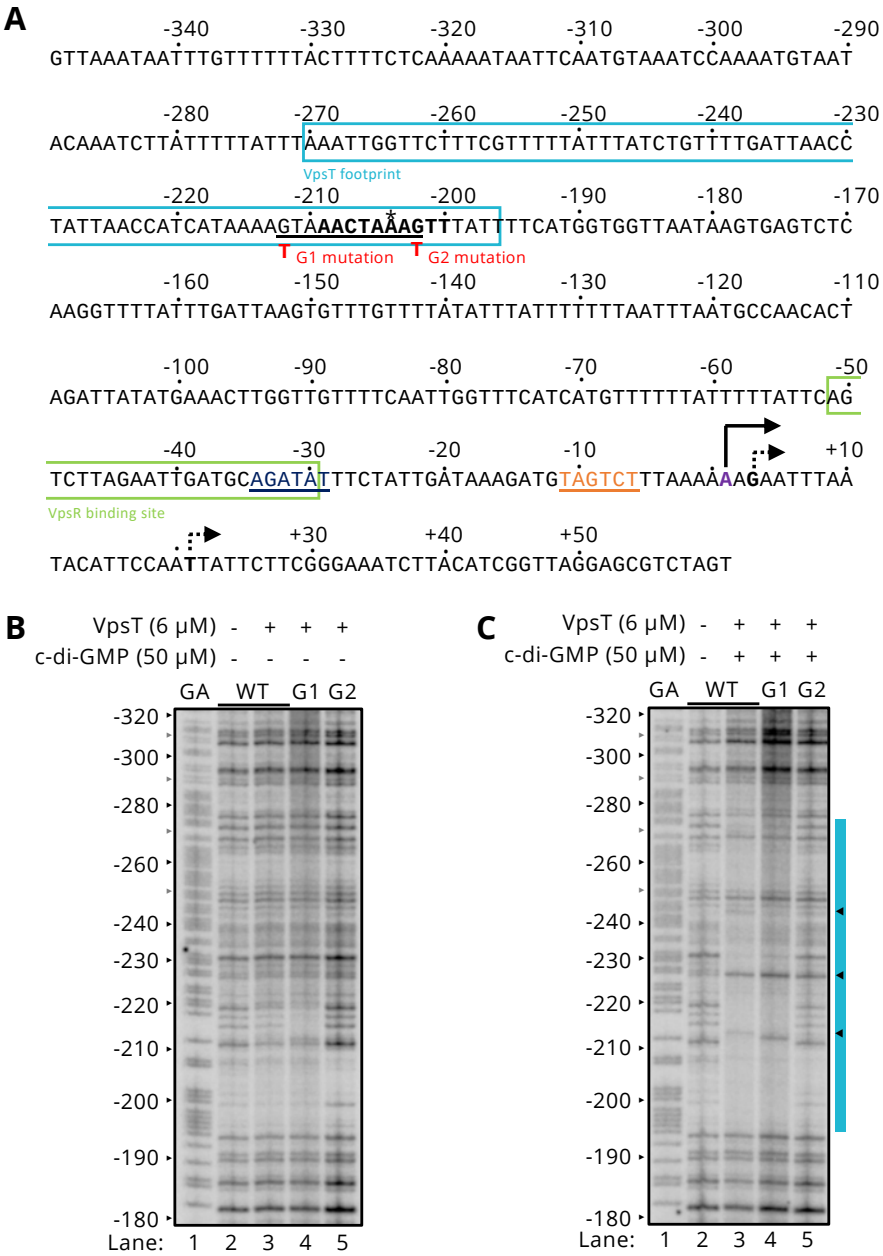

Figure 3.7: DNase I footprint of the promoter DNA upstream of *vpsL*. A) DNA sequence of the *vpsL* promoter. The VpsT binding motif (bold) is -203 nt from the *vpsL* transcription start site (bent arrow). Promoter -10 and -35 elements are coloured and underlined. VpsR and VpsT binding sites are boxed. The centre of the ChIP-seq peak is marked by an asterisk. Two point mutations made at the VpsT binding site are shown in red. B and C) DNase I footprints done with purified VpsT (6 μM). In C the experiment was done with 50 μM c-di-GMP. In both panels Lane 1 is a Maxam-Gilbert G+A ladder showing the distance from the transcription start site. Lanes 2 and 3 are the wildtype *vpsL* promoter DNA sequence without and with VpsT. Lanes 4 and 5 are mutated at either G1 or G2.

3.4 Validation of VpsT DNA targets identified by ChIP-seq

To check the validity of the ChIP-seq results eight of the novel VpsT targets were selected to confirm binding of VpsT to these regions of DNA in vitro. The regulatory region upstream of each was amplified by PCR and then used with purified VpsT for electrophoretic mobility shift assays (Figure 3.8). VpsT did not bind to a control DNA sequence taken from the promoter region of the *E. coli lac* operon (Figure 3.8A) but did bind to all the VpsT targets tested. Binding was unaffected by the absence of exogenous c-di-GMP which we expect is due to the presence of c-di-GMP that co-purified with VpsT. The atypical 'V'-shaped migration pattern has been observed with VpsT before and appears to be due to the dissociation of VpsT-bound c-di-GMP during electrophoresis (Ayala et al., 2015).

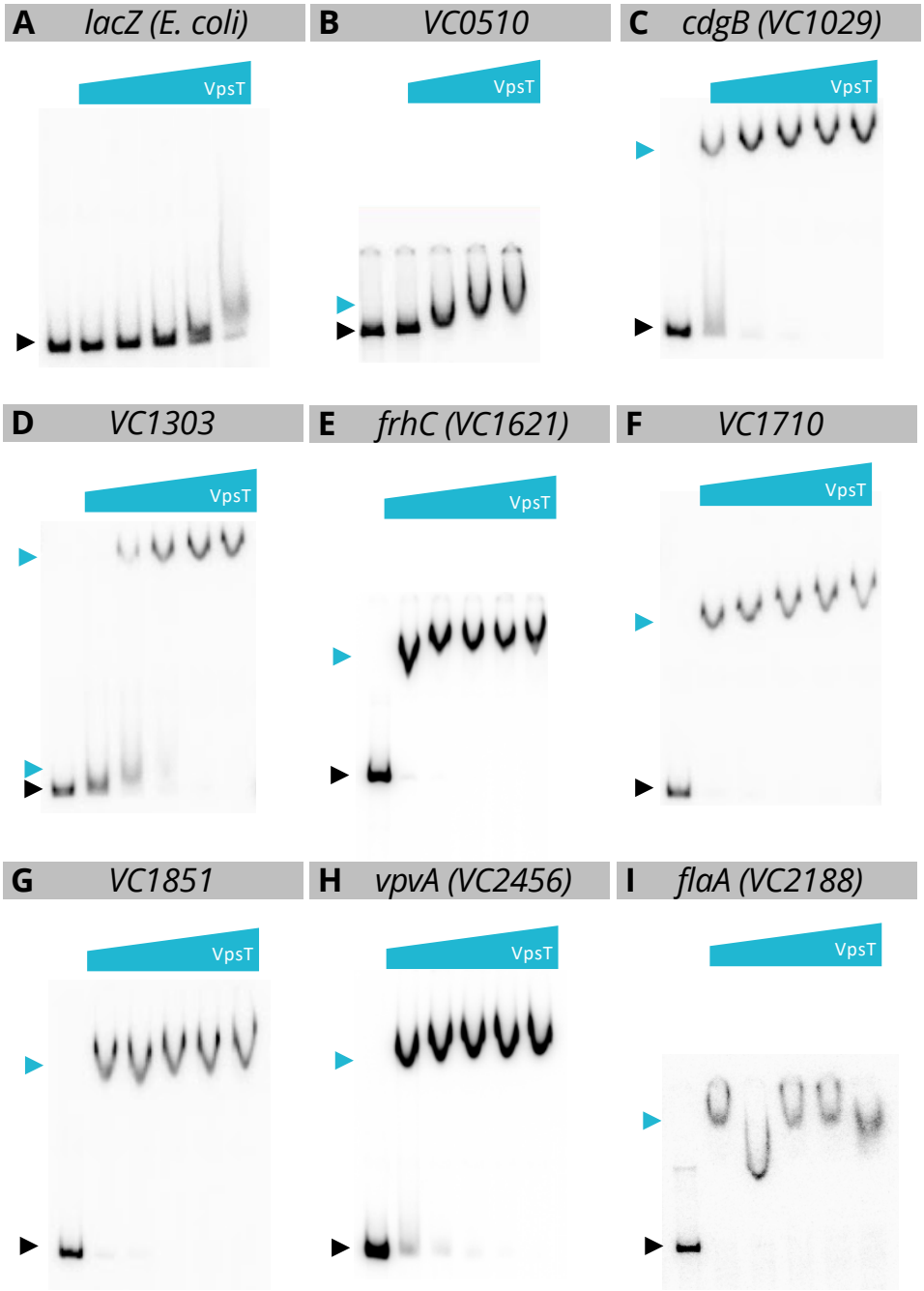

Figure 3.8: Electrophoretic mobility shift assays with VpsT and DNA from the ChIP-seq targets. DNA from the ChIP-seq peaks was amplified by PCR and then used for EMSA with a gradient of purified VpsT (0 μM - 7.5 μM) plus 50 μM c-di-GMP. Unbound DNA is indicated by a black-filled triangle and VpsT-DNA complexes with a blue-filled triangle. A) Promoter DNA from the *E. coli lac* operon was used as a negative control.

3.5 Regulatory effects of VpsT at DNA targets identified by ChIP-seq

To test the role of VpsT binding to these promoter regions each was cloned in plasmid pRW50T, upstream of a copy of the *lacZ* gene lacking its native promoter (Manneh-Roussel et al., 2018). The pRW50T derivatives were introduced into the wildtype E7946 and *vpsT* mutant by conjugation. β-galactosidase assays were done to infer promoter activity using lysates of cultures grown to mid-log phase at 37 °C in LB media. No apparent differences in promoter activity for any of the constructs was observed in the wildtype or *ΔvpsT* mutant in these conditions.

VpsT is active when the intracellular concentration of c-di-GMP is elevated so we added exogenous c-di-GMP to cultures before β-galactosidase assays were done but found this had no effect. We thought this might be because c-di-GMP is not taken up from the media. Therefore, to increase the intracellular concentration of c-di-GMP we introduced the plasmid pBAD33 encoding a diguanylate cyclase and induced expression using arabinose. However, consistent with increased production of c-di-GMP, these strains formed aggregates and it was not possible to record reliable optical density measurements required for the experiment.

Alternatively, the pRW50T derivatives were introduced into wildtype E7946 that constitutively expressed *vpsT* from the pAMNF plasmid. These cells were grown to mid-log phase and the promoter activity assayed, reproducing the setup used for the ChIP-seq experiment. In these conditions VpsT repressed transcription from the regulatory DNA upstream of *VC1303, VC1304, frhA (VC1620), frhC (VC1621)* and *VC1711* (Figure 3.9). Transcription from DNA upstream of *cdgB (VC1029), VC1031* and *vpvA (VC2456)* was activated in cells where

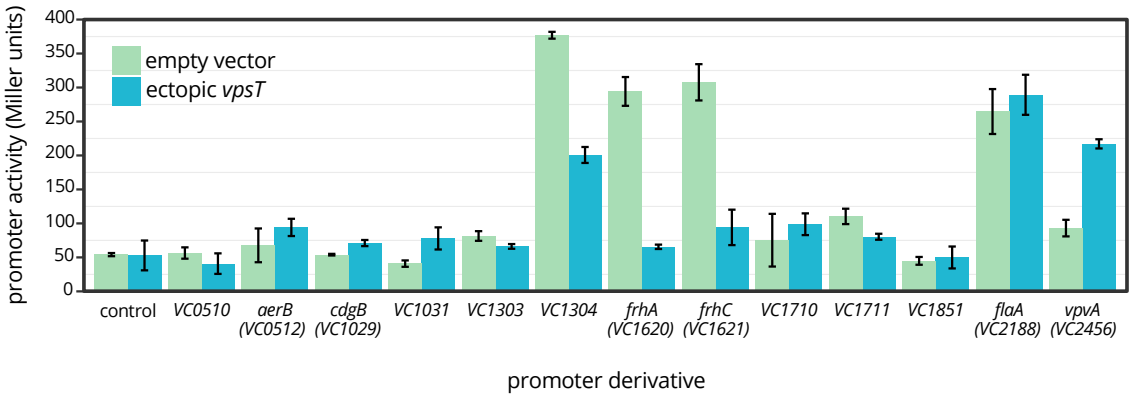

Figure 3.9: Promoter activity in wildtype E7946 with ectopic expression of *vpsT*. Results of β-galactosidase assays done using lysates of E7946 cells transformed with a pRW50T derivative containing promoter DNA from a VpsT target and either an empty pAMNF or one that constitutively expresses *vpsT*. Empty pRW50T was used as a control. Assays were done using lysates of mid-log phase cultures grown in LB media. Data are mean activity in Miller units for 3 biological replicates and error bars represent the standard deviation.

vpsT was ectopically expressed. Strains with an empty pAMNF plasmid were used as controls. The amplitude of the VpsT-induced effect on transcription from these promoters varied and there was little effect at a number of the targets which were similar to the background from the negative control.

The activity of the promoters was also tested in vitro using purified VpsT, *V. cholerae* RNA polymerase and RpoD. For a number of the targets there were no discernible transcripts, however this was typically for promoters activated by VpsT in vivo (data not shown). In the absence of H-NS VpsT is not required to activate *vpsL* gene since it displaces and relieves repression by H-NS (Ayala et al., 2015). VpsT could activate transcription of these targets in a similar way, however, the absence of transcripts in these assays suggests an additional regulatory factor is required for activation.

To see if there was any pattern between the production of transcripts in vitro and H-NS occupancy we compared data from an H-NS ChIP-seq with VpsT ChIP-seq (Figure 3.10). Six of the 23 targets bound by VpsT overlapped with H-NS binding (Table 3.1). For example, the *vpsA* promoter DNA is bound by both VpsT and H-NS (Figure 3.10A). However, H-NS is absent from the promoter region of *rpoS* (Figure 3.10B), a known target for repression by VpsT. We noticed that H-NS is also absent from the promoter DNA upstream of the *vpv* operon (Figure 3.10C). Transcription from this promoter is activated by VpsT which suggests that this is achieved by an alternative mechanism.

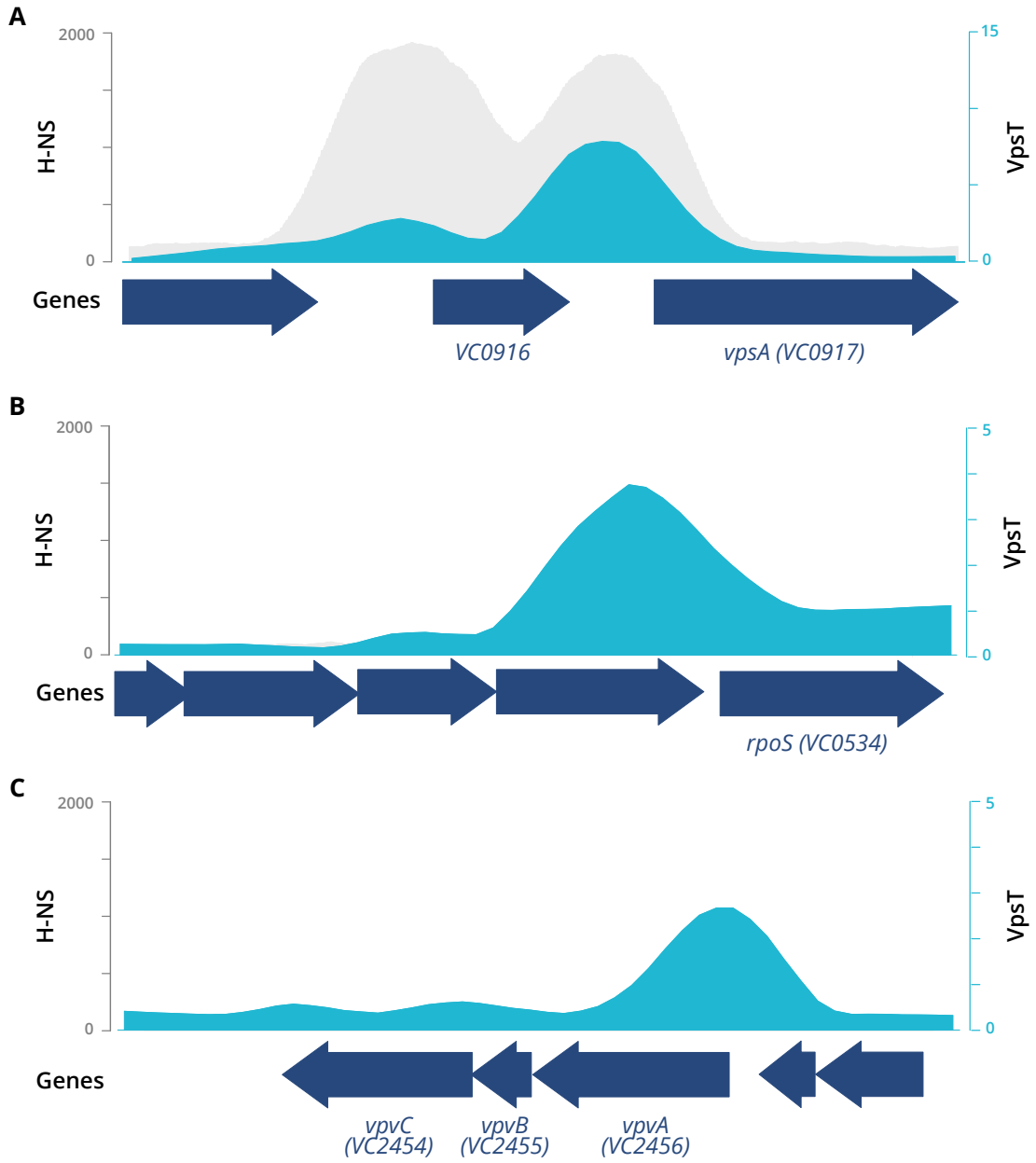

Figure 3.10: ChIP-seq average read profile at VpsT targets for H-NS compared with VpsT. A) The *vps-I* operon showing the average reads from VpsT ChIP-seq (light blue) and the average reads from H-NS ChIP-seq (grey). B) The *rpoS* promoter region, a target that is known to be repressed by VpsT (located within the preceding gene) where there is no H-NS binding. C) The *vpv* operon which is activated by VpsT where there is no H-NS binding.

3.6 Discussion

We identified 23 binding target sites for VpsT using ChIP-seq, the majority of which occurred upstream of genes. Previously, six targets (*vpsA, vpsL, rbmA, rbmE, tag* and *katB*) were known to be directly regulated by VpsT (Zamorano-Sánchez et al., 2015, Fernandez et al., 2018, Fernandez and Waters, 2019). A larger number of potential targets were identified by microarray transcriptomics and we identified six of these as additional direct VpsT targets (Krasteva et al., 2010). We identified a further 11 novel VpsT binding sites. The majority of the target sites we identified are found upstream of genes associated with motility, attachment, formation and maintenance of biofilm. Together, this suggests VpsT has a wider role in the transition from motile cell to biofilm community.

Using the DNA sequence at VpsT peaks we were able to refine the DNA binding motif 'T-box' for VpsT. We found that it frequently appears in the *V. cholerae* genome but unlike the ChIP-seq peaks occurred frequently in intragenic regions of DNA. All the binding targets we identified are on the larger chromosome I despite the fact that *vpsT* itself is on chromosome II. In contrast to earlier reports we found no evidence for VpsT auto-regulation as there was no binding peak upstream of *vpsT* (Zamorano-Sánchez et al., 2015). However, poor expression of *vpsT* in Δ*vpsT* mutants might instead be attributed to a lower c-di-GMP concentration rather than a direct relationship. Another potential target for VpsT on chromosome II is the sRNA VadR. Over-expression of VpsT is known to reduce production of CrvA, a regulator of cell curvature, and lead to growth of straightened rods (Fernandez et al., 2020). Recently, Peschek et al. (2020) identified VadR, an sRNA that directly regulates the abundance of *crvA* mRNA. While our analysis did not identify a peak

upstream of *vadR,* there is some weak enrichment for VpsT at this locus.

The majority of the VpsT target promoters showed little change in gene expression or transcript production in the conditions we tested and we suspect this is because the presence of VpsT is not by itself sufficient to induce regulatory effects at these promoters. This is sometimes observed for binding sites identified using ChIP (Wade et al., 2007). VpsT may function in only specific contexts, under certain conditions, or require the presence of an additional co-operative regulator. Other regulators such as VpsR are known to share VpsT targets and the full extent of this relationship is unknown. For example, at the *vpsA* and *vpsL* promoters VpsT acts by de-repressing gene transcription via H-NS displacement whereas VpsR directly activates transcription (Hsieh et al., 2020). However, activation by VpsT at the *vibrio phase variation* (*vpvA*) promoter, where there is no H-NS co-localisation, suggests the possibility of an alternative mechanism of activation. We noted that H-NS was also absent from the promoter of *tag,* another target gene activated by VpsT.

The maintenance of cellular c-di-GMP concentration is complex, involving a balance between the activity of the c-di-GMP synthesising diguanylate cyclases and degrading phosphodiesterases (Conner et al., 2017). Four of VpsT targets we identified may contribute to c-di-GMP metabolism. The diguanylate cyclase *cdgB (VC1029)* is over-expressed in rugose colonies (Yildiz et al., 2004). Another, *vpvC (VC2454)* is the third gene of the vibrio phase variation operon which by a single SNP can lock cells in the rugose state (Beyhan and Yildiz, 2007). Two potential phosphodiesterase genes: *VC1710* and *VC1851* were also identified.

Motility is inhibited by c-di-GMP but is vital for locating a suitable surface for cell attachment prior to biofilm formation (Utada et al., 2014). We identified the flagellin *flaA (VC2188)*, a class III σ-54 and FlrC dependent

flagellar gene as a target for VpsT binding (Klose and Mekalanos, 1998, Prouty et al., 2008). In addition, *frhA (VC1620)* and *frhC (VC1621)* are also regulated as part of the flagellar regulatory hierarchy (Syed et al., 2009). FrhA mediates adhesion to surfaces such as chitin and erythrocytes while FrhC likely facilitates the activity of FrhA by anchoring it to the cell.

Two VpsT target genes are predicted to encode proteins involved in chemotaxis; *aerB (VC0512)* and *cheY (VC2065)*. The function of *aerB* is poorly understood, it is activated during zinc starvation and its expression is repressed in vivo (Murphy et al., 2021, Cakar et al., 2018). *VC2065* is a *cheY* homologue and has a role in controlling flagellar rotation (Hyakutake et al., 2005). The *VC2065* gene is also required in vivo for the production of cholera toxin (Lee et al., 2001).

The general stress response sigma-factor (RpoS) additionally regulates diverse functions in *V. cholerae* such as natural transformation and the mucosal escape program (Nielsen et al., 2006, Dalia, 2016). There are two growth phase specific promoters for *rpoS,* P1 is located in the intergenic DNA upstream of *rpoS* and P2 is located within the preceding gene, *nlpD* (Yildiz and Schoolnik, 1998). VpsT has been shown to repress *rpoS* gene transcription by occluding the P2 transcription start site (Wang et al., 2014). However, the ChIP-seq binding peak is instead centred within the preceding *nlpD* gene, overlapping the P1 transcription start site. This discrepancy is likely a consequence of the conditions in which the ChIP was done but suggests that VpsT could interact with promoters at both transcription start sites.

Other genes involved in resistance to stress are activated by VpsT including *katB* and the DNA repair gene *tag* (Fernandez et al., 2018, Fernandez and Waters, 2019). Although we did not identify peaks at these genes in our analysis there is some enrichment for VpsT at these

loci. We found another target gene involved in DNA repair (*VC0510*), a predicted RadC-like protein which is encoded in the Vibrio seventh pandemic island-II (VSP-II). Given the diverse functions that are subject to regulation by *rpoS* VpsT could be fine-tuning the stress response to suit the different environmental niches occupied by *V. cholerae*.

Genes involved in virulence with VpsT binding sites included two genes in the toxin co-regulated pilus (TCP) operon (*tcpP* and *tcpS*). To date, VpsT has not been associated with regulating virulence factors. The regulatory effect of this binding is unknown but could contribute to the hyper infectivity of *V. cholerae* biofilms. There is also the accessory colonisation factor C, *acfC (VC0841)*, a chemotactic protein that contributes to intestinal colonisation (Valiente et al., 2018).

Several target genes are poorly understood. For example, there are no conserved protein domains in *VC2667.* The VpsT binding site at this locus is downstream of the annotated *VC2667* start codon but upstream of two intragenic start sites mapped by differential RNA-seq that could be indicatives of the presence of non-coding RNAs (Papenfort et al., 2015).

4 | Molecular characterisation of VpsT binding targets

4.1 Introduction

In the previous chapter we used ChIP-seq to map VpsT binding across the genome and identified 23 direct binding targets. These targets included genes with known roles in biofilm formation, motility and cyclic-di-GMP metabolism. Additionally, many VpsT binding sites were upstream of genes with no known function. We used EMSAs to confirm VpsT bound to a subset of the the novel targets and showed examples of VpsT both repressing and activating transcription of its targets in vivo.

In this chapter we narrowed our focus to four of the VpsT binding targets. These targets were selected because VpsT had clear affects on their transcription and they encompass a broad range of biological roles (see Figure 3.9). The first target selected was between the divergent genes *VC0510* and *aerB (VC0512)*. The second VpsT target was between *VC1303* and *VC1304*. The third target was between *frhA (VC1620)* and *frhC (VC1621)*. The final target is upstream of *vpvA (VC2456)*, the first gene of the *vibrio phase variation (vpv)* operon. We have used a combination molecular methods to examine the interaction of VpsT with the DNA at these targets in detail. DNase I footprinting analysis was done to identify the position of VpsT binding to the DNA at these sites. Primer extension was used to identify transcription start sites and confirm regulatory effects of VpsT. In vitro transcription assays were used to measure direct effects of VpsT on transcription in the absence of other factors. We also mutated promoter DNA sequences to remove and/or disrupt VpsT binding to these altered promoters in vivo.

4.2 VpsT targets in the vibrio seventh pandemic island-II (VSP-II)

The first of the VpsT targets we investigated is located in the Vibrio seventh pandemic island-II (VSP-II). The VSP-II is one of two genomic islands (GI) that are characteristic of the El Tor biotype, responsible for the current seventh cholera pandemic (O'Shea, 2004). The VSP-II includes 24 open-reading frames (*VC0490* - *VC0516*) and is highly conserved in wave 1 and 2 strains, while a shorter variant (without *VC0496*-*VC0498*) is more common in wave 3 strains (Nguyen et al., 2018). Only three genes have characterised functions, the integrase *VC0516,* peptidoglycan endopeptidase *VC0503* and the transcription factor *verA (VC0513)* (Murphy and Boyd, 2008, Murphy et al., 2019, 2021). The remaining genes have a variety of predicted functions: DNA repair (*VC0510*), motility (methyl-accepting chemotaxis proteins: *aerB (VC0512)* and *VC0514*), transcription regulation (*VC0497*) and c-di-GMP metabolism (phosphodiesterase, *VC0515*). The VSP-II genes are not expressed during growth in laboratory conditions, though recently zinc starvation has been shown to play a role in regulating *aerB* transcription via ZurA and VerA (Mandlik et al., 2011, Murphy et al., 2021).

Characterisation of the VpsT binding site at the intergenic DNA between *VC0510* and *aerB (VC0512)*

There are two VpsT ChIP-seq peaks in the VSP-II GI. The first is at the intergenic region between the divergent genes *VC0510* and *aerB (VC0512),* and the second is upstream of *verA* (*VC0513)* (Figure 4.1A). There is also some enrichment of VpsT between the two convergent

genes *VC0515* and *VC0516*. We focussed on the peak between the divergent genes *VC0510* and *aerB (VC0512)*.

The 1.3 kb intergenic DNA was too large for DNase I footprinting analysis so we made a shorter DNA fragment, corresponding to the centre of the VpsT ChIP-seq binding peak, by making use of a naturally occurring restriction site (Figure 4.1B). The DNA fragment was radio-labelled at one [γ^{32}P]-ATP and T4 polynucleotide kinase. The footprinting experiment was done with and without 50 μM c-di-GMP. A Maxam Gilbert GA ladder was used to calibrate the gel, which is shown numbered from the position of the *aerB (VC0512)* start codon. In the absence of c-di-GMP there was no VpsT footprint (Figure 4.2A). In the presence of c-di-GMP there was protection of the DNA between nucleotide position 915 and 990 (Figure 4.2B). VpsT also induces DNase I hypersensitivity at nucleotide positions 940, 950, ~955 and ~960. This indicates alteration of the DNA conformation by VpsT. We identified two sites that matched the VpsT consensus binding motif within the protected DNA (Figure 4.1B, bold). One site is positioned between 944 - 951 nt and another between 968 - 976 nt which is located 8 nt from the centre of the VpsT ChIP-seq peak.

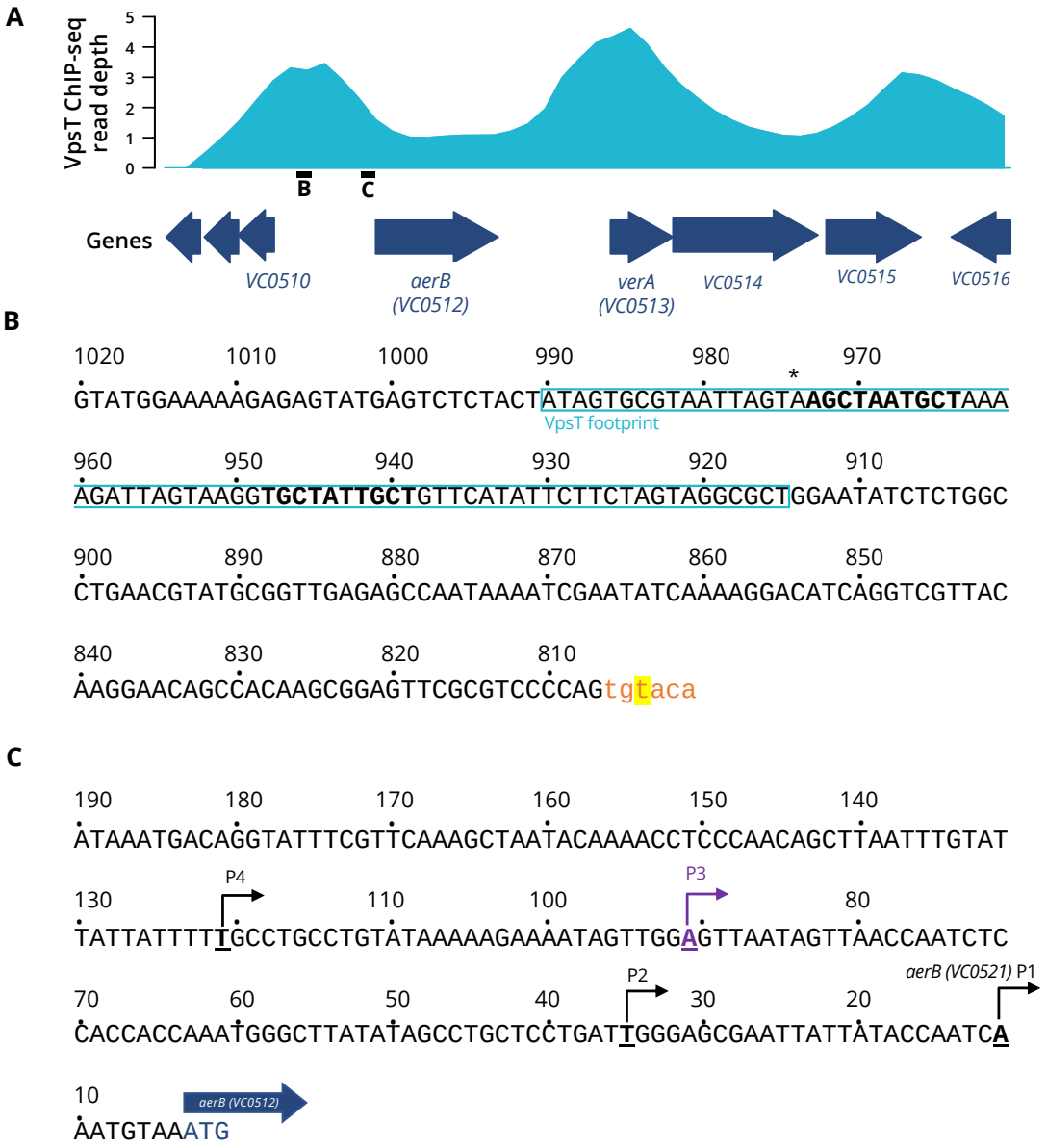

Figure 4.1: VpsT binding at the intergenic DNA between *VC0510* and *aerB (VC0512)*. A) VpsT ChIP-seq average read depth at part of the VSP-II. Genes are shown as dark blue arrows. The black bars show the extent of the DNA sequence shown in panels B and C. The DNA sequences are numbered from the *aerB (VC0512)* start codon. B) The centre of the ChIP-seq peak is marked with an asterisk. The VpsT DNase I footprint is boxed and bold are predicted VpsT binding motifs. C) Transcription start sites indicated with bent arrow.

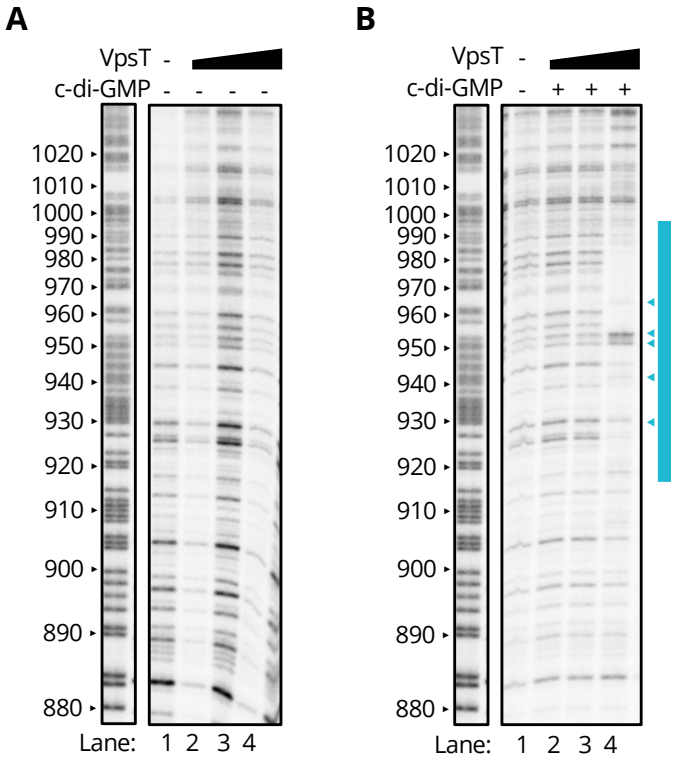

Figure 4.2: DNase I footprint of the promoter DNA upstream of *aerB (VC0512)*. Panels A and B show DNase I footprints done with purified VpsT (0, 2, 4, 6 μM). In B the experiment was done with 50 μM c-di-GMP. In both panels a Maxam-Gilbert G+A ladder shows the distance from the *aerB (VC0512)* start codon. Lane 1 is without VpsT and Lanes 2 - 4 show the experiments done with increasing concentrations of VpsT. The blue bar indicates the VpsT footprint.

VpsT represses transcription from the *aerB (VC0512)* promoter in vitro

To confirm the position of transcription start sites upstream of *VC0510* and *aerB (VC0512)* we used primer extension. We isolated RNA from *V. cholerae* str. E7946 with or without ectopic expression of *vpsT* from a low level constitutive promoter (plasmid pAMNF). The strain also contained a pRW50T derivative, described in Chapter 3, encoding the intergenic DNA in either the *VC0510* or *aerB (VC0512)* orientation, fused to *lacZ*. After RNA isolation from mid-log phase cultures primer extension was done using an oligonucleotide primer that anneals to the *lacZ* sequence encoded by pRW50T.

The results of the primer extension assays are shown in Figure 4.3. In the *VC0510* orientation there was a single 194 nt transcript produced in the absence of VpsT (Lane 5). The production of this transcript is repressed by VpsT (Lane 6). We mapped this start site to a position -108 nt upstream of the *VC0510* start codon. In the *aerB (VC0512)* orientation there were four transcripts produced in the absence of VpsT (Lane 7), these ranged from 97 nt up to 207 nt in length. We named the transcription start sites *aerB P1-4* and were mapped to positions upstream of the aerB (VC0512) start codon (Figure 4.1). No transcripts were produced in the presence of VpsT (Lane 8).

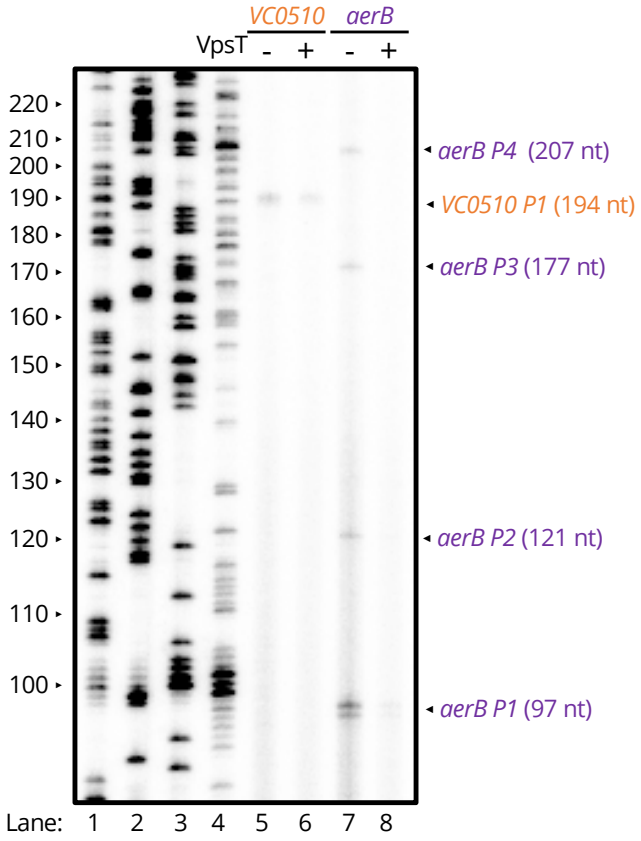

Figure 4.3: Transcription start sites in the intergenic DNA between *VC0510* and *aerB (VC0512).* Primer extension assay done using RNA extracted from cultures grown to mid-log phase in LB using a control strain (empty pAMNF plasmid) or in cells with ectopic expression of *vpsT*. Transcription start sites are indicated with an arrow. Lanes 1-4 show sanger sequencing reactions used to measure the size of the primer extension products. Lanes 5-8 show primer extension products, from strains with the intergenic DNA in either the *VC0510* and *aerB (VC0512)* orientation in pRW50T, with or without ectopic expression of *vpsT*.

To assess the activity of the *aerB (VC0512)* promoter in vitro the intergenic DNA sequence between *VC0510* and *aerB (VC0512),* in both orientations, was cloned in plasmid pSR upstream of a □*oop* terminator sequence (Figure 4.1C)). The pSR derivatives were used as a template for in vitro transcription. The pSR plasmid encodes a 108 nt transcript, called RNAI, that serves as a useful internal control. We used *V. cholerae* RNAP with and without VpsT (2 μM) plus 50 μM c-di-GMP (Figure 4.4). In the *VC0510* orientation a 197 nt transcript from *VC0510 P1* was produced in the absence of VpsT (Lane 5) and transcription of this product was reduced when VpsT was present (Lane 6). There are additional bands that did not match transcription start sites identified in the primer extension assay. These spurious transcripts are most likely due to the high AT content of the VSP-II and the lack of silencing proteins, such as TsrA which is known to repress transcription of VSP-II genes (Caro et al., 2020). We examined TSSs mapped by RNA-seq to see if there might be additional start sites that were not captured in our data, however the C6706 strain that was used for this does not have the complete VSP-II (Papenfort et al., 2015). In the *aerB (VC0512)* orientation a transcript of ~180 nt matches the length of the product expected from *aerB P3* (Lanes 7 and 8).

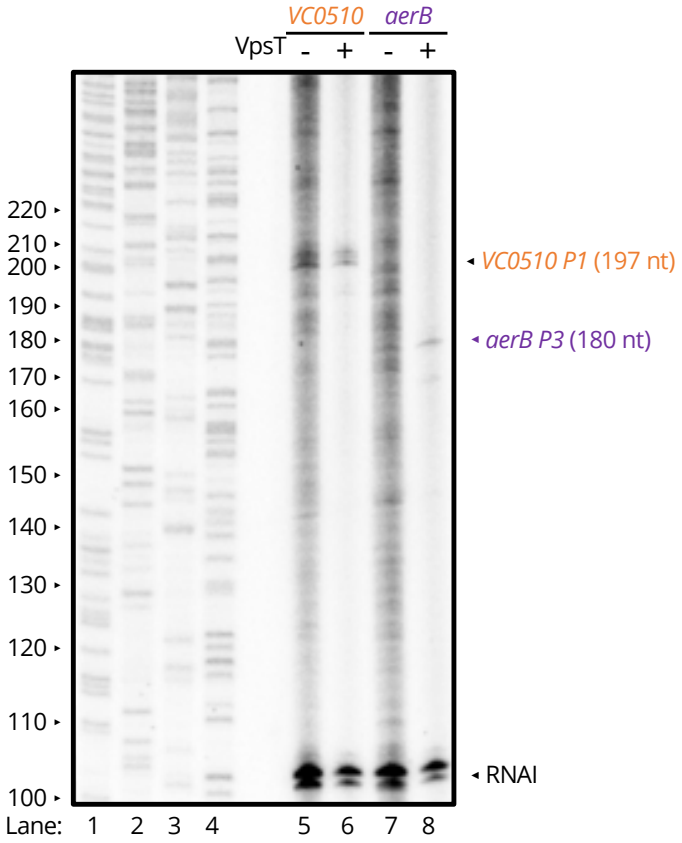

Figure 4.4: Transcription from the *aerB (VC0512)* promoter DNA in vitro. In vitro transcription assays were done using *V. cholerae* RNAP/RpoD holoenzyme with and without purified VpsT (2 μM) plus c-di-GMP (50 μM). The regulatory DNA between *VC0510* (Lanes 5 and 6) and *aerB (VC0512*, Lanes 7 and 8) was cloned in plasmid pSR, in both orientations, and used as a template. Transcripts are marked with arrows. The RNAI transcript is derived from the plasmid and acts as an internal control. Sanger sequencing reactions (Lanes 1-4) were used to measure the size of the transcripts.

4.3 Characterisation of the VpsT binding site at the intergenic DNA between *VC1303* and *VC1304*

The *VC1303* and *VC1304* genes are poorly characterised but are predicted to encode proteins that could be involved in carbon metabolism. *VC1303* encodes a predicted para-aminobenzoate synthase which synthesises para-aminobenzoic acid (PABA), a folate intermediate from chorismate. *VC1304* is predicted to encode a class I fumarate hydratase, that can participate in the citric acid cycle. Expression of *VC1304* is repressed by the iron-regulated small RNA RyhB in the *V. cholerae* classical biotype (Davis et al., 2005).

VpsT binds to the intergenic region between *VC1303* and *VC1304* (Figure 4.5A). In the previous chapter we showed that VpsT repressed transcription from the intergenic DNA in both orientations (Figure 3.9). To further characterise VpsT binding between *VC1303* and *VC1304* we used DNase I footprints to precisely map the VpsT binding site. There was no VpsT footprint in the absence of c-di-GMP (Figure 4.5C). However, a VpsT footprint was detected in the presence of c-di-GMP; protection of the DNA was evident between nucleotide positions 90 and 120 (Figure 4.5D). The position of the VpsT footprint exactly overlaps with the centre of the ChIP-seq peak (Figure 4.5B).

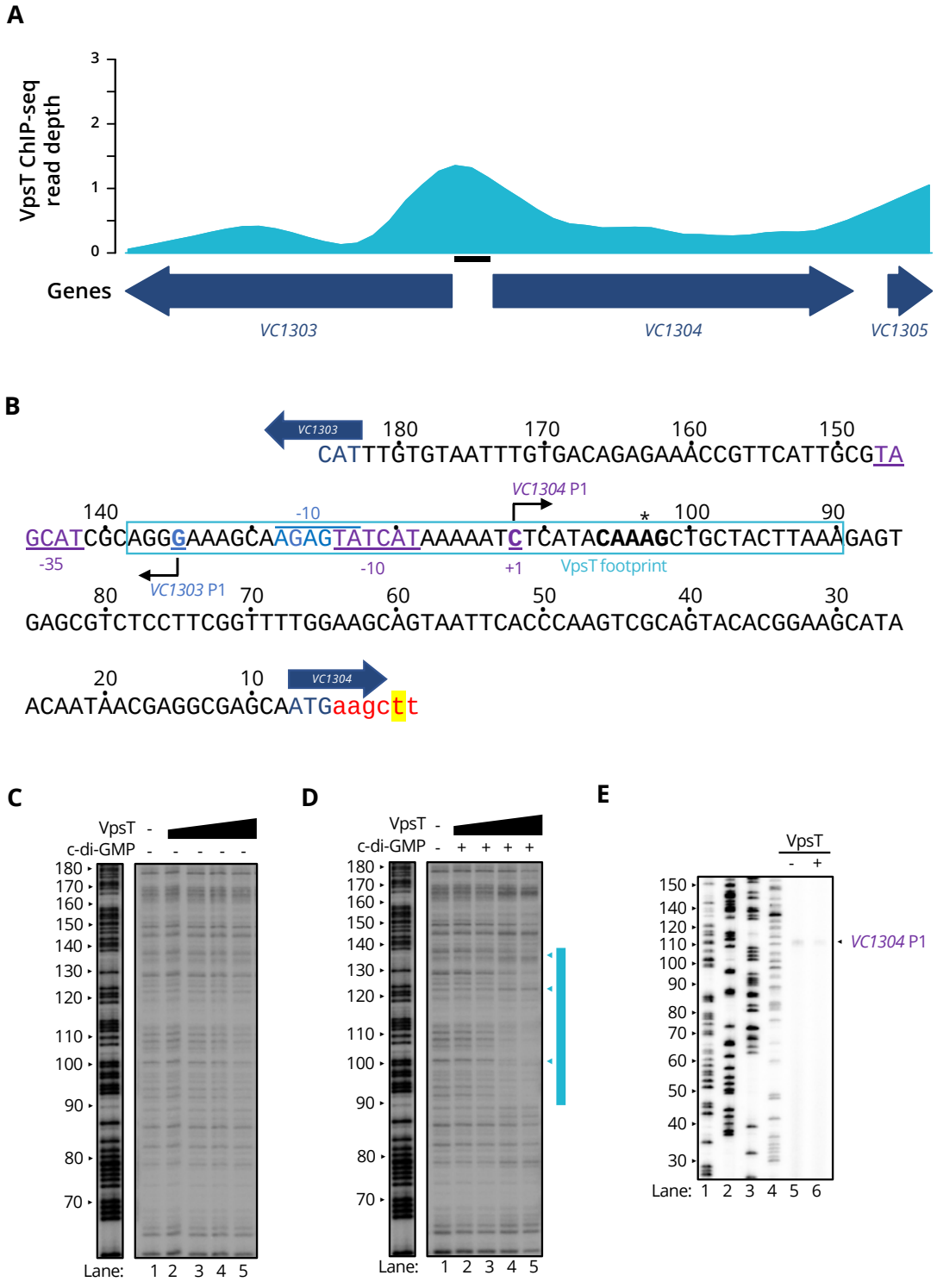

Figure 4.5: Molecular characterisation of VpsT binding at the intergenic DNA between *VC1303* and *VC1304*. A) VpsT ChIP-seq average read depth at the intergenic region between *VC1303* and *VC1304*. Genes are shown as dark blue arrows. The black bar shows the position of the DNA sequence in panel B. *(continued...)*

Figure 4.5 *(continued)*: B) The centre of the ChIP-seq peak is marked with an asterisk. The VpsT binding site is boxed and in bold is the predicted VpsT binding motif. Transcription start sites are indicated by a bent arrow and the -10 and -35 elements are coloured and underlined. The DNA sequence is numbered from the radio-labelled nucleotide used for DNase I footprinting. C and D) DNase I footprints done using purified VpsT (0, 2, 4, 6 μM). In D the experiment was done with 50 μM c-di-GMP. A Maxam-Gilbert G+A ladder shows the distance from the radio-labelled nucleotide. Lane 1 is without VpsT and Lanes 2-5 show the experiments done with increasing concentrations of VpsT. The blue bar indicates the VpsT footprint. Arrowheads indicate VpsT-induced DNase I hypersensitivity. E) Primer extension assays done using RNA extracted from cultures grown to mid-log phase in LB using a control strain (empty pAMNF plasmid) or in cells with ectopic expression of *vpsT*. The *VC1304* transcription start site is indicated with an arrow. Lanes 1-4 show Sanger sequencing reactions, Lane 5 primer extension products from pAMNF vector strain, and Lane 6 from a strain with ectopic expression of *vpsT*.

Identification of transcription start sites upstream of *VC1304*

We searched data from differential RNA-seq (dRNA-seq) to find transcription starts sites in the intergenic DNA between *VC1303* and *VC1304*. There was a single transcription start site upstream of *VC1303* and two transcription start sites reported upstream of *VC1304* (Papenfort et al., 2015). A primer extension assay was done to confirm the position of start sites upstream of *VC1304*, as described earlier. M13 phage DNA sequencing reactions were used to calibrate the gel. A single 112 nt transcript was produced (Figure 4.5E). Consistent with promoter activity measurements in Chapter 3, transcript production was repressed by VpsT (Figure 3.9). We did not observe a transcript from the second transcription start site that was reported in the dRNA-seq dataset. The *VC1304* P1 transcription start site is 8 nt from the centre of the ChIP-seq peak and overlaps with the VpsT DNase I footprint (Figure 4.5B). We were also able to identify -10 and -35 hexamers upstream of the start site (Figure 4.5B).

VpsT regulates transcription from a bidirectional promoter to repress two divergent genes, *VC1303* and *VC1304*

The arrangement of the two divergent transcription start sites, 23 nt apart, means that part of the -10 promoter element is shared. Promoters with shared promoter elements, between two divergent genes, are bidirectional promoters (Warman et al., 2021). To investigate the regulatory role of VpsT at this promoter we cloned the intergenic DNA between *VC1303* and *VC1304*, in plasmid pSR, in both orientations.

The production of transcripts was tested using the in vitro transcription system described earlier. In the absence of VpsT, there was a single 140 nt transcript in the *VC1303* orientation and a 195 nt transcript in the opposite orientation (Figure 4.6). Consistent with the position of VpsT DNase I footprint overlapping the shared -10 promoter element, VpsT repressed transcription in both orientations.

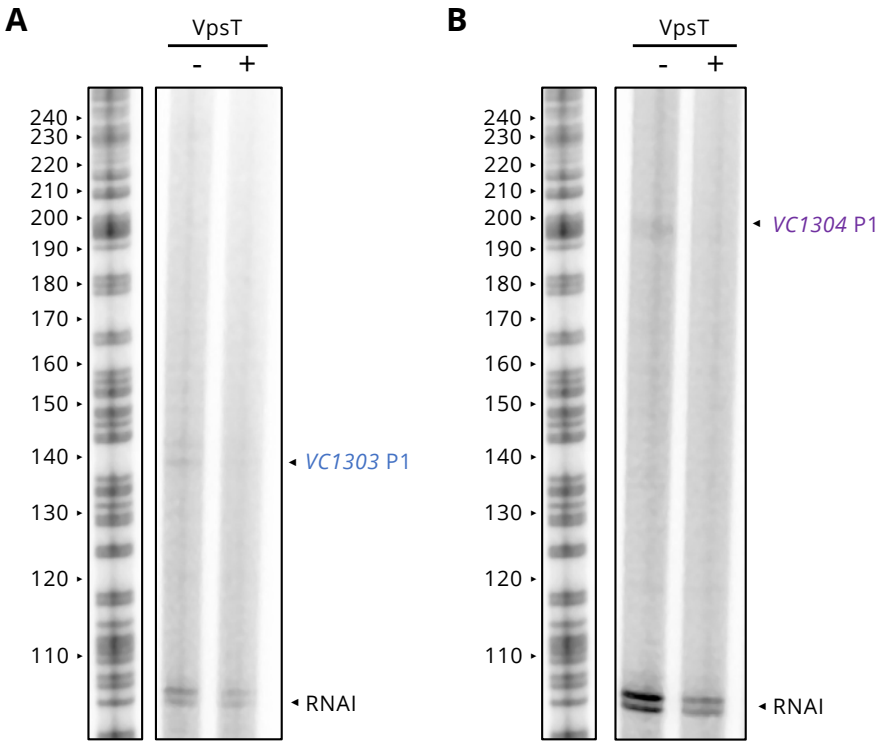

Figure 4.6: Transcription from the *VC1303* and *VC1304* promoter DNA is repressed by VpsT in vitro. In vitro transcription assays were done using *V. cholerae* RNAP/RpoD holoenzyme with and without purified VpsT (2 μM) plus c-di-GMP (50 μM). The DNA was cloned in plasmid pSR, in both orientations, and used as a template. A) *VC1303*. B) *VC1304*. Transcripts are marked with arrows. The RNAI transcript is derived from the plasmid and acts as an internal control. A Maxam Gilbert G+A ladder is arbitrarily labelled according to the ladder DNA sequence and was used to measure the length of the transcripts.

4.4 The *flagellum-regulated haemagglutinin A (frhA VC1620)*

The flagellum-regulated haemagglutinin A (*frhA*, VC1620) and the divergent gene *frhC (VC1621)* are regulated by the Class I flagellar transcription factor FlrA (Syed et al., 2009). FrhA is a large 2,334 kDa protein with RTX-like repeats and a Type-1 secretion system motif at the C terminus. The divergent gene *frhC VC1621* is predicted to encode a LapE-like Type I secretion protein (Kitts et al., 2019). The following gene in the operon *VC1622* shares homology with the outer membrane porin OmpA.

FrhA and FrhC contribute to haemagglutination, biofilm formation and host colonisation by mediating adherence to epithelial cells and chitin (Syed et al., 2009). These traits are less pronounced in *V. cholerae* O1 El Tor strains due to the presence of more dominant agglutination proteins, such as the MSHA pilus (Kitts et al., 2019). The reference *V. cholerae* El Tor str.N16961 has an intragenic stop codon in *frhA (VC1620)* (Chatterjee et al., 2008).

Transcription of *frhA* is activated indirectly by the DGC CdgD due to elevated levels of c-di-GMP (Syed et al., 2009). VpsT is active at elevated c-di-GMP concentrations but a recent study found that VpsT repressed *frhA (VC1620)* transcription (Kitts et al., 2019). Intriguingly, this regulation only occurred during exponential growth. We found that the intergenic DNA between *frhA (VC1620)* and *frhC (VC1621)* is a direct binding target for VpsT (Figure 4.7A). We did β-galactosidase assays using lysates from mid-exponential cultures and confirmed that VpsT repressed transcription in both orientations (Figure 3.9). A previous study from our group identified a binding site for the cyclic-AMP receptor

protein (CRP) (Manneh-Roussel et al., 2018).

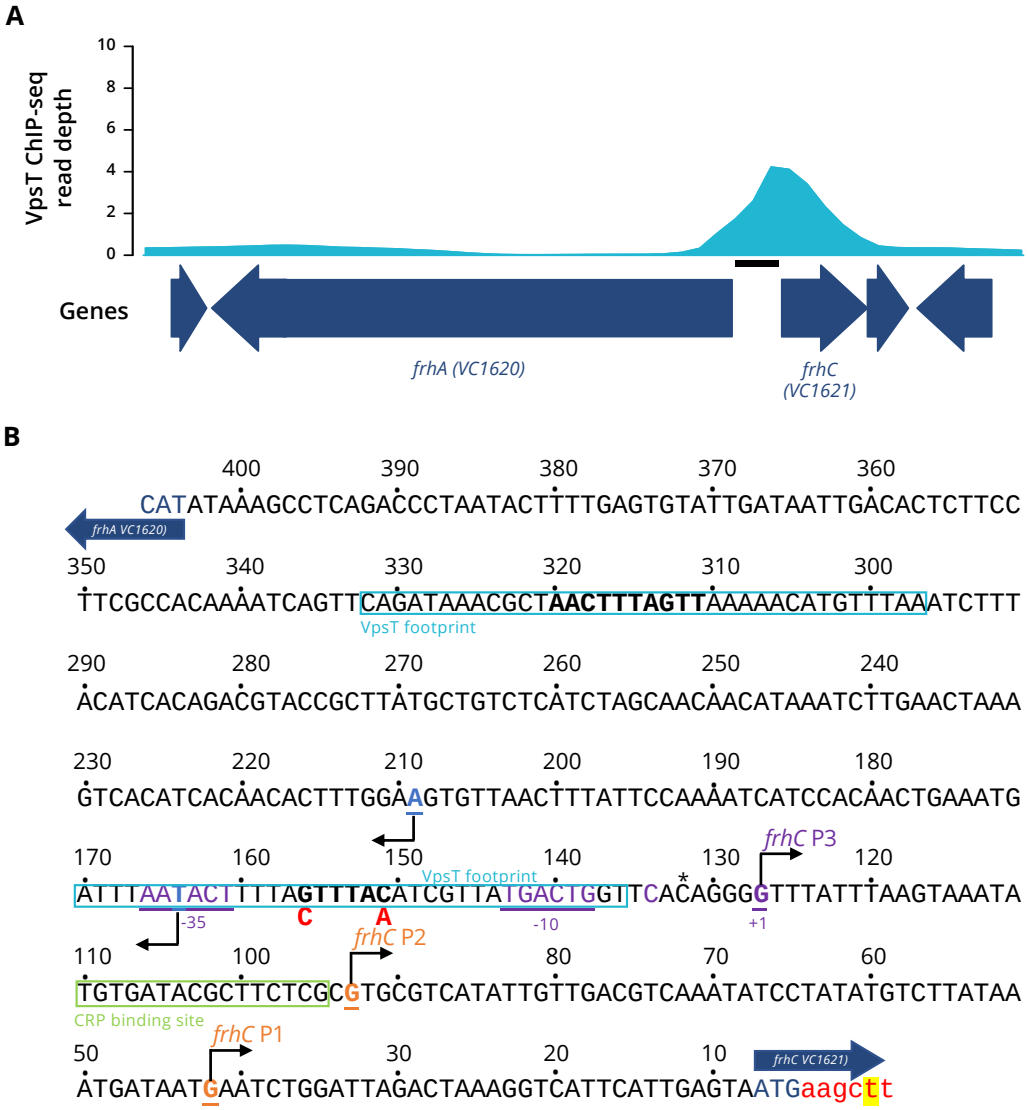

Figure 4.7: VpsT binding at the integernic DNA between *frhA (VC1620)* and *frhC (VC1621)*. A) VpsT ChIP-seq read depth between the divergent genes *frhA (VC1620)* and *frhC (VC1621)*. Genes are shown as dark blue arrows. The black bar shows the position of the DNA sequence in panel B. B) The centre of the ChIP-seq peak is marked with an asterisk. The VpsT binding sites are boxed and in bold are the predicted VpsT binding motifs. Point mutations made to the VpsT binding motif are shown in red. Mapped transcription start sites are indicated by bent arrows and the predicted -10 and -35 promoter elements are underlined. The DNA sequence is numbered from the radio-labelled nucelotide used for DNase I footprinting.

Characterisation of the VpsT binding site at the intergenic DNA between *frhA (VC1620)* and *frhC (VC1621)*

To further characterise the binding of VpsT to this region, DNase I footprints were done. To allow visualisation of the entire *frhA (VC1620)-frhC (VC1621)* intergenic region the template DNA was radio-labelled at one end or the other. For simplicity, the GA ladders are labelled according to the position upstream of the *frhC (VC1621)* start codon, regardless of the DNA fragment orientation (Figure 4.7B). There were no VpsT footprints without the addition of c-di-GMP (Figure 4.8A and C). When c-di-GMP was added there were two VpsT footprints (Figure 4.8B). There was protection of the DNA between between nucleotide position 135 - 180, and between positions 305 - 315. In the reverse orientation only the footprint between nucleotides 135 and 180 was resolved (Figure 4.8D). There are two hypersensitive bands at 145 nt (Lanes 3-5) and 160 nt (Lane 3). The band at 160 nt was not present at higher concentrations of VpsT.

Identification of transcription start sites upstream of *frhC (VC1621)*

Primer extension was done as described earlier using RNA from cells with the *frhC (VC1621)* promoter DNA fused to *lacZ*, in plasmid pRW50T. Three transcription start sites (TSS) were identified upstream of *frhC (VC1621)* (Figure 4.9A). The *frhC* P1 and P2 TSSs match those reported using dRNA-seq (Papenfort et al., 2015). TSSs in the *frhA (VC1620)* orientation reported in the dRNA-seq study are marked in blue (Figure 4.9B). The VpsT binding site is in close proximity to *frhC (VC1621)* P3

and on the reverse strand a transcription start site for *frhA (VC1620)*.

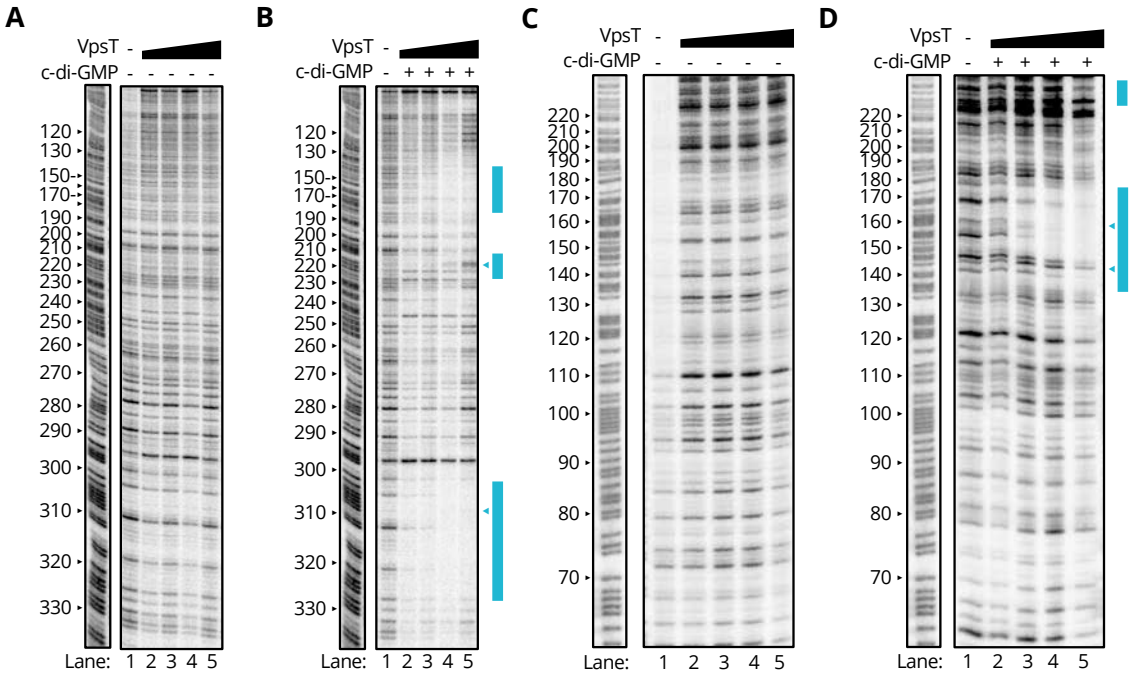

Figure 4.8: DNase I footprints for the promoter DNA of *frhA (VC1620)* and *frhC (VC1621)*. DNase I footprints done with purified VpsT (0, 2, 4, 6 μM). In B and D the experiments were done with 50 μM c-di-GMP. In all panels a Maxam-Gilbert G+A ladder shows the distance from the radio-labelled nucleotide in the *frhC (VC1621)* orientation. Panels A and B show DNase footprints with the DNA in the *frhA (VC1620)* orientation. In C and D the orientation of the DNA sequence is reversed (*frhC (VC1621)*). In all panels Lane 1 is without VpsT and Lanes 2 - 5 show the experiments done with increasing concentrations of VpsT. The blue bar indicates a VpsT footprint and arrowheads indicate VpsT-induced DNase I hypersensitivity.

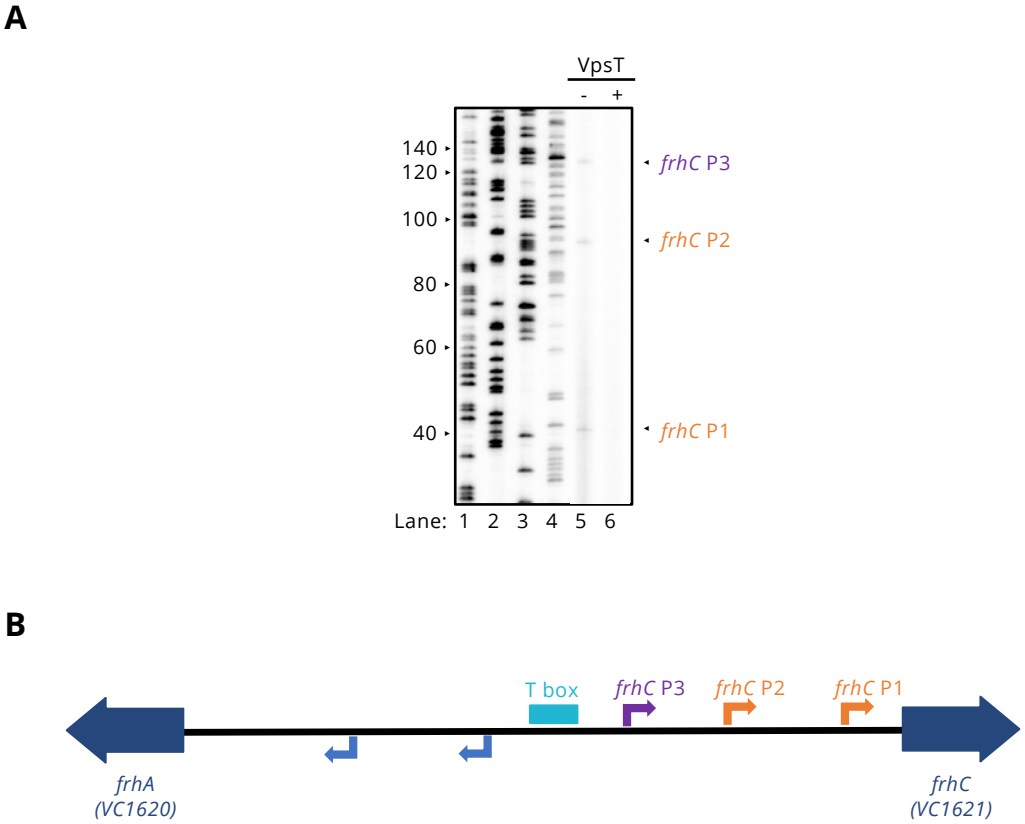

Figure 4.9: Transcription start sites from promoter DNA upstream of *frhC(VC1621)*. A) Primer extension assays done using RNA extracted from cultures grown to mid-log phase in LB using a control strain (empty pAMNF plasmid) or in cells with ectopic expression of *vpsT*. Transcription start sites are indicated with an arrow. Lanes 1-4 show Sanger sequencing reactions, Lane 5 primer extension products from pAMNF vector strain, and Lane 6 from a strain with ectopic expression of *vpsT*. B) Schematic of the integernic promoter DNA between *frhA(VC1620)* and *frhC(VC1621)* labelled with transcription start sites and VpsT footprint (not to scale). Transcription start sites in blue are from differential RNA-seq (Papenfort et al., 2015).

Transcription from the *frhA (VC1620)* and *frhC (VC1621)* promoters is repressed by VpsT in vivo

To confirm VpsT binding to the site we identified upstream of *frhC (VC1621)* P3 we made a series of pRW50T constructs for β-galactosidase assays (Figure 4.10A). As before, the assays were done in cells which either constitutively expressed *vpsT* or carried the empty plasmid (pAMNF). Two point mutations to nucleotides at position 151 and 156 were made to disrupt the predicted VpsT T-box (marked in red, Figure 4.7A). Transcription from the wildtype *frhA (VC1620)* promoter was repressed by VpsT (Figure 4.10B). When the T-box was mutated the basal expression from this promoter was greatly reduced (presumably due to inadvertent mutation of a sequence required for maximal promoter activity) and there was no further repressive effect of VpsT. Transcription in the reverse orientation, upstream of *frhC (VC1621)*, was also repressed by VpsT (Figure 4.10B). Mutating the T-box had no effect on basal promoter activity and VpsT-mediated repression was completely lost. This showed that this T-box is required only for repression of *frhC (VC1621)*. VpsT-mediated repression of transcription in the *frhA (VC1620)* orientation must therefore involve an alternative VpsT binding site.

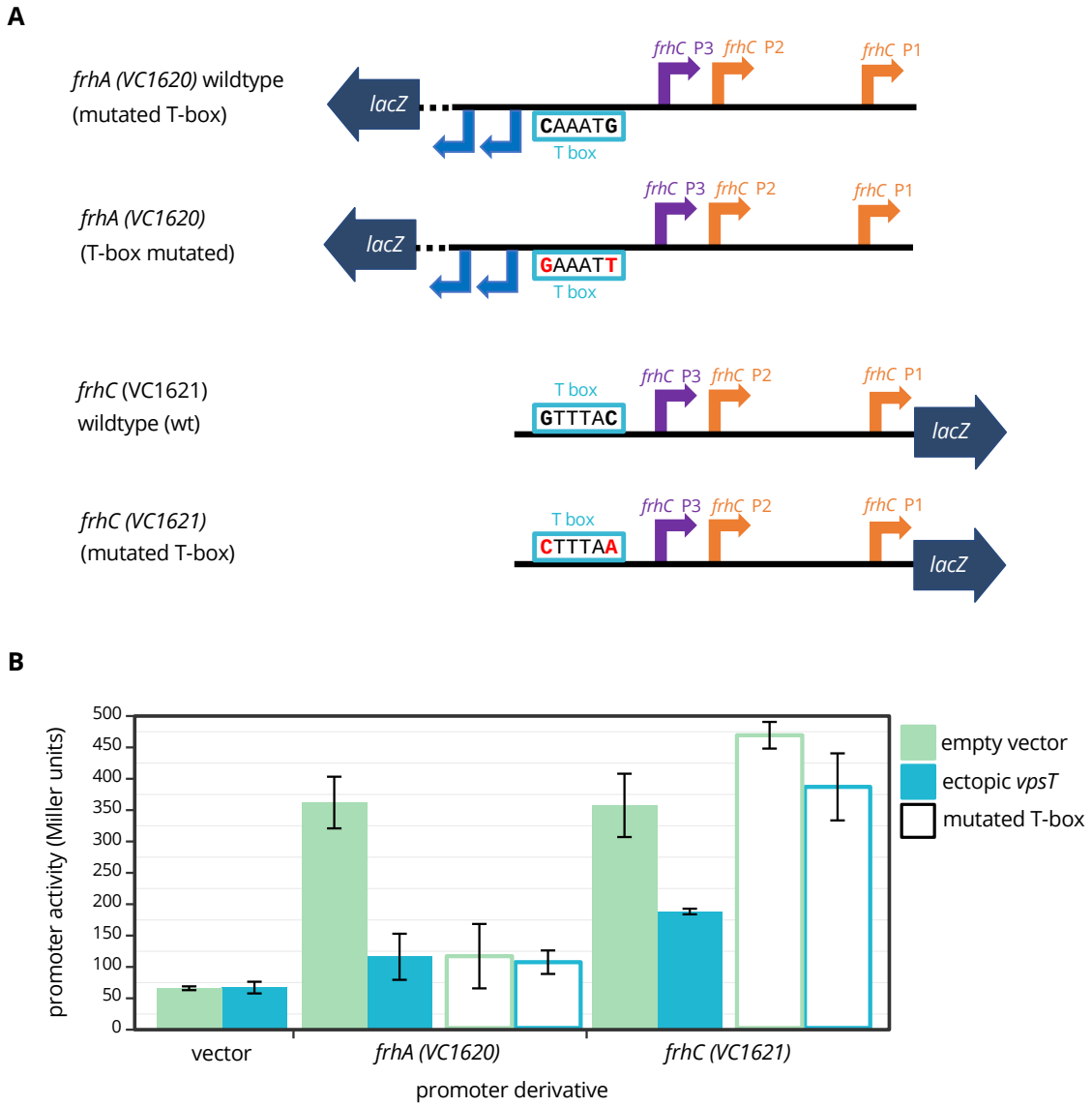

Figure 4.10: Point mutations to the *frhC (VC1621)* promoter affect VpsT regulatory activity A) Schematics show how the promoter region upstream of *frhA (VC1620)* and *frhC (VC1621)* was cloned upstream of a promoterless *lacZ* in pRW50T. Mutations to the sequence were made to mutate key residues in the VpsT binding site (T-box). B) The promoter activity (Miller units) was assayed after growth to mig-log phase in LB using a control strain (empty pAMNF plasmid, green bars) or in cells that constitutively express *vpsT* ectopically (light-blue bars). Filled bars have wildtype T-box, unfilled bars are mutated. The empty pRW50T vector was used as a negative control. Data are means from three biological replicates and error bars represent the standard deviation.

4.5 The vibrio phase variation (*vpv*) operon

Phase variation is a reversible process in which cells switch between different phase variants to help overcome environmental stresses (Ali et al., 2002). In *V. cholerae*, formation of rugose variants is a result of increased production of the vibrio polysaccharide (Yildiz and Schoolnik, 1999). Increased intracellular c-di-GMP modulates this increase via the biofilm master regulators VpsR and VpsT (Yildiz et al., 2004, Lim et al., 2006). The vibrio phase variation operon, *vpvABC,* is encoded on the larger chromosome I (Beyhan and Yildiz, 2007). The functions of VpvA and VpvB are unknown but VpvC is a GGDEF domain-containing protein, typical of diguanylate cyclases (DGC) involved in the production of c-di-GMP.

Characterisation of VpsT binding at the *vibrio phase variation (vpv)* operon

ChIP-seq data showed VpsT bound upstream of *vpvA (VC2456)*, the first gene of the *vpv* operon (Figure 4.11A). In the previous chapter we showed using β-galactosidase assays, using the *vpvA* promoter DNA, that VpsT activates transcription at this locus (Figure 3.9). We also noted that this regulatory DNA is not bound by H-NS. Hence, activation by VpsT is not due to H-NS displacement (Figure 3.10C).

To identify the VpsT binding site at the promoter DNA upstream of *vpvA (VC2456)* DNase I footprinting was done. There was no VpsT footprint without the addition of c-di-GMP (Figure 4.11C). There were two VpsT footprints when c-di-GMP was added (Figure 4.11D). There was protection of DNA between nucleotide position 110 and 127 and a

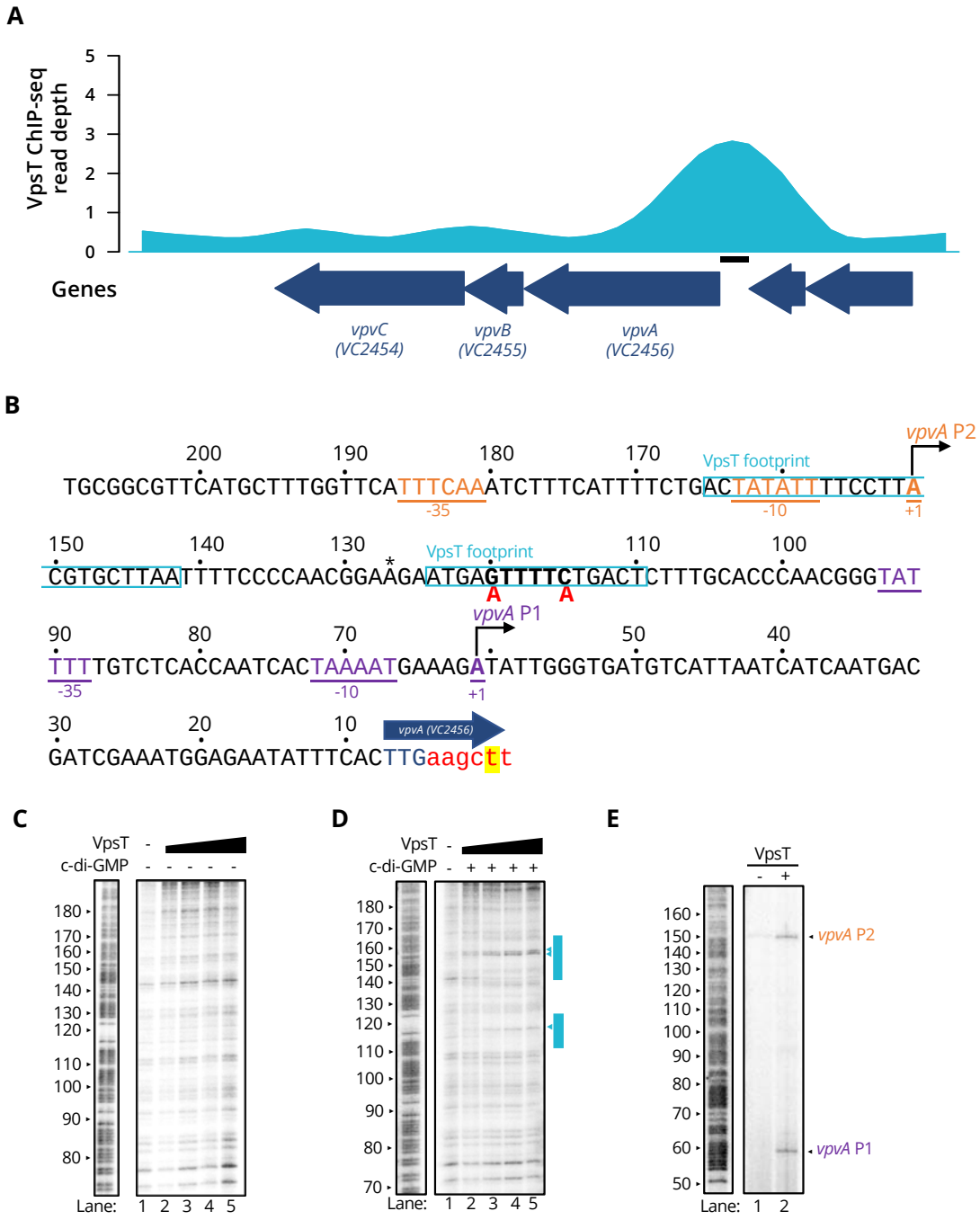

Figure 4.11: Molecular characterisation of VpsT binding at the promoter DNA upstream of *vpvA (VC2456)*. A) VpsT ChIP-seq average read depth upstream of *vpvA (VC2456)*. Genes are shown as dark blue arrows. The black bar shows the position of the DNA sequence in panel B. B) The centre of the ChIP-seq peak is marked with an asterisk. The VpsT binding sites are boxed and in bold is the predicted VpsT binding motif. *(continued...)*

Figure 4.11 *(continued)*: Point mutations made to the VpsT binding motif are shown in red. Mapped transcription start sites are indicated by bent arrows and the predicted -10 and -35 promoter elements are underlined. The DNA sequence is numbered from the radio-labelled nucleotide used for DNase I footprinting. C and D) DNase I footprints done using purified VpsT (0, 2, 4, 6 μM). In D the experiment was done with 50 μM c-di-GMP. A Maxam-Gilbert G+A ladder shows the distance from the radio-labelled nucleotide. Lane 1 is without VpsT and Lanes 2-5 show the experiments done with increasing concentrations of VpsT. The blue bar indicates VpsT footprints. E) Primer extension assays done using RNA extracted from cultures grown to mid-log phase in LB using a control strain (empty pAMNF plasmid) or in cells with ectopic expression of *vpsT*. Transcription start sites are indicated with arrows. A Maxam-Gilbert G+A ladder used to calculate the length of the transcripts. Lane 1 shows primer extension products from pAMNF vector strain, Lane 2 from a strain with ectopic expression of *vpsT*.

hypersensitive band at ~ 118 nt. The second VpsT footprint occurred between positions 140 - 170 with two closely positioned hypersensitive bands at ~155 nt.

Identification of the *vpvA* transcription start sites

To identify TSS upstream of *vpvA (VC2456)* primer extension was done (Figure 4.11E). In the absence of VpsT there was a single transcript, we named this distal promoter *vpvA* P2 (Figure 4.11E, Lane 1). Using RNA from bacteria with ectopic expression of *vspT* there was an additional transcript from a promoter we named *vpvA* P1 (Figure 4.11E, Lane 2). We mapped the *vpvA* P1 TSS to nucleotide position 61 and the *vpvA* P2 TSS to position 151 (Figure 4.11B). Both TSSs are activated by VpsT but P2 was active at a low level without VpsT. We were able to identify -10 and -35 elements upstream of each start site. The VpsT binding site (T-box) is located 20 nt upstream of the -35 hexamer of *vpvA* P1 (Figure 4.11B).

Transcription of the *vpv* operon is activated by VpsT

To investigate the relationship between VpsT and the two TSSs identified upstream of *vpvA (VC2456)* we prepared pRW50T constructs with mutated versions of the promoter DNA for β-galactosidase assays (Figure 4.12A). As before, assays were done using *V. cholerae* str. E7946 which either had low level constitutive expression of *vpsT* or carried empty plasmid (pAMNF) as a control. Transcription from the wildtype promoter sequence was activated by VpsT (Figure 4.12B). When *vpvA* P2 was deleted transcription was still activated by VpsT but overall activity was reduced. Activation by VpsT was abolished when the T- box was mutated

or deleted. Collectively these results indicate that VpsT regulation occurs via binding to the T-box upstream of *vpvA* P1.

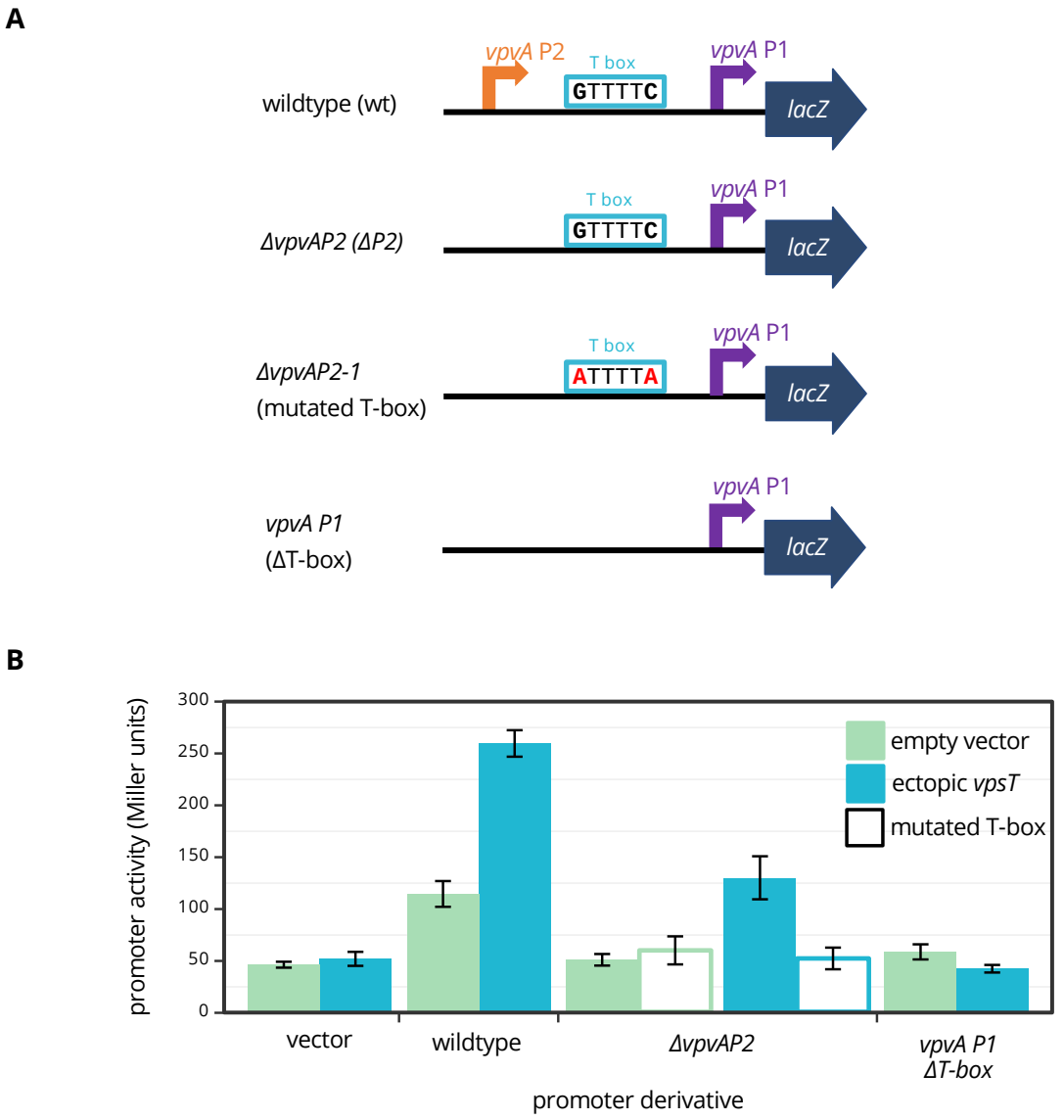

Figure 4.12: Point mutations to the *vpvA (VC2456)* promoter affect VpsT regulatory activity. A) Schematics show how the promoter region upstream of *vpvA (VC2456)* was cloned upstream of a promoterless *lacZ* in pRW50T. Mutations to the sequence were made to delete *vpvA P2* and/or mutate key residues in the VpsT T-box. B) The promoter activity (Miller units) was assayed after growth to mig-log phase in LB using a control strain (empty pAMNF plasmid, green bars) or in cells that constitutively express an ectopic copy of *vpsT* (light-blue bars). Filled bars have wildtype T-box, unfilled bars are mutated. The empty pRW50T vector was used as a negative control. Data are means from three biological replicates and error bars represent the standard deviation.

Mechanism for VpsT activation of the *vpv* operon

VpsT activates transcription at the *vps* promoter by displacing H-NS, relieving H-NS repression of this operon (Ayala et al., 2015). In contrast, the *vpvA* promoter is not bound by H-NS and so VpsT activation is independent of H-NS. NarL, a closely related LuxR type transcription factor in *E. coli* is also able to activate transcription using different mechanisms (Squire et al., 2009). The first is via displacement of the NAP Fis. Alternatively, NarL can activate transcription by directly interacting with the RNAP holoenzyme via two surface residues (R178 and R179) (Ruanto et al., 2020). We used Clustal and ESPript3 to align the NarL and VpsT proteins (Robert and Gouet, 2014). The D181 and K182 amino acids of VpsT align to the same positions as R178 and R179 of NarL (Figure 4.13A). In order to test if VpsT might also interact with the RNAP holoenzyme via residues at these positions we introduced two alanine substitutions to VpsT, D181A and K182A. The mutated versions of VpsT, like the wildtype, were supplied by ectopic expression from plasmid pAMNF. We used these plasmid constructs to measure β-galactosidase activity from the *vpv* promoter fused upstream of the *lacZ* gene in pRW50T (Figure 4.13B). We also included the *lacZ* promoter-fusion constructs for the *frhC (VC1621)* promoter region as a negative control, to assess if VpsT was still able to repress transcription. The wildtype VpsT activates transcription from the *vpvA* promoter. The D181A and K182A mutations did not affect the ability of VpsT to activate transcription. VpsT represses transcription from the *frhC (VC1621)* promoter. The D181A mutated VpsT was still able to repress transcription. However, repression of transcription from *frhC (VC1621)* promoter was weakened with the K182A mutated VpsT, that suggests this mutation has affected DNA binding, rather than any interaction with the RNAP holoenzyme.

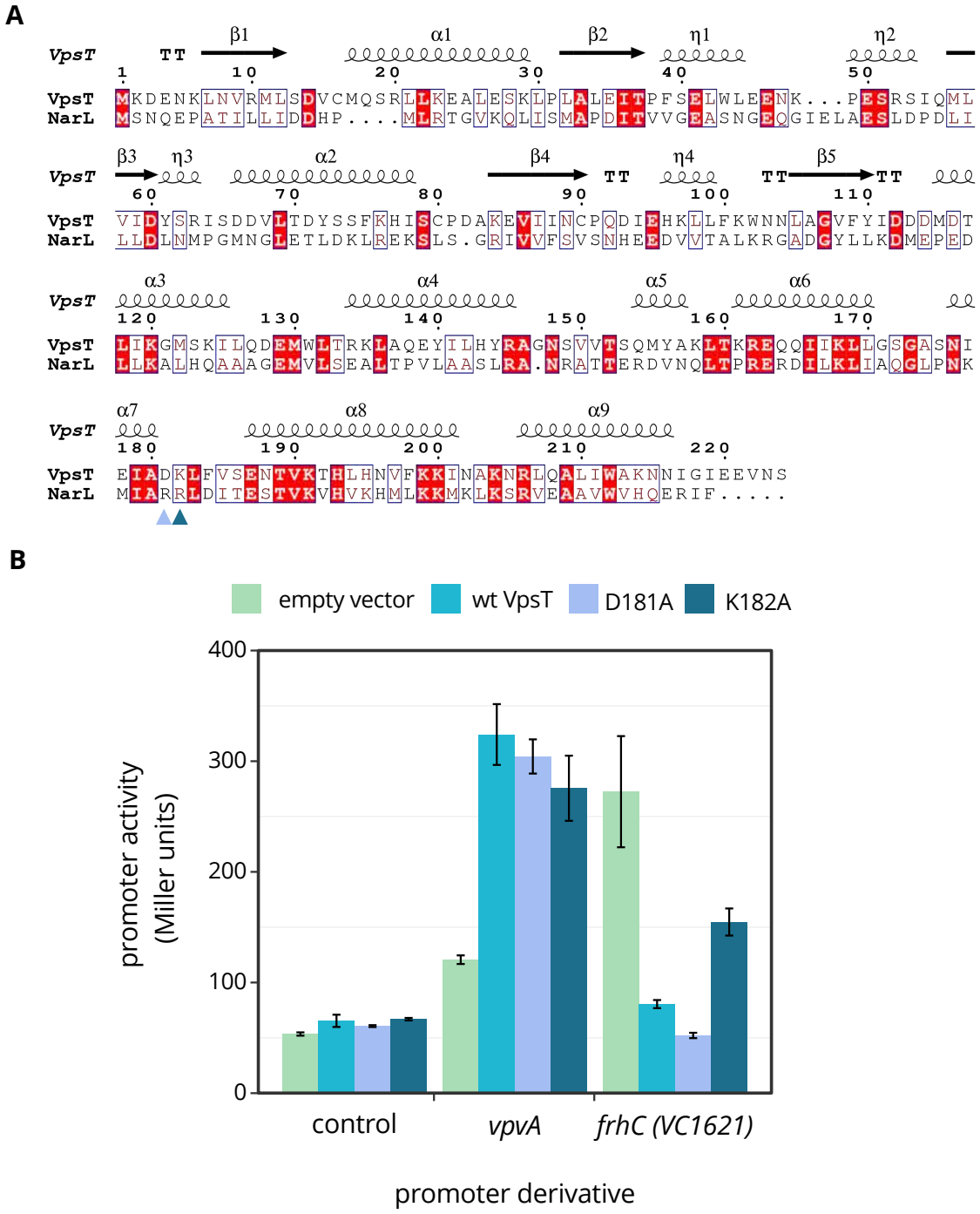

Figure 4.13: VpsT alanine substitutions. A) Protein alignment of VpsT and *E. coli* NarL. Alanine substitutions were made to VpsT at positions D181 and K182. B) The promoter activity (Miller units) was assayed after growth to mig-log phase in LB using a control strain (empty pAMNF plasmid, green bars) or in cells that constitutively express an ectopic copy of wildtype or a mutant *vpsT* (see key). The empty pRW50T vector was used as a negative control. Data are means from three biological replicates and error bars represent the standard deviation.

4.6 Discussion

The VSP-II remains a poorly understood but potentially important characteristic of seventh pandemic El Tor *V. cholerae*. Recent research found the zinc starvation response regulator (Zur) is an activator of *verA* transcription (Murphy et al., 2021). VerA, in turn, activates transcription of *aerB*, that promotes migration away from the liquid-air interface. How VpsT contributes to the regulation of these genes remains unclear. In our experimental conditions, in which transcription of *VC0510* and *aerB (VC0512)* was low, the regulatory role of VpsT is difficult to discern. Promoter activity assays showed little difference, with or without VpsT (Figure 3.9), and this was similar to our results using an in vitro transcription system (Figure 4.4). We identified five transcription start sites (TSS) upstream of *aerB (VC0512)* but we did not observe the *aerB (VC0512)* TSS reported by Murphy et al. (2021), that was >1 kb from the AerB start codon. It is possible that VpsT also regulates *verA* (*VC0513*) since its promoter (located -225 nt upstream of the start codon) is in close proximity to the second VpsT ChIP-seq peak in the VSP-II (see Figure 4.1). Altogether this suggests VpsT could play an important role in regulating transcription of genes in the VSP-II genome island.

The biological roles of *VC1303* and *VC1304* are unknown, though it is tempting to speculate that might be related in some way. There is a single TSS in each orientation, spaced 23 nt apart. This arrangement results in reciprocal -10 promoter element. This is a bidirectional promoter and they occur frequently in prokaryote genomes (Warman et al., 2021). Bidirectional promoters provide the opportunity for a single regulatory event to co-ordinate transcription of two divergent genes. We showed that VpsT repressed transcription of the two divergent

genes, *VC1303* and *VC1304* by binding to a site overlapping the shared bidirectional promoter.

The flagellum-regulated haemagluttinin FrhA (VC1621) is an adhesin that can mediate binding to a range of different surface types (Syed et al., 2009). The divergent gene, *frhC* (*VC1621)* is predicted to encode a type I secretion protein that could function to anchor FrhC into the cell membrane. The VpsT ChIP-seq peak was centred nearer to the *frhC* start codon and we focussed on regulation in this orientation. Using primer extension we identified three TSSs upstream of *frhC* and two VpsT binding sites by DNase I footprinting. Intriguingly, one the VpsT footprints was closely located between *frhC P3* and in the reverse orientation a TSS for *frhA*. However, we found that this site was only involved in regulating transcription of *frhC.* Mutating the T box abolished *frhA* transcription, presumably we had unintentionally disrupted the promoter sequence. A second VpsT footprint is located nearer the *frhA* start codon and may be responsible for VpsT repression of *frhA*. The regulation of *frhA* and *frhC* shows the interconnected, paradoxical nature of c-di-GMP regulatory systems. Other groups have shown that c-di-GMP promotes activation of *frhA* transcription (Syed et al., 2009). More recently, VpsT was identified as a repressor of *frhA* transcription, but only during exponential growth (Kitts et al., 2019). We suggest that the interplay between CRP and VpsT co-ordinate growth-phase specific regulation of these two genes.

Phase variation in *V. cholerae* results in the growth of two variants, which are differentiated by their wrinkled appearance and increased c-di-GMP production (Lim et al., 2006). We showed there are two TSSs upstream of *vpvA,* the first gene of the *vibrio phase variation* operon. VpsT activation occurs predominantly via a T box located 70 nt upstream of *vpv P1,* that is inactive in the absence of VpsT. A second VpsT binding

site, as shown by DNase I footprinting is located at a site overlapping the -10 element of *vpvA P2.* Promoter 2 is active in the absence of VpsT, but transcription activity does increase in the presence of VpsT.

5 | VpsT controls a positive feedback loop to maintain cyclic-di-GMP levels

5.1 Introduction

In the previous chapter we showed that VpsT activates transcription of the vibrio phase variation (*vpv*) operon. As mentioned previously, phase variation results in two variant types that differ by their levels of c-di-GMP, VPS and biofilm production (Lim et al., 2006). The two variants can be easily distinguished by the morphology of colonies grown on agar plates (Figure 5.1).

The *vpv* operon has three genes; *vpvA (VC2456)*, *vpvB (VC2455)* and *vpvC (VC2454)*. VpvC is a diguanylate cyclase, while the functions of VpvA and VpvB are unknown. A single nucleotide polymorphism in *vpvC* can lock cells in the rugose state, but this is lost in *vpvA* or *vpvC* deletion mutants (Beyhan and Yildiz, 2007). Deletion of *vpvC* reduces c-di-GMP levels and transcription of *vpsL* (Townsley and Yildiz, 2015).

The level of c-di-GMP is maintained by the activity of diguanylate cyclases (DGC) and phosphodiesterases (PDE) (Conner et al., 2017). Typically, elevated cyclic-di-GMP promotes biofilm formation and the repression of motility. Many phenotypic changes reported in *V. cholerae,* resulting from mutation of genes involved in c-di-GMP metabolism, have used strains and rugose variants that have high levels of c-di-GMP (Beyhan et al., 2008, Beyhan and Yildiz, 2007). However, c-di-GMP levels are known to vary among *V. cholerae* strains (Satchell et al., 2016).

In this chapter we characterise the phenotypic consequences of VpsT mediated activation of *vpv* transcription.

Figure 5.1: The colony morphology of *V. cholerae* rugose and smooth phase variants. Photograph of an agar plate showing smooth and rugose phase variants. Smooth colonies are circular, convex, glossy and slightly translucent. Rugose colonies are more irregularly shaped, raised with an uneven surface and opaque.

5.2 Clinical isolates of *V. cholerae* have differing levels of c-di-GMP

To measure the intracellular concentration of c-di-GMP we used a reporter plasmid (pFY4535, gifted from the Yildiz laboratory, UCSC) that encodes two fluorescent reporter proteins (Zamorano-Sánchez et al., 2019). The production of the red fluorescent protein, TurboRFP, is regulated by two c-di-GMP sensitive riboswitches arranged in tandem. Therefore, TurboRFP expression is proportional to intracellular c-di-GMP (Zhou et al., 2016). The cyan fluorescent protein, AmCyan, is produced constitutively and so serves as a control.

V. cholerae derivatives transformed with plasmid pFY4535 were plated on LB agar and the fluorescence due to each reporter protein was measured. The TurboRFP:AmCyan relative fluorescence intensity (RFI) was then calculated as the ratio of the two fluorescent signals. First, we compared intracellular c-di-GMP from a selection of frequently studied El Tor strains of *V. cholerae* to see if the E7946 strain we use is typical. We included three different wildtype E7946 stocks from our laboratory, as well as N16961, A1552 and C6706. The mean RFI and the standard deviation are plotted in Figure 5.2. E7946 (stock 1 and 3), N16961 and A1552 all had similar levels of c-di-GMP. The E7946 (stock 2) had much lower levels of c-di-GMP compared to the other two E7946 stocks. C6706 also had lower c-di-GMP compared to the other strains tested.

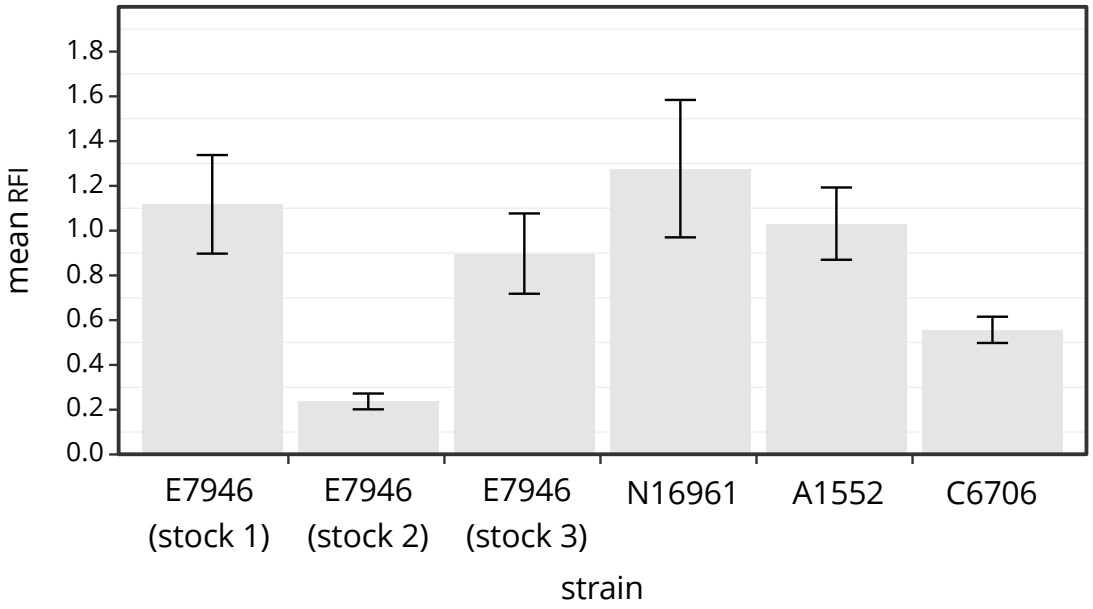

Figure 5.2: The intracellular concentration of c-di-GMP varies amongst *V. cholerae* strains. Results of c-di-GMP assay using the pFY4535 biosensor plasmid. Bacteria were spotted onto LB agar plates. Data are the mean relative fluorescence intensity of four replicates and errors bar show the standard deviation.

5.3 The intracellular c-di-GMP concentration of *vpsT* mutants is reduced

We next used the same c-d-GMP assay to determine if *vpsT* deletion changed c-di-GMP levels. Visual inspection of the plates can be used as an indicator of larger differences between strains, since production of the fluorescence proteins affects the colour of colonies (Figure 5.3A). Note the orange colour of wildtype colonies (left, labelled 'E7') compared to the green colour of *ΔvpsT* colonies. The mean RFI are plotted in Figure 5.3B. Compared to the wildtype the *ΔvpsT* mutant had a low c-di-GMP concentration.

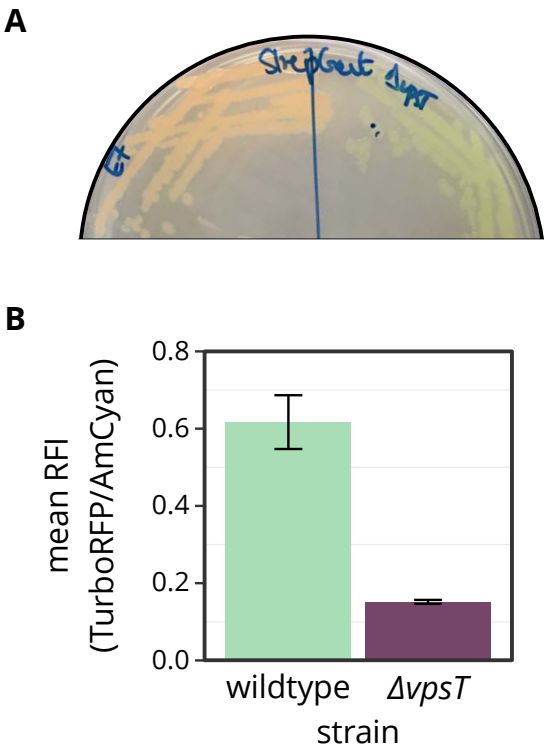

Figure 5.3: The Δ*vpsT* mutant has reduced c-di-GMP. A) Photo of agar plate comparing wildtype E7946 colonies with the Δ*vpsT* mutant grown on LB agar plates. B) Mean relative fluorescence intensity (RFI) of bacterial colonies. The strains carry the c-di-GMP reporter plasmid and incubated for 18 hours at 30 °C. Error bars represent standard deviation.

5.4 VpsT regulation of the *vpv* operon contributes to maintaining c-di-GMP levels

Given that the third gene of the *vpv* operon, *vpvC,* encodes a diguanylate cyclase we surmised that it could contribute to maintaining c-di-GMP levels. We constructed derivatives of the E7946 (stock 1) to determine how VpsT regulation of *vpv* transcription might impact intracellular c-di-GMP. In Chapter 4 we characterised the promoter region upstream of *vpvA,* the first gene of the operon, and found that VpsT activated transcription via *vpvA* P1. Hence, we constructed a *ΔvpvABC* strain by deleting the DNA region from upstream of *vpvA* to the stop codon of *vpvC.* In addition, we made strains lacking the *vpvA P2* promoter*,* with or without the VpsT T box (Figure 5.4). We cultured these *vpv* mutant derivatives, transformed with plasmid pFY4535, and spotted the suspension onto LB agar plates (Figure 5.5A. To normalise the RFI of multiple replicates, the mean RFI for the wildtype strain was made equal to 1, as indicated by the dotted horizontal line. The plot shows the mean RFI of four biological replicates, relative to the wildtype. A representative plate is shown in the photograph and the fluorescence intensity due to each of the fluorescent proteins, AmCyan and TurboRFP, is indicated by the greyscale images. Deletion of the complete *vpvABC* operon resulted in no change in c-di-GMP levels, while *vpvA P2* deletion resulted in a modest reduction of c-di-GMP compared to the wildtype. Mutations made to the VpsT T box further reduced c-di-GMP levels.

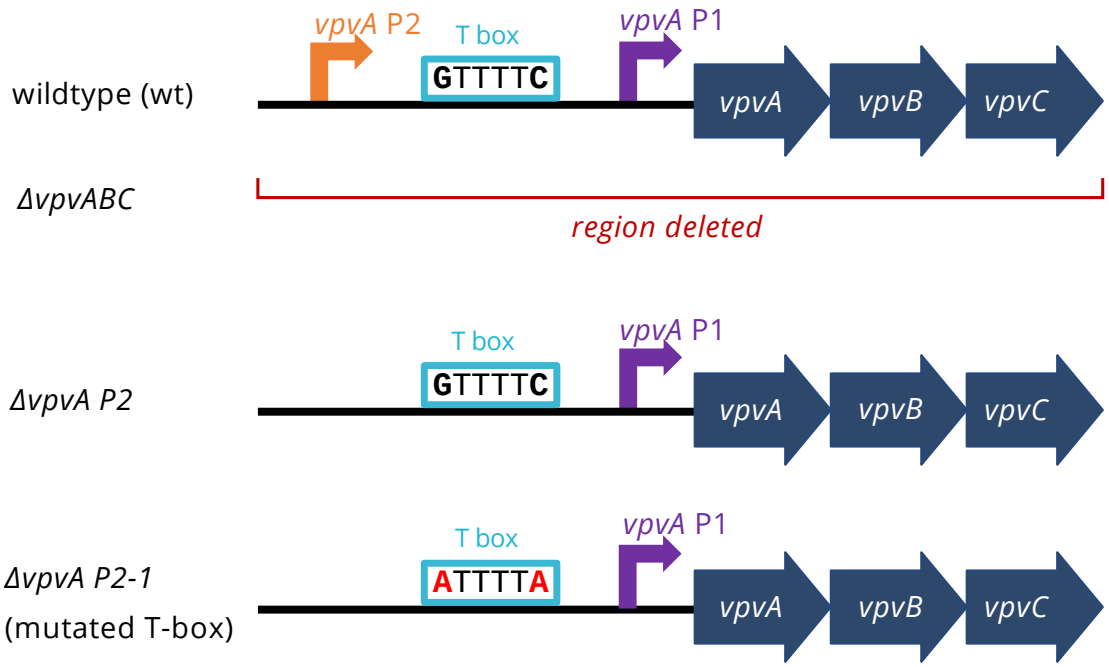

Figure 5.4: Mutagenesis of the *vpv* locus. Cartoon showing the mutations made to the *vpv* locus, compared to the wildtype. Transcription start sites are shown as bent arrows, the location of the VpsT T box is shown as a blue box with key sequence residues. Genes are shown as navy arrows.

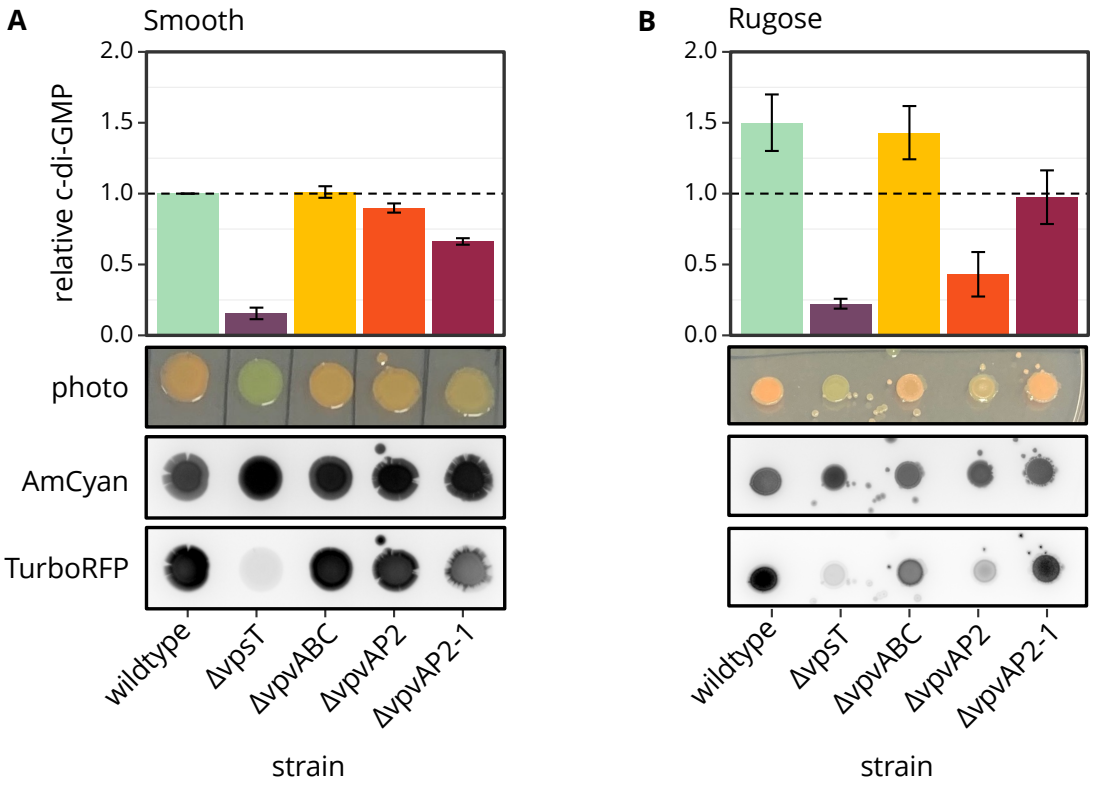

Figure 5.5: VpsT regulation of the *vpv* operon affects intracellular c-di-GMP. Plots show the c-di-GMP levels relative to the wildtype smooth phase variant. Images show representative colonies used for quantification, as a photograph showing the colony morphology and the signal from the AmyCyan and TurboRFP fluorescent proteins, respectively. A) C-di-GMP levels of smooth phase variants. B) C-di-GMP levels of rugose phase variants.

We then repeated this analysis using rugose phase variants (Figure 5.5B). The RFI was normalised to the wildtype smooth bacteria. The wildtype rugose variant had higher c-di-GMP levels compared to the smooth variant. The *ΔvpsT* rugose strain was much less corrugated than the others and had very low c-di-GMP. The *ΔvpvABC* rugose variant, like the smooth, was unchanged compared to the wildtype. The rugose *ΔvpvA P2* variant had lower c-di-GMP compared both to the rugose wildtype and the smooth *ΔvpvA P2* bacteria. The *ΔvpvA P2-1* rugose variant had lower c-di-GMP compared to the wildtypes, but higher c-di-GMP than its smooth equivalent.

5.5 Motility is unaffected by mutations to the *vpv* operon

C-di-GMP levels are inversely related to motility. Given the observed differences in c-di-GMP levels in the *vpv* mutants we tested to see if this would impact bacterial motility. The bacteria were cultured overnight in LB broth and then the suspension was used to inoculate motility LB plates (containing 0.3 % agarose). The plates were imaged 18 hours after inoculation and the diameter of migration was recorded (Figure 5.6). All the strains migrated up to 10 mm from the inoculation site and there was no change in motility compared to the wildtype in these conditions.

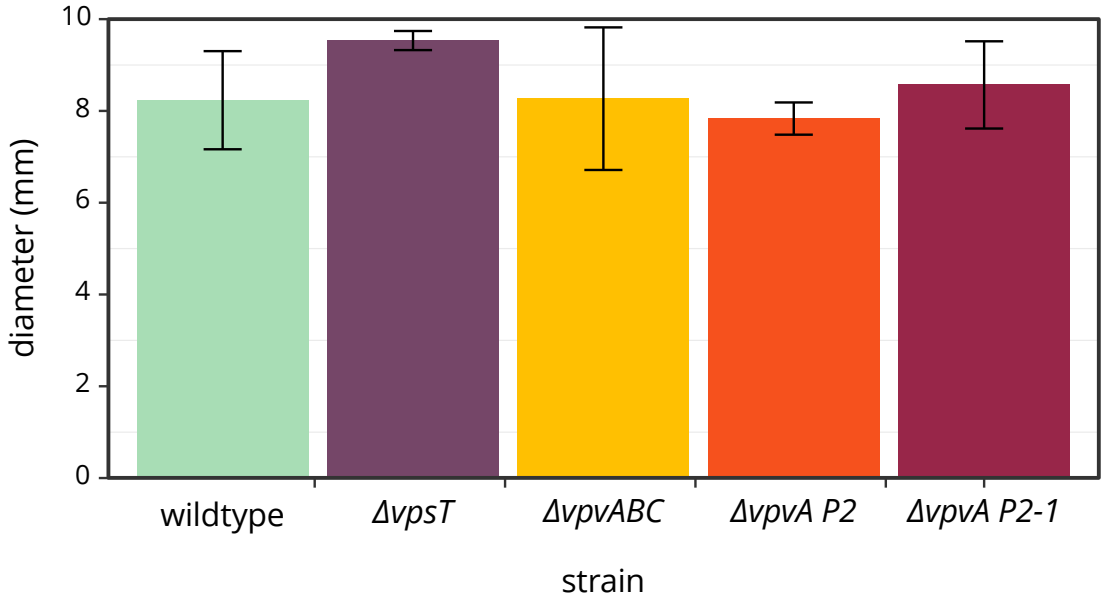

Figure 5.6: VpsT regulation of the *vpv* operon does not affect bacterial motility. Plot shows the diameter of migration in soft agar plates. Data are means from three replicates and errors show the standard deviation.

5.6 The frequency of rugose phase switching is increased by disrupting the *vpv* operon

Phase variation can be induced by growing cells in nutrient deplete media (Ali et al., 2002). We inoculated Alkaline phosphate water (APW) using a single smooth colony and incubated statically for 72 hours. The culture was plated and the number of smooth and rugose colonies was counted. Rugose colonies have a corrugated morphology and are raised compared to smooth colonies (Figure 5.1). However, the degree of rugosity does vary. The frequency of phase switching is represented as a percentages of colonies that have the rugose morphology (Table 5.1). We found that mutating the *vpv* locus altered the rate of smooth to rugose switching. More than 50 % of the colonies lacking *vpvA P2,* with or without the T box, were rugose. For comparison, 34 % of wildtype colonies were rugose. The *ΔvpvABC* mutant had a smaller increase in the number of rugose colonies (35 %). As expected, the Δ*vpsT* mutant had the lowest percentage of rugose phase colonies.

Table 5.1: Frequency of rugose colony formation

strain	% rugose colonies[1,2]
wildtype	34 (1.9)
ΔvpsT	12 (2.2)
ΔvpvABC	35 (4.3)
ΔvpvA P2	61 (3.5)
ΔvpvA P2-1	56 (2.1)

[1] Values represent mean average percentages of colonies shifting to rugose colony morphology from three independent experiments

[2] Values in parantheses are standard errors

5.7 Biofilm formation is reduced in *vpv* mutants

Biofilm formation is typically enhanced by elevated c-di-GMP. We measured biofilm formation using crystal violet assays. We noted previously that three wildtype E7946 stocks had differing levels of c-di-GMP so we tested to see if this meant they would also produce different amounts of biofilm (Figure 5.7). The E7946 (stock 1) produced the most biofilm, while the other two stocks produced less biofilm than the *ΔvpsT* mutant.

We then used the different *vpv* locus derivatives to see if they had an effect on biofilm formation (Figure 5.8). As shown previously, the *ΔvpsT* mutant produced less biofilm compared to the wildtype. The *ΔvpvABC, ΔvpvA P2* and *ΔvpvA P2-1* mutants also produced less biofilm compared to the wildtype, similar to *ΔvpsT.*

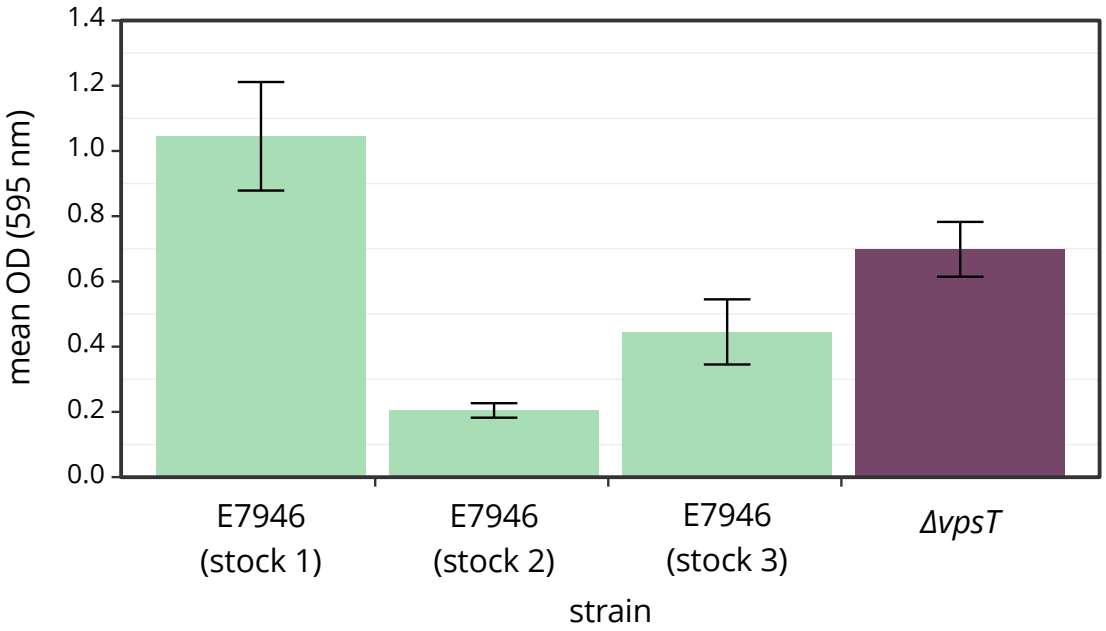

Figure 5.7: Wildtype E7946 stocks produce different amounts of biofilm. Results of crystal violet biofilm formation assays using wildtype *V. cholerae* E7946 stocks and *vpsT* mutant. Bacteria were grown in static 96-well plates. Surface-attached biofilms were stained with crystal violet and then solubilised in ethanol. Data are mean optical density (595 nm) values for three replicates and error bars show the standard deviation.

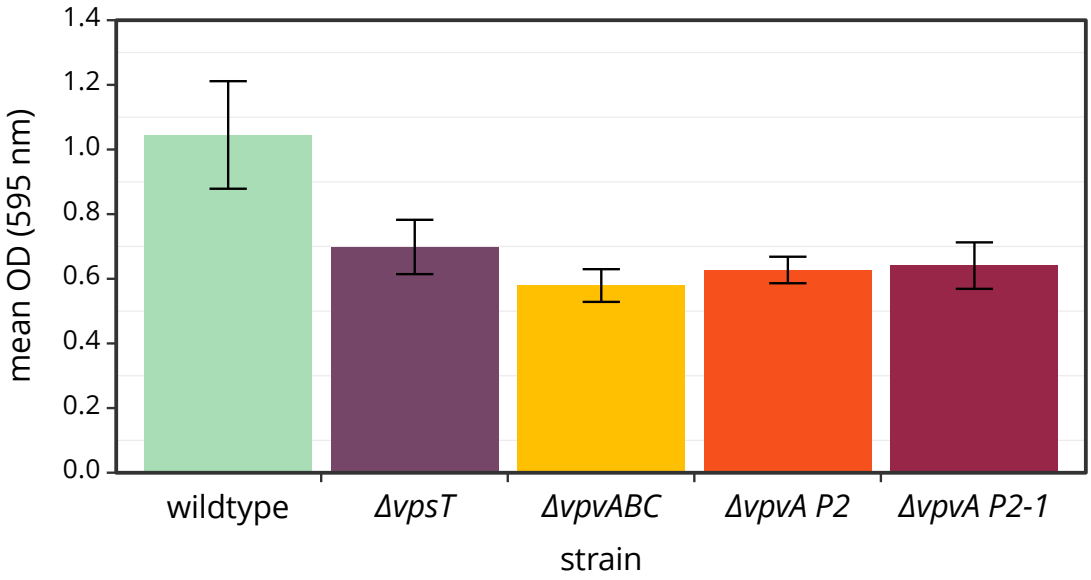

Figure 5.8: Deletion of *vpsT* and mutating the *vpv* locus reduces biofilm formation. Results of crystal violet biofilm formation assays. Bacteria were grown in static 96-well plates. Surface-attached biofilms were stained with crystal violet and then solubilised in ethanol. Data are mean optical density (595 nm) values for three replicates and error bars show the standard deviation.

5.8 *rpoS* SNPs inhibit biofilm formation in lab-adapted bacteria

We were intrigued by the c-di-GMP and biofilm deficient wildtype E7946 (stocks 2 and 3). A recent publication describing lab adaptation of *V. cholerae* strains prompted us to isolate DNA from these different wildtype stocks for whole genome sequencing (Dorman and Thomson, 2020). We also included a separate lab-adapted stock (stock 4), which was no longer naturally transformable, in our analysis. Following alignment of the sequenced genomes to the reference E7946 sequence we checked for single nucleotide polymorphisms (SNP) in coding DNA regions. Table 5.2 summarises the occurrence of SNPs found in the four sequenced genomes. In total, we identified 12 SNPs in gene coding DNA on chromosome I and a single SNP on chromosome II. There were eight SNPs that occurred in all of the sequenced strains, compared to the reference sequence. Four SNPs are predicted to result in nonsynonymous mutations that alter the sequence of amino acids. Two of these resulted in mutations to *VC0059,* the others in *tuf* and *rpsL.*

The E7946 stock 4 had three SNPs, two of which were unique, including a synonymous mutation to *frhA (VC1620)*. The *frhA* gene is a VpsT target, described in the previous Chapter, involved in bacterial adhesion. There are also SNPs in *hapR* (the quorum-sensing master-regulator) and *rpoS* which are likely to contribute to the loss of ability to undergo natural transformation. RpoS is essential for natural transformation but is commonly mutated in lab-adapted bacteria (Dalia, 2016, Dorman and Thomson, 2020).

Table 5.2: Results from whole-genome sequencing

reference		E7946 stock					
position	base	1	2	3	4	gene	type
chromosome I							
53608	C	T	T	T	T	*VC0059*	nonsynonymous
53611	C	A	A	A	A	*VC0059*	nonsynonymous
254864	G	A	A	A	A	*tagH (VC0247)*	synonymous
254867	G	A	A	A	A	*tagH (VC0247)*	synonymous
334400	G	C	C	C	C	*tuf*	nonsynonymous
377373	A	G	G	G	G	*rpsL*	nonsynonymous
399804	A	G	G	G	G	*VC0386*	synonymous
564244	G	G	T	G	T	*rpoS (VC0534)*	nonsense
611884	C	C	C	C	T	*hapR (VC0583)*	nonsynonymous
934754	T	T	A	A	T	*VC01969*	synonymous
1735813	C	C	A	C	C	*frhA(VC1620)*	synonymous
1737187	C	C	C	C	A	*frhA(VC1620)*	synonymous
chromosome II							
343650	A	G	G	G	G	*gspI*	nonsynonymous

Note:

Red coloured nucleotides indicate a change compared to the base at the same position in the reference E7946 genome sequence. Green indicates the same base occurs at that position.

The E7946 stock 1 had no SNPs, except the eight shared by all sequenced strains. E7946 stock 2 had the same nonsense *rpoS* mutation as the LA strain. There were additional SNPs resulting in synonymous mutations in *VC01969*, a gene of unknown function, and a second unique SNP in *frhA (VC1620).* Compared with the other E7946 wildtype stocks, stock 2 had lower c-di-GMP (Figure 5.2) and reduced ability to form biofilm (Figure 5.7). E7946 stock 3 has the same SNP in *VC01969*, but no others except from eight shared by all the strains. E7946 stock 3 had similar c-di-GMP levels compared to stock 1, but produced less biofilm, though the effect was not as pronounced as for stock 2.

5.9 Discussion

Our initial observation that *vpsT* mutants had lower c-di-GMP levels led us to focus our attention on VpsT targets with roles in c-di-GMP metabolism (Figure 5.3). In the VpsT ChIP-seq experiment we identified four targets with potential roles in c-di-GMP metabolism. We did not find a strong VpsT regulatory effect on transcription of the two targets with predicted phosphodiesterase domains, *VC1710* and *VC1851* (Figure 3.9). There was also little change in transcription from the regulatory DNA upstream of *cdgB*, a diaguanylate cyclase, with or without VpsT. However, transcription activity from the regulatory region upstream of *vpvA,* the first gene of the *vibrio phase variation (vpvABC)* operon, was strongly activated by VpsT. The third gene, *vpvC*, encodes a diguanylate cyclase. We thought this could be important for allowing VpsT to maintain c-di-GMP levels.

In Chapter 4 we showed that VpsT activated transcription via *vpvA P1* (Figure 5.4). To assess the role of VpsT mediated regulation of the *vpv* operon in maintaining c-di-GMP levels, we introduced these mutations

to the chromosome. We used these mutant derivatives, with the c-di-GMP biosensor plasmid to measure c-di-GMP levels (Figure 5.5). This was done in both smooth and rugose phase variants. For the smooth variants, deletion of *vpvA P2* reduced the level of c-di-GMP, and this was reduced still further when the T box was also mutated. In contrast, rugose variants with the *vpvA P2* deletion had much lower c-di-GMP compared to the rugose wildtype, which slightly recovered when the T box was mutated. These results suggest that the primary promoter for *vpvA* transcription might differ between smooth (*vpvA P1*) and rugose (*vpvA P2*) variants. There was no difference when the *vpvABC* operon was deleted in smooth or rugose bacteria. We expect this , could be due to redundancy, given the large number of diguanylate cyclases in *V. cholerae.* Often individual deletions do not elicit a large change in c-di-GMP levels (Dalia et al., 2017). Typically, studies of c-di-GMP metabolism in *V. cholerae* use derivatives in which a large number of DGCs are deleted (Shikuma et al., 2012).

We found VpsT did not bind upstream of *vpsT,* despite previous reports indicating auto-regulation (Casper-Lindley and Yildiz, 2004). In these studies *vpsT-lacZ* fusions were less active in *vpsT* mutants. However, VpsT is not required for c-di-GMP induction of *vpsT* (Srivastava et al., 2011). Here we showed VpsT is important for maintaining c-di-GMP levels and this could be an indirect mechanism for auto-regulation, perhaps via VpsR, another c-di-GMP sensitive transcription factor that directly regulates transcription of *vpsT* (Hsieh et al., 2018).

The model we proposed in Figure 5.9 suggests a positive feedback loop in which VpsT, activated by binding c-di-GMP, enhances the expression of *vpvC,* to increase the production of c-di-GMP. Given the role of c-di-GMP in regulating *V. cholerae* motility, biofilm formation and phase variation, we next studied if the change in c-di-GMP levels would alter

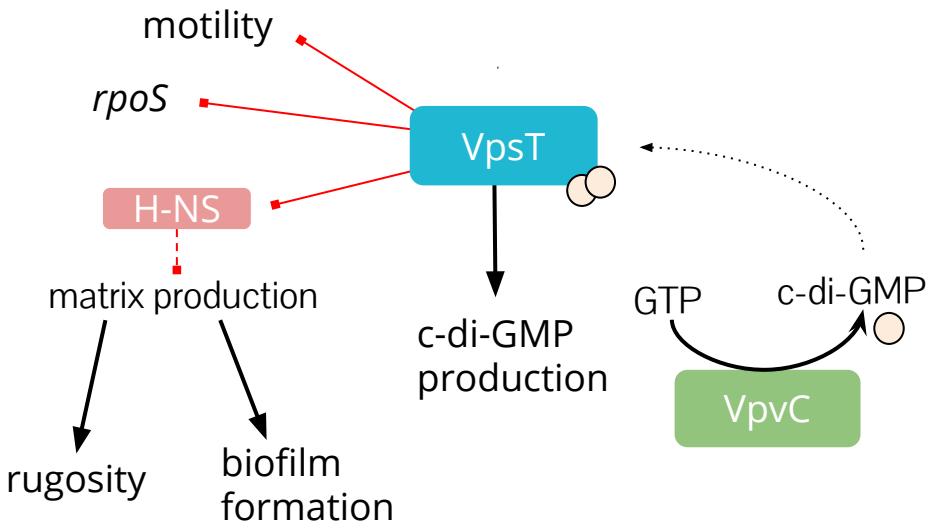

Figure 5.9: Proposed model for regulation at the *vpv* operon. VpsT promotes rugosity and biofilm formation by inhibiting the H-NS repression of matrix gene transcription (red lines indicate inhibition or repression of transcription). VpsT also represses transcription of motility genes and *rpoS*. VpsT promotes c-di-GMP production by directly activating transcription of the *vpv* operon including the digaunylate cyclase *vpvC*. VpvC increases the intracellular concentration of c-di-GMP that enhances VpsT activity, creating a feed-forward loop (as indicated by the dotted line).

these behaviours.

Motility is typically regulated inversely with biofilm formation, but we found that the mutations made to the *vpv* locus did not alter motility in soft agar plates (Figure 5.6). However, we did show that biofilm formation in all of the *vpv* mutants was reduced (Figure 5.8). Deletion of *vpvABC* resulted in a decrease in biofilm formation, but not in c-di-GMP levels. The rate of phase variation in mutants where the *vpv* operon had been mutated was higher than wildtype strain (Figure 5.5). Phase variation and biofilm formation are similar; both require the production of the vibrio polysaccharide and have many regulators in common (Beyhan et al., 2007). Rugose colonies are frequently used as an indicator for the ability to form biofilm, though the extent to whether phase variation and biofilm formation are the same is unclear. Biofilms are defined as surface-attached communities but *V. cholerae* is known to form floating aggregates as well as pellicles at liquid:air interfaces, these are both enhanced in cultures of rugose phase variants (Yildiz and Schoolnik, 1999). Furthermore, cell density is intricately linked to the regulation of biofilm formation, which is a low-cell density behaviour, while the formation of floating aggregates has been shown to occur at high-cell density (Jemielita et al., 2018).

We also compared c-di-GMP levels in a range of widely studied isolates, as well as multiple stocks of our wildtype E7946 strain. We did this to see if the levels of c-di-GMP and biofilm produced by E7946 was typical of *V. cholerae.* Most research investigating biofilm formation use the rugose variant of A1552 because it produces robust biofilm. In our assays we found that smooth variants of E7946, A552 and N16961 had similar c-di-GMP levels but C6706 was lower. We then compared biofilm formation between the different E7946 stocks. E7946 stocks 2 and 3 produced lower levels of biofilm compared E7946 stock 1. Whole

genome sequencing of the stocks suggest the differences are the result of a nonsense mutation in *rpoS.* We included in our analysis an additional lab-adapted strain and this too had the same *rpoS* mutation as well as a mutation in *hapR.*

6 | Final conclusions

Cholera remains a significant and growing cause of disease and mortality in countries lacking the water and healthcare infrastructure to prevent disease transmission. Outbreaks in Yemen, that began during civil war, have been the most severe on record (WHO, 2018). As an aquatic bacterium *V. cholerae* can persist in the environment between outbreaks. The bacteria can switch lifestyles from a planktonic motile cell to form communities in surface-attached biofilms that allow them to tap into scarce resources and also primes them for infection. VpsT is an activator of genes that produce the vibrio polysaccharide (VPS) and is at the centre of a regulatory network that governs this lifestyle switch. Earlier research suggested a wider role for VpsT in this transition but its regulon remained poorly defined. Here, we used ChIP-seq, biochemical and phenotypic assays to define the global binding profile for VpsT, characterise its binding sequence motif and understand the contribution of VpsT target genes in lifestyle switching.

Using ChIP-seq we identified 23 binding targets of VpsT on the larger chromosome I and none on chromosome II (chrII). Interestingly, the *vpsT* gene is on chrII and previous reports had suggested that VpsT could auto-regulate its own transcription (Casper-Lindley and Yildiz, 2004). Instead, we propose that VpsT indirectly auto-regulates *vpsT* transcription via modulation of c-di-GMP levels. This supports an earlier observation that found VpsT was not required for c-di-GMP induction of

vpsT transcription (Srivastava et al., 2011). This regulation is likely due to VpsR, a known direct activator of *vpsT* and other genes involved in lifestyle switching. VpsR is less sensitive to c-di-GMP than VpsT but c-di-GMP binding enhances VpsR activity at some promoters (Hsieh et al., 2020). The full extent of the overlap between VpsT and VpsR regulons is unknown but *vpsR,* unlike *vpsT*, is essential for biofilm formation (Yildiz et al., 2001). Further characterisation of VpsR should be a major focus for understanding the regulation underpinning lifestyle switching in *V. cholerae.*

We used the DNA sequences at the VpsT ChIP-seq targets to redefine the 'T-box', improving on previous predictions that relied on much less data (Zamorano-Sánchez et al., 2015). We found that the T-box is highly degenerate, except the requirement for a C or G in positions 3 and 8 within the 10 nucleotide sequence motif. We validated this binding motif using the well-characterised *vpsL* promoter, which is located upstream of the second VPS operon. We also showed that this arrangement was essential in all the novel VpsT target promoters we studied. The identification of this motif overcomes the unexplained degeneracy of identified VpsT binding sites that was not evident in the longer, highly conserved motif published previously. We then searched for T-boxes across chrII and found 97 putative VpsT binding sites. We did not identify any VpsT ChIP-seq peaks on chrII in our analysis. This could be because VpsT does not bind to any targets on chrII, despite the occurrence of T-box sequences. Alternatively, VpsT targets on chrII may not be as accessible or may only be weakly bound by VpsT in the experimental conditions we used. There are differences in global gene expression between the two chromosomes. For example genes on chrII are expressed at low levels in typical laboratory conditions, but more highly expressed in vivo (Xu et al., 2003). Visual inspection of

the ChIP-seq dataset shows there was some weak enrichment for VpsT at sites on chrII but whether or not these are indicative of real VpsT binding sites remains undetermined.

We characterised transcriptional regulation by VpsT at four of the novel targets where we observed the greatest change in promoter activity in response to VpsT. The first target is located in the poorly understood genomic island VSP-II. Given the uniqueness of the VSP-II to El Tor strains it could have important implications for understanding the enhanced environmental fitness of this biotype (O'Shea, 2004). There has been a renewed interested in the VSP-II and a number of its genes have been characterised recently by Murphy et al. (2021) who showed that zinc-starvation played an important role in regulating its expression. We showed that there are multiple VpsT binding sites at the VSP-II but we focussed on only one target between *VC0510* and *aerB (VC0512),* so the full extent of VpsT regulation of genes in the VSP-II remains unexplored. Furthermore, because the VSP-II is so poorly expressed in typical laboratory conditions we were unable to clearly define the regulatory role of VpsT but recent insights into the zinc-starvation response provide possible experimental conditions that can be exploited in future to study transcriptional regulation of the VSP-II.

The VpsT target site between the divergent genes *VC1303* and *VC1304* happened to be an example of a bidirectional promoter. We reported in Warman et al. (2021) on the widespread occurrence of bidirectional promoters in prokaryotes. Bidirectional promoters drive transcription of divergent genes on opposite strands via shared -10 promoter elements. We demonstrated that VpsT represses transcription of both divergent genes by binding to a position overlapping the promoter. Unfortunately, the role of these genes in *V. cholerae* are unknown. Predictions suggest

they could function in carbon metabolism and therefore could contribute to the altered metabolic state of cells within biofilms, including the dormant viable non-culturable bacteria found in aquatic environments.

FrhA is a large adhesin that has been shown to bind to numerous surfaces types including chitin and erythrocytes (Syed et al., 2009). Recent research found that transcription of *frhA* was growth-phase dependent in *vpsT* mutants (Kitts et al., 2019). Previously, ChIP-seq analysis of CRP in *V. cholerae* identified a binding site for CRP upstream of *frhA* (Manneh-Roussel et al., 2018). We showed that repression of *frhA* and the divergent gene *frhC* is directly regulated by VpsT binding. Furthermore, that regulation of *frhC* was achieved via a separate VpsT binding site. This suggests the transcription of these two genes is tightly regulated in a complex arrangement involving multiple VpsT binding sites and additional regulatory proteins.

There is a multiplicity of proteins involved in the production and degradation of c-di-GMP. The competition between DGCs and PDEs modulates c-di-GMP levels in the cell (Conner et al., 2017). The c-di-GMP effectors are also diverse, including riboswitches, proteolytic pathways and transcription factors. Many DGC and PDE proteins are membrane-bound and typically comprise a sensory domain. However, the relevant environmental stimuli remain poorly defined. Lower temperatures stimulate the activity of six DGCs in an additive manner, however the mechanism for this is unclear (Townsley and Yildiz, 2015). Also unclear is how the global level of c-di-GMP results in specific regulatory responses, given the large number of proteins able to manipulate it (Hengge, 2021). Local pools of c-di-GMP have been proposed as a solution to allow for targeted regulation by a c-di-GMP effector. For example, in *E. coli*, some modulator-effector (protein-protein) interactions have been characterised and in *V. cholerae* individual DGCs correlate more strongly

with biofilm formation than overall c-di-GMP levels (Massie et al., 2012, Sarenko et al., 2017).

We demonstrated a role for VpsT in maintaining c-di-GMP levels via a positive feedback loop. We observed a large decrease in c-di-GMP in *vpsT* deletion mutants and there are four VpsT targets that have predicted or characterised roles in c-di-GMP metabolism. Transcription from the *vpvA (VC2456)* regulatory region was activated by VpsT so we focussed attention on this target, though the other targets may also contribute to c-di-GMP levels. There are two promoters upstream of *vpvA*. Disruption of the T-box upstream of the *vpvA P1* promoter reduced c-di-GMP levels and biofilm formation. Our results also suggest that TSS selection at *vpv* operon could be phase-dependent, though this requires further investigation. To date there has been no evidence for VpsT either co-localising or directly interacting with proteins involved in c-di-GMP metabolism. The majority of VpsT localises within the cytoplasmic fraction of the cell in a c-di-GMP dependent manner, though this was independent of the activity of five DGCs (Shikuma et al., 2012). However, they also identified a smaller amount of VpsT localised at the membrane. This could allow VpsT to make use of local pools of c-di-GMP at sites associated with membrane-bound DGCs or PDEs.

An additional complexity to studying lifestyle switching in *V. cholerae* is the differentiation between biofilm formation, phase variation and the formation of aggregates. These terms are often used interchangeably and rugose colonies are used as an indicator for biofilm formation. Rugose variants form floating aggregates when grown in broth culture, not unlike the aggregates described by Jemielita et al. (2018), that contrary to biofilm formation occur at high-cell density. Rugose variants are identified by their wrinkled colony morphology. This morphological change is due to increased production of vibrio polysaccharide (VPS) but

the molecular mechanisms behind this are poorly understood. Regulation of VPS production in rugose colonies involves the interplay of the quorum sensing regulators HapR and AphA, as well as VpsR and VpsT (Yildiz et al., 2004). However, the heritability of the rugose colony morphology suggests genetic changes are responsible for the appearance of some rugose colonies. For example, a single point mutation to the diguanylate cyclase (DGC) gene *vpvC (VC2454)* was found to be responsible for the wrinkled morphology of one rugose colony (Beyhan and Yildiz, 2007). Presumably, mutations in any number of other loci might result in the over-production of c-di-GMP or conversely down-regulation of phosphodiesterases (PDEs), or otherwise alter and remove the regulatory constraints that modulate production of VPS.

Lifestyle switching is an important factor in the success of *V. cholerae,* influencing its environmental persistence, ability to colonise the small intestine and cause disease. Understanding the molecular mechanisms promoting lifestyle switching remains an important field of research with many outstanding questions. Furthermore, biofilms are the predominant lifestyle of bacteria throughout the natural world, with costly implications due to biofouling in industrial and medical settings, hence understanding the regulatory mechanisms that govern this switch are important beyond the suffering caused by cholera.

References

Alam, M., Hasan, N. A., Sadique, A., Bhuiyan, N. A., Ahmed, K. U., Nusrin, S., Nair, G. B., Siddique, A. K., Sack, R. B., Sack, D. A., Huq, A., and Colwell, R. R. (2006). Seasonal Cholera Caused by *Vibrio cholerae* Serogroups O1 and O139 in the Coastal Aquatic Environment of Bangladesh. *Applied and Environmental Microbiology*, 72(6):4096–4104.

Alam, M., Sultana, M., Nair, G. B., Siddique, A. K., Hasan, N. A., Sack, R. B., Sack, D. A., Ahmed, K. U., Sadique, A., Watanabe, H., and others (2007). Viable but nonculturable *Vibrio cholerae* O1 in biofilms in the aquatic environment and their role in cholera transmission. *Proceedings of the National Academy of Sciences*, 104(45):17801–17806.

Ali, A., Rashid, M. H., and Karaolis, D. K. R. (2002). High-Frequency Rugose Exopolysaccharide Production by *Vibrio cholerae*. *Applied and Environmental Microbiology*, 68(11):5773–5778.

Au, K. F., Jiang, H., Lin, L., Xing, Y., and Wong, W. H. (2010). Detection of splice junctions from paired-end rna-seq data by splicemap. *Nucleic Acids Research*, 38(14):4570–4578. PMID:20371516.

Ayala, J. C., Wang, H., Silva, A. J., and Benitez, J. A. (2015). Repression by H-NS of genes required for the biosynthesis of the *Vibrio cholerae* biofilm matrix is modulated by the second messenger cyclic diguanylic acid: Regulation of biofilm formation by H-NS. *Molecular Microbiology*, 97(4):630–645.

Bailey, T. and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pages 28–36.

Bartlett, D. and Azam, F. (2005). Chitin, Cholera, and Competence. *Science*, 310(5755):1775–1776.

Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-Mediated General Stress Response in *Escherichia coli*. *Annual Review of Microbiology*, (65):41.

Beyhan, S., Bilecen, K., Salama, S. R., Casper-Lindley, C., and Yildiz, F. H. (2007). Regulation of Rugosity and Biofilm Formation in *Vibrio cholerae*: Comparison of VpsT and VpsR Regulons and Epistasis Analysis of vpsT, vpsR, and hapR. *Journal of Bacteriology*, 189(2):388–402.

Beyhan, S., Odell, L. S., and Yildiz, F. H. (2008). Identification and Characterization of Cyclic Diguanylate Signaling Systems Controlling Rugosity in *Vibrio cholerae*. *Journal of Bacteriology*, 190(22):7392–7405.

Beyhan, S. and Yildiz, F. H. (2007). Smooth to rugose phase variation in *Vibrio cholerae* can be mediated by a single nucleotide change that targets c-di-GMP signalling pathway. *Molecular Microbiology*, 63(4):995–1007.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene*, 2(2):95–113.

Borukhov, S. and Nudler, E. (2008). RNA polymerase: the vehicle of transcription. *Trends in Microbiology*, 16(3):126–134.

Brenzinger, S., van der Aart, L. T., van Wezel, G. P., Lacroix, J.-M., Glatter, T., and Briegel, A. (2019). Structural and Proteomic Changes in Viable but Non-culturable *Vibrio cholerae*. *Frontiers in Microbiology*, 10.

Bridges, A. A., Fei, C., and Bassler, B. L. (2020). Identification of signaling pathways, matrix-digestion enzymes, and motility components controlling *Vibrio cholerae* biofilm dispersal.

Browning, D. F. and Busby, S. J. W. (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2(1):57–65.

Browning, D. F. and Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology*, 14(10):638–650.

Browning, D. F., Butala, M., and Busby, S. J. (2019). Bacterial Transcription Factors: Regulation by Pick "N" Mix. *Journal of Molecular Biology*.

Butala, M., Sonjak, S., Kamenšek, S., Hodošček, M., Browning, D. F., Žgur Bertok, D., and Busby, S. J. W. (2012). Double locking of an *Escherichia coli* promoter by two repressors prevents premature colicin expression and cell lysis. *Molecular Microbiology*, page 11.

Cakar, F., Zingl, F. G., Moisi, M., Reidl, J., and Schild, S. (2018). In vivo repressed genes of *Vibrio cholerae* reveal inverse requirements of an H^+/Cl^- transporter along the gastrointestinal passage. *Proceedings of the National Academy of Sciences*, 115(10):E2376–E2385.

Caro, F., Caro, J. A., Place, N. M., and Mekalanos, J. J. (2020). Transcriptional Silencing by TsrA in the Evolution of Pathogenic *Vibrio cholerae* Biotypes. *mBio*, 11(6).

Casper-Lindley, C. and Yildiz, F. H. (2004). VpsT Is a transcriptional Regulator Required for Expression of *vps* Biosynthesis Genes and the Development of Rugose Colonial Morphology in *Vibrio cholerae* O1 El Tor. *Journal of Bacteriology*, 186(5):1574–1578.

Cech, T. and Steitz, J. (2014). The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell*, 157(1):77–94.

Charles, R. C. and Ryan, E. T. (2011). Cholera in the 21st century:. *Current Opinion in Infectious Diseases*, 24(5):472–477.

Chatterjee, R., Nag, S., and Chaudhuri, K. (2008). Identification of a new RTX-like gene cluster in *Vibrio cholerae*. *FEMS Microbiology Letters*, 284(2):165–171.

Clemens, J. D., Nair, G. B., Ahmed, T., Qadri, F., and Holmgren, J. (2017). Cholera. *The Lancet*, 390(10101):1539–1549.

Colwell, R. R. (1996). Global Climate and Infectious Disease: The Cholera Paradigm. *Science*, 274(5295):2025–2031.

Comstock, L. E., Johnson, J. A., Michalski, J. M., Morris Jr, J. G., and Kaper, J. B. (1996). Cloning and sequence of a region encoding a surface polysaccharide of *Vibrio cholerae* O139 and characterization of the insertion site in the chromosome of *Vibrio cholerae*O1. *Molecular Microbiology*, 19(4):815–826.

Conner, J. G., Zamorano-Sánchez, D., Park, J. H., Sondermann, H., and Yildiz, F. H. (2017). The ins and outs of cyclic di-GMP signaling in *Vibrio cholerae*. *Current Opinion in Microbiology*, 36:20–29.

Cont, A., Rossy, T., Al-Mayyah, Z., and Persat, A. (2020). Biofilms deform soft surfaces and disrupt epithelia. *eLife*, 9:e56533.

Cvjetanovic, B. and Barua, D. (1972). The Seventh Pandemic of Cholera. *Nature*, 239:137–138.

Dalia, A. B. (2016). RpoS is required for natural transformation of *Vibrio cholerae* through regulation of chitinases. *Environmental Microbiology*, 18(11):3758–3767. _eprint: https://sfamjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/1462-2920.13302.

Dalia, A. B., McDonough, E., and Camilli, A. (2014). Multiplex genome editing by natural transformation. *Proceedings of the National Academy of Sciences*, 111(24):8937–8942.

Dalia, T. N., Yoon, S. H., Galli, E., Barre, F.-X., Waters, C. M., and Dalia, A. B. (2017). Enhancing multiplex genome editing by natural transformation (MuGENT) via inactivation of ssDNA exonucleases. *Nucleic Acids Research*, 45(12):7527–7537.

Dambach, M., Irnov, I., and Winkler, W. C. (2013). Association of RNAs with *Bacillus subtilis* Hfq. *PLOS ONE*, 8(2):18.

Dame, R. T., Rashid, F.-Z. M., and Grainger, D. C. (2020). Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nature Reviews Genetics*, 21(4):227–242.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2). giab008.

Davis, B. M., Quinones, M., Pratt, J., Ding, Y., and Waldor, M. K. (2005). Characterization of the Small Untranslated RNA RyhB and Its Regulon in *Vibrio cholerae*. *Journal of Bacteriology*, 187(12):4005–4014.

Dennis, P. P. and Bremer, H. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, 3(1).

Dillon, S. C. and Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology*, 8(3):185–195.

Domman, D., Quilici, M.-L., Dorman, M. J., Njamkepo, E., Mutreja, A., Mather, A. E., Delgado, G., Morales-Espinosa, R., Grimont, P. A. D., Lizárraga-Partida, M. L., Bouchier, C., Aanensen, D. M., Kuri-Morales, P., Tarr, C. L., Dougan, G., Parkhill, J., Campos, J., Cravioto, A., Weill, F.-X., and Thomson, N. R. (2017). Integrated view of *Vibrio cholerae* in the Americas. *Science*, 358(6364):789–793.

Dorman, M. J. and Thomson, N. R. (2020). 'Community evolution' – laboratory strains and pedigrees in the age of genomics. *Microbiology*, 166(3):233–238.

Duchi, D. (2016). RNA Polymerase Pausing during Initial Transcription. *Molecular Cell*, 63(6):939–950.

Eickhoff, M. J. and Bassler, B. L. (2018). SnapShot: Bacterial Quorum Sensing. *Cell*, 174(5):1328–1328.e1.

Escobar, L. E., Ryan, S. J., Stewart-Ibarra, A. M., Finkelstein, J. L., King, C. A., Qiao, H., and Polhemus, M. E. (2015). A global map of suitability for coastal *Vibrio cholerae* under current and future climate conditions. *Acta Tropica*, 149:202–211.

Faruque, S. M., Biswas, K., Udden, S. M. N., Ahmad, Q. S., Sack, D. A., Nair, G. B., and Mekalanos, J. J. (2006). Transmissibility of cholera: In vivo-formed biofilms and their relationship to infectivity and persistence in the environment. *Proceedings of the National Academy of Sciences*, 103(16):6350–6355.

Fernandez, N. L., Hsueh, B. Y., Nhu, N. T. Q., Franklin, J. L., Dufour, Y. S., and Waters, C. M. (2020). *Vibrio cholerae* adapts to sessile and motile lifestyles by cyclic di-GMP regulation of cell shape. *Proceedings of the National Academy of Sciences*, 117(46):29046–29054.

Fernandez, N. L., Srivastava, D., Ngouajio, A. L., and Waters, C. M. (2018). Cyclic di-GMP Positively Regulates DNA Repair in *Vibrio cholerae*. *Journal of Bacteriology*, 200(15).

Fernandez, N. L. and Waters, C. M. (2019). Cyclic di-GMP Increases Catalase Production and Hydrogen Peroxide Tolerance in *Vibrio cholerae*. *Applied and Environmental Microbiology*, 85(18).

Flemming, H.-C., Wingender, J., Szewzyk, U., Steinberg, P., Rice, S. A., and Kjelleberg, S. (2016). Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology*, 14(9):563–575.

Fong, J. C. N., Syed, K. A., Klose, K. E., and Yildiz, F. H. (2010). Role of Vibrio polysaccharide (*vps*) genes in VPS production, biofilm formation and *Vibrio cholerae* pathogenesis. *Microbiology*, 156(9):2757–2769.

Fong, J. C. N. and Yildiz, F. H. (2008). Interplay between Cyclic AMP-Cyclic AMP Receptor Protein and Cyclic di-GMP Signaling in *Vibrio cholerae* Biofilm Formation. *Journal of Bacteriology*, 190(20):6646–6659.

Gallego-Hernandez, A. L., DePas, W. H., Park, J. H., Teschler, J. K., Hartmann, R., Jeckel, H., Drescher, K., Beyhan, S., Newman, D. K., and Yildiz, F. H. (2020). Upregulation of virulence genes promotes *Vibrio cholerae* biofilm hyperinfectivity. *Proceedings of the National Academy of Sciences*, 117(20):11010–11017.

Gao, R., Bouillet, S., and Stock, A. M. (2019). Structural Basis of Response Regulator Function. *Annual Review of Microbiology*, 73:175–197.

Garrine, M., Mandomando, I., Vubil, D., Nhampossa, T., Acacio, S., Li, S., Paulson, J. N., Almeida, M., Domman, D., Thomson, N. R., Alonso, P., and Stine, O. C. (2017). Minimal genetic change in *Vibrio cholerae* in Mozambique over time: Multilocus variable number tandem repeat analysis and whole genome sequencing. *PLOS Neglected Tropical Diseases*, 11(6):e0005671.

Goldberg, J. B. and Ohman, D. E. (1984). Cloning and expression in *Pseudomonas aeruginosa* of a gene involved in the production of alginate. *Journal of bacteriology*, 158(3):1115–1121.

Grainger, D. C. (2016). Structure and function of bacterial H-NS protein. *Biochemical Society Transactions*, page 9.

Grainger, D. C., Goldberg, M. D., Lee, D. J., and Busby, S. J. W. (2008). Selective repression by Fis and H-NS at the *Escherichia coli dps* promoter. *Molecular Microbiology*, 68(6):1366–1377.

Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics*, 27(7):1017–1018.

Grigorova, I. L. (2004). Fine-tuning of the *Escherichia coli* sigma E envelope stress response relies on multiple mechanisms to inhibit signal-independent proteolysis of the transmembrane anti-sigma factor, RseA. *Genes & Development*, 18(21):2686–2697.

Hahne, F. and Ivanek, R. (2016). *Statistical Genomics: Methods and Protocols*, chapter Visualizing Genomic Data Using Gviz and Bioconductor, pages 335–351. Springer New York, New York, NY.

Harden, T. T., Herlambang, K. S., Chamberlain, M., Lalanne, J.-B., Wells, C. D., Li, G.-W., Landick, R., Hochschild, A., Kondev, J., and Gelles, J. (2020). Alternative transcription cycle for bacterial RNA polymerase. *Nature Communications*, 11(1).

Haseltine, W. A. and Block, R. (1973). Synthesis of Guanosine Tetra- and Pentaphosphate Requires the Presence of a Codon-Specific, Uncharged Transfer Ribonucleic Acid in the Acceptor Site of Ribosomes. *Proc. Nat. Acad. Sci. USA*, 70(5):1564–1568.

Hauryliuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T., and Gerdes, K. (2015). Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nature Reviews Microbiology*, 13(5):298–309.

Haycocks, J. R. J., Sharma, P., Stringer, A. M., Wade, J. T., and Grainger, D. C. (2015). The Molecular Basis for Control of ETEC Enterotoxin Expression in Response to Environment and Host. *PLoS Pathogens*, 11(1):e1004605.

Haycocks, J. R. J., Warren, G. Z. L., Walker, L. M., Chlebek, J. L., Dalia, T. N., Dalia, A. B., and Grainger, D. C. (2019). The quorum sensing transcription factor AphA directly regulates natural competence in *Vibrio cholerae*. *PLOS Genetics*, 15(10):e1008362.

Hengge, R. (2009). Principles of c-di-GMP signalling in bacteria. *Nature Reviews Microbiology*, 7(4):263–273.

Hengge, R. (2021). High-specificity local and global c-di-GMP signaling. *Trends in Microbiology*, page S0966842X21000378.

Hollenbeck, E., Fong, J., Lim, J., Yildiz, F., Fuller, G., and Cegelski, L. (2014). Molecular Determinants of Mechanical Properties of *V. cholerae* Biofilms at the Air-Liquid Interface. *Biophysical Journal*, 107(10):2245–2252.

Hounmanou, Y. M. G., Mdegela, R. H., Dougnon, T. V., Madsen, H., Withey, J. H., Olsen, J. E., and Dalsgaard, A. (2019). Tilapia (*Oreochromis niloticus*) as a Putative Reservoir Host for Survival and Transmission of *Vibrio cholerae* O1 Biotype El Tor in the Aquatic Environment. *Frontiers in Microbiology*, 10:1215.

Hsieh, M.-L., Hinton, D. M., and Waters, C. M. (2018). VpsR and cyclic di-GMP together drive transcription initiation to activate biofilm formation in *Vibrio cholerae*. *Nucleic Acids Research*, 46(17):8876–8887.

Hsieh, M.-L., Waters, C. M., and Hinton, D. M. (2020). VpsR directly activates transcription of multiple biofilm genes in *Vibrio cholerae*. *Journal of Bacteriology*, 202(18).

Hsu, L. M., Vo, N. V., Kane, C. M., and Chamberlin, M. J. (2003). *In Vitro* Studies of Transcript Initiation by *Escherichia coli* RNA Polymerase. 1. RNA Chain Initiation, Abortive Initiation, and Promoter Escape at Three Bacteriophage Promoters †. *Biochemistry*, 42(13):3777–3786.

Hsueh, B. Y. and Waters, C. M. (2019). Combating Cholera. *F1000Research*, 8:589.

Hung, D. T., Zhu, J., Sturtevant, D., and Mekalanos, J. J. (2006). Bile acids stimulate biofilm formation in *Vibrio cholerae*. *Molecular Microbiology*, 59(1):193–201.

Huq, A., Yunus, M., Sohel, S. S., Bhuiya, A., Emch, M., Luby, S. P., Russek-Cohen, E., Nair, G. B., Sack, R. B., and Colwell, R. R. (2010). Simple Sari Cloth Filtration of Water Is Sustainable and Continues To Protect Villagers from Cholera in Matlab, Bangladesh. *mBio*, 1(1).

Hyakutake, A., Homma, M., Austin, M. J., Boin, M. A., Hase, C. C., and Kawagishi, I. (2005). Only One of the Five CheY Homologs in *Vibrio cholerae* Directly Switches Flagellar Rotation. *Journal of Bacteriology*, 187(24):8403–8410.

Jemielita, M., Wingreen, N. S., and Bassler, B. L. (2018). Quorum sensing controls *Vibrio cholerae* multicellular aggregate formation. *eLife*, 7:25.

Jenal, U., Reinders, A., and Lori, C. (2017). Cyclic di-GMP: second messenger extraordinaire. *Nature Reviews Microbiology*, 15(5):271–284.

Joelsson, A., Liu, Z., and Zhu, J. (2006). Genetic and Phenotypic Diversity of Quorum-Sensing Systems in Clinical and Environmental Isolates of *Vibrio cholerae*. *Infection and Immunity*, 74(2):1141–1147.

Kamruzzaman, M., Udden, S. M. N., Cameron, D. E., Calderwood, S. B., Nair, G. B., Mekalanos, J. J., and Faruque, S. M. (2010). Quorum-regulated biofilms enhance the development of conditionally viable, environmental *Vibrio cholerae*. *Proceedings of the National Academy of Sciences*, 107(4):1588–1593.

Karatan, E., Duncan, T. R., and Watnick, P. I. (2005). NspS, a Predicted Polyamine Sensor, Mediates Activation of *Vibrio cholerae* Biofilm Formation by Norspermidine. *Journal of Bacteriology*, 187(21):7434–7443.

Kariisa, A. T., Weeks, K., and Tamayo, R. (2016). The RNA Domain Vc1 Regulates Downstream Gene Expression in Response to Cyclic Diguanylate in *Vibrio cholerae*. *PLOS ONE*, 11(2):e0148478.

Kitts, G., Giglio, K. M., Zamorano-Sánchez, D., Park, J. H., Townsley, L., Cooley, R. B., Wucher, B. R., Klose, K. E., Nadell, C. D., Yildiz, F. H., and Sondermann, H. (2019). A Conserved Regulatory Circuit Controls Large Adhesins in *Vibrio cholerae*. *mBio*, 10(6).

Klose, K. E. and Mekalanos, J. J. (1998). Differential Regulation of Multiple Flagellins in *Vibrio cholerae*. *Journal of Bacteriology*, 180:14.

Kolb, A., Kotlarz, D., Kusanol, S., and Ishihamal, A. (1995). Selectivity of the *Escherichia coli* RNA polymerase Ec738 for overlapping promoters and ability to support CRP activation. *Nucleic Acids Research*, 23(5):819–826.

Komissarova, N., Becker, J., Solter, S., Kireeva, M., and Kashlev, M. (2002). Shortening of RNA:DNA Hybrid in the Elongation Complex of RNA Polymerase Is a Prerequisite for Transcription Termination. *Molecular Cell*, 10(5):1151–1162.

Krasteva, P. V., Fong, J. C. N., Shikuma, N. J., Beyham, S., Navarro, M. V. A. S., Yildiz, F. H., and Sondermann, H. (2010). *Vibrio cholerae* VpsT Regulates Matrix Production and Motility by Directly Sensing Cyclic di-GMP. *Science*, 327(5967):866–868.

Krebs, S. J. and Taylor, R. K. (2011). Nutrient-dependent, rapid transition of *Vibrio cholerae* to coccoid morphology and expression of the toxin co-regulated pilus in this form. *Microbiology*, 157(10):2942–2953.

Lamberte, L. E., Baniulyte, G., Singh, S. S., Stringer, A. M., Bonocora, R. P., Stracy, M., Kapanidis, A. N., Wade, J. T., and Grainger, D. C. (2017). Horizontally acquired

AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nature Microbiology*, 2.

Laviad-Shitrit, S., Izhaki, I., and Halpern, M. (2019). Accumulating evidence suggests that some waterbird species are potential vectors of *Vibrio cholerae*. *PLOS Pathogens*, 15(8):e1007814.

Lee, D. J., Minchin, S. D., and Busby, S. J. (2012). Activating Transcription in Bacteria. *Annual Review of Microbiology*, 66(1):125–152.

Lee, S. H., Butler, S. M., and Camilli, A. (2001). Selection for *in vivo* regulators of bacterial virulence. *Proceedings of the National Academy of Sciences*, 98(12):6889–6894.

Lim, B., Beyhan, S., Meir, J., and Yildiz, F. H. (2006). Cyclic-diGMP signal transduction systems in *Vibrio cholerae*: modulation of rugosity and biofilm formation. *Molecular Microbiology*, 60(2):331–348.

Lin, H.-Y., Bledsoe, P. J., and Stewart, V. (2007). Activation of *yeaR-yoaG* Operon Transcription by the Nitrate-Responsive Regulator NarL Is Independent of Oxygen-Responsive Regulator Fnr in *Escherichia coli* K-12. *J. BACTERIOL.*, 189(21):7539–7548.

Lodge, J., Fear, J., Busby, S., Gunasekaran, P., and Kamini, N. R. (1992). Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiology Letters*, 95(2-3):271–276.

Lutz, C., Erken, M., Noorian, P., Sun, S., and McDougald, D. (2013). Environmental reservoirs and mechanisms of persistence of *Vibrio cholerae*. *Frontiers in Microbiology*, 4.

Mandlik, A., Livny, J., Robins, W., Ritchie, J., Mekalanos, J., and Waldor, M. (2011). RNA-Seq-Based Monitoring of Infection-Linked Changes in *Vibrio cholerae* Gene Expression. *Cell Host and Microbe*, 10(2):165–174.

Manneh-Roussel, J., Haycocks, J. R. J., Magán, A., Perez-Soto, N., Voelz, K., Camilli, A., Krachler, A.-M., and Grainger, D. C. (2018). cAMP Receptor Protein Controls *Vibrio cholerae* Gene Expression in Response to Host Colonization. *mBio*, 9(4):15.

Maris, A. E., Kaczor-Grzeskowiak, M., Ma, Z., Kopka, M. L., Gunsalus, R. P., and Dickerson, R. E. (2005). Primary and Secondary Modes of DNA Recognition by the NarL Two-Component Response Regulator. *Biochemistry*, 44(44):14538–14552.

Mascher, T. (2013). Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. *Current Opinion in Microbiology*, 16(2):148–155.

Massie, J. P., Reynolds, E. L., Koestler, B. J., Cong, J.-P., Agostoni, M., and Waters, C. M. (2012). Quantification of high-specificity cyclic diguanylate signaling. *Proceedings of the National Academy of Sciences*, 109(31):12746–12751.

Matz, C., McDougald, D., Moreno, A. M., Yung, P. Y., Yildiz, F. H., and Kjelleberg, S. (2005). Biofilm formation and phenotypic variation enhance predation-driven persistence of *Vibrio cholerae*. *Proceedings of the National Academy of Sciences*, 102(46):16819–16824.

Meibom, K. L. (2005). Chitin Induces Natural Competence in *Vibrio cholerae*. *Science*, 310(5755):1824–1827.

Meibom, K. L., Li, X. B., Nielsen, A. T., Wu, C.-Y., Roseman, S., and Schoolnik, G. K. (2004). The *Vibrio cholerae* chitin utilization program. *Proceedings of the National Academy of Sciences*, 101(8):2524–2529.

Merrell, D. S., Butler, S. M., Qadri, F., Dolganov, N. A., Alam, A., Cohen, M. B., Calderwood, S. B., Schoolnik, G. K., and Camilli, A. (2002). Host-induced epidemic spread of the cholera bacterium. *Nature*, 417:4.

Miller, J. (1972). *Experiments in molecular genetics*. Bacterial genetics. Cold Spring Harbor Laboratory.

Miller, V. L., DiRita, V. J., and Mekalanos, J. J. (1989). Identification of *toxS*, a regulatory gene whose product enhances *toxR*-mediated activation of the cholera toxin promoter. *Journal of Bacteriology*, 171(3):1288–1293.

Mitchell, J. E. (2003). Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Research*, 31(16):4689–4695.

Murphy, R. A. and Boyd, E. F. (2008). Three Pathogenicity Islands of *Vibrio cholerae* Can Excise from the Chromosome and Form Circular Intermediates. *Journal of Bacteriology*, 190(2):636–647.

Murphy, S. G., Alvarez, L., Adams, M. C., Liu, S., Chappie, J. S., Cava, F., and Dörr, T. (2019). Endopeptidase Regulation as a Novel Function of the Zur- Dependent Zinc Starvation Response. 10(1):15.

Murphy, S. G., Johnson, B. A., Ledoux, C. M., and Dörr, T. (2021). *Vibrio cholerae* 's mysterious Seventh Pandemic island (VSP-II) encodes novel Zur-regulated zinc starvation genes involved in chemotaxis and autoaggregation. *bioRxiv (preprint)*.

Mutreja, A., Kim, D. W., Thomson, N. R., Connor, T. R., Lee, J. H., Kariuki, S., Croucher, N. J., Choi, S. Y., Harris, S. R., Lebens, M., Niyogi, S. K., Kim, E. J., Ramamurthy, T., Chun, J., Wood, J. L. N., Clemens, J. D., Czerkinsky, C., Nair, G. B., Holmgren, J., Parkhill, J., and Dougan, G. (2011). Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*, 477(7365):462–465.

Naser, I. B., Hoque, M. M., Faruque, S. N., Kamruzzaman, M., Yamasaki, S., and Faruque, S. M. (2019). *Vibrio cholerae* strains with inactivated *cqsS* gene overproduce autoinducer-2 which enhances resuscitation of dormant environmental *V. cholerae*. *PLOS ONE*, 14(10):e0223226.

Nguyen, T. H., Pham, T. D., Higa, N., Iwashita, H., Takemura, T., Ohnishi, M., Morita, K., and Yamashiro, T. (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: VSP-II, GIs and *AttL* of *V. cholerae* strains. *Microbiology and Immunology*, 62(3):150–157.

Nielsen, A. T., Dolganov, N. A., Otto, G., Miller, M. C., Wu, C. Y., and Schoolnik, G. K. (2006). RpoS Controls the *Vibrio cholerae* Mucosal Escape Response. *PLoS Pathogens*, 2(10).

Obenchain, V., Lawrence, M., Carey, V., Gogarten, S., Shannon, P., and Morgan, M. (2014). Variantannotation: a bioconductor package for exploration and annotation of genetic variants. *Bioinformatics*, 30(14):2076–2078.

Oehler, S., Eismann, E., Krämer, H., and Müller-Hill, B. (1990). The three operators of the *lac* operon cooperate in repression. *The EMBO Journal*, 9(4):973–979.

O'Shea, Y. A. (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(12):4053–4063.

Page, L., Griffiths, L., and Cole, J. A. (1990). Different physiological roles of two independent pathways for nitrite reduction to ammonia by enteric bacteria. *Archives of microbiology*, 154(4):349–354.

Paget, M. (2015). Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules*, 5(3):1245–1265.

Paget, M. S. and Helmann, J. D. (2003). The sigma70 family of sigma factors. *Genome Biology*, 4(203).

Panda, D., Dasgupta, U., and Das, J. (1991). Transformation of *Vibrio cholerae* by plasmid DNA. *Gene*.

Papenfort, K. and Bassler, B. L. (2016). Quorum sensing signal–response systems in Gram-negative bacteria. *Nature Reviews Microbiology*, 14(9):576–588.

Papenfort, K., Förstner, K. U., Cong, J.-P., Sharma, C. M., and Bassler, B. L. (2015). Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proceedings of the National Academy of Sciences*, 112(7).

Papenfort, K., Silpe, J. E., Schramma, K. R., Cong, J.-P., Seyedsayamdost, M. R., and Bassler, B. L. (2017). A *Vibrio cholerae* autoinducer–receptor pair that controls biofilm formation. *Nature Chemical Biology*, 13(5):551–557.

Paul, B. J., Berkmen, M. B., and Gourse, R. L. (2005). DksA potentiates direct activation of amino acid promoters by ppGpp. *Proceedings of the National Academy of Sciences*, 102(22):7823–7828.

Perez-Soto, N., Moule, L., Crisan, D. N., Insua, I., Taylor-Smith, L. M., Voelz, K., Fernandez-Trillo, F., and Krachler, A. M. (2017). Engineering microbial physiology with synthetic polymers: cationic polymers induce biofilm formation in *Vibrio cholerae* and downregulate the expression of virulence genes. *Chemical Science*, 8(8):5291–5298.

Peschek, N., Herzog, R., Singh, P. K., Sprenger, M., Meyer, F., Fröhlich, K. S., Schröger, L., Bramkamp, M., Drescher, K., and Papenfort, K. (2020). RNA-mediated control of cell shape modulates antibiotic resistance in *Vibrio cholerae*. *Nature Communications*, 11(1).

Peters, J. M., Mooney, R., Grass, J., Jessen, E., Tran, F., and Landick, R. (2012). Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes and Development*, 26(23):2621–2633.

Peters, J. M., Vangeloff, A. D., and Landick, R. (2011). Bacterial Transcription Terminators: The RNA 3′-End Chronicles. *Journal of Molecular Biology*, 412(5):793–813.

Prouty, M. G., Correa, N. E., and Klose, K. E. (2008). The novel s54- and s28-dependent flagellar gene transcription hierarchy of *Vibrio cholerae*. *Molecular Microbiology*, 39(6):1595–1609.

Pursley, B. R., Maiden, M. M., Hsieh, M.-L., Fernandez, N. L., Severin, G. B., and Waters, C. M. (2018). Cyclic di-GMP Regulates TfoY in *Vibrio cholerae* To Control Motility by both Transcriptional and Posttranscriptional Mechanisms. *Journal of Bacteriology*, 200(7).

Rawlings, T. K., Ruiz, G. M., and Colwell, R. R. (2007). Association of *Vibrio cholerae* O1 El Tor and O139 Bengal with the Copepods Acartia tonsa and Eurytemora affinis. *Applied and Environmental Microbiology*, 73(24):7926–7933.

Richardson, J. P. (1982). Activation of rho protein ATPase requires simultaneous interaction at two kinds of nucleic acid-binding sites. 257(10):7.

Robert, X. and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42(W1):W320–W324.

Roberts, J. W. (2019). Mechanisms of Bacterial Transcription Termination. *Journal of Molecular Biology*, 431(20):4030–4039.

Roberts, J. W., Yarnell, W., Bartlett, E., Guo, J., Marr, M., Ko, D. C., Sun, H., and Roberts, C. W. (1998). Antitermination by bacteriophage lambda Q protein. *Cold Spring Harbor symposia on quantitative biology*, 63:319–326.

Roelofs, K. G., Jones, C. J., Helman, S. R., Shang, X., Orr, M. W., Goodson, J. R., Galperin, M. Y., Yildiz, F. H., and Lee, V. T. (2015). Systematic Identification of Cyclic-di-GMP Binding Proteins in *Vibrio cholerae* Reveals a Novel Class of Cyclic-di-GMP-Binding ATPases Associated with Type II Secretion Systems. *PLOS Pathogens*, 11(10).

Ruanto, P., Chismon, D. L., Hothersall, J., Godfrey, R. E., Lee, D. J., Busby, S. J. W., and Browning, D. F. (2020). Activation by NarL at the *Escherichia coli ogt* promoter. *Biochemical Journal*, 477(15):2807–2820.

Sack, D. A., Sack, R. B., Nair, G. B., and Siddique, A. K. (2004). Cholera. *The Lancet*, 363:223–233.

Safa, A., Nair, G. B., and Kong, R. Y. (2010). Evolution of new variants of *Vibrio cholerae* O1. *Trends in Microbiology*, 18(1):46–54.

Sarenko, O., Klauck, G., Wilke, F. M., Pfiffer, V., Richter, A. M., Herbst, S., Kaever, V., and Hengge, R. (2017). More than Enzymes That Make or Break Cyclic Di-GMP—Local Signaling in the Interactome of GGDEF/EAL Domain Proteins of *Escherichia coli*. *mBio*, 8(5).

Satchell, K. J. F., Jones, C. J., Wong, J., Queen, J., Agarwal, S., and Yildiz, F. H. (2016). Phenotypic Analysis Reveals that the 2010 Haiti Cholera Epidemic Is Linked to a Hypervirulent Strain. *Infection and Immunity*, 84(9):2473–2481.

Saxena, S., Myka, K. K., Washburn, R., Costantino, N., Court, D. L., and Gottesman, M. E. (2018). *Escherichia coli* transcription factor NusG binds to 70S ribosomes: NusG binding to ribosomes couples transcription with translation. *Molecular Microbiology*, 108(5):495–504.

Schmidt, C. and Chamberlin, M. J. (1987). nusA Protein of *Escherichia coli* is an Efficient Transcription Termination Factor for Certain Terminator Sites. 195(4):809–818.

Senderovich, Y., Izhaki, I., and Halpern, M. (2010). Fish as Reservoirs and Vectors of *Vibrio cholerae*. *PLoS ONE*, 5(1).

Seper, A., Fengler, V. H. I., Roier, S., Wolinski, H., Kohlwein, S. D., Bishop, A. L., Camilli, A., Reidl, J., and Schild, S. (2011). Extracellular nucleases and extracellular DNA play important roles in *Vibrio cholerae* biofilm formation: Nucleases of *Vibrio cholerae*. *Molecular Microbiology*, 82(4):1015–1037.

Sharma, P., Haycocks, J. R. J., Middlemiss, A. D., Kettles, R. A., Sellars, L. E., Ricci, V., Piddock, L. J. V., and Grainger, D. C. (2017). The multiple antibiotic resistance operon of enteric bacteria controls DNA repair and outer membrane integrity. *Nature Communications*, 8(1).

Shikuma, N. J., Fong, J. C. N., and Yildiz, F. H. (2012). Cellular Levels and Binding of c-di-GMP Control Subcellular Localization and Activity of the *Vibrio cholerae* Transcriptional Regulator VpsT. *PLoS Pathogens*, 8(5):e1002719.

Shin, J.-H., Lanz, M., Smolka, M. B., and Dörr, T. (2020). Characterization of an anti-FLAG antibody binding protein in *V. cholerae*. *Biochemical and Biophysical Research Communications*, 528(3):493–498.

Singh, S. S. and Grainger, D. C. (2013). H-NS Can Facilitate Specific DNA-binding by RNA Polymerase in AT-rich Gene Regulatory Regions. *PLoS Genetics*, 9(6).

Skorupski, K. and Taylor, R. K. (1996). Positive selection vectors for allelic exchange. *Gene*, 169(1):47–52.

Squire, D., Xu, M., Cole, J., Busby, S., and Browning, D. (2009). Competition between NarL-dependent activation and Fis-dependent repression controls expression from the *Escherichia coli yeaR* and *ogt* promoters. *Biochemical Journal*, 420(2):249–257.

Srivastava, D., Harris, R. C., and Waters, C. M. (2011). Integration of Cyclic di-GMP and Quorum Sensing in the Control of *vpsT* and *aphA* in *Vibrio cholerae*. *Journal of Bacteriology*, 193(22):6331–6341.

Srivastava, D. and Waters, C. M. (2012). A Tangled Web: Regulatory Connections between Quorum Sensing and Cyclic Di-GMP. *Journal of Bacteriology*, 194(17):4485–4493.

Sudarsan, N., Lee, E. R., Weinberg, Z., Moy, R. H., Kim, J. N., Link, K. H., and Breaker, R. R. (2008). Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP. *Science*, 321(5887):411–413.

Syed, K. A., Beyhan, S., Correa, N., Queen, J., Liu, J., Peng, F., Satchell, K. J. F., Yildiz, F., and Klose, K. E. (2009). The *Vibrio cholerae* Flagellar Regulatory Hierarchy Controls Expression of Virulence Factors. *Journal of Bacteriology*, 191(21):6555–6570.

Teschler, J. K., Zamorano-Sánchez, D., Utada, A. S., Warner, C. J. A., Wong, G. C. L., Linington, R. G., and Yildiz, F. H. (2015). Living in the matrix: assembly and control of *Vibrio cholerae* biofilms. *Nature Reviews Microbiology*, 13(5):255–268.

Tischler, A. D. and Camilli, A. (2004). Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation. *Molecular Microbiology*, 53(3):857–869.

Townsley, L. and Yildiz, F. H. (2015). Temperature affects c-di-GMP signalling and biofilm formation in *Vibrio cholerae*: Environmental regulation of c-di-GMP signalling. *Environmental Microbiology*, 17(11):4290–4305.

Treviño-Quintanilla, L., Freyre-González, J., and Martínez-Flores, I. (2013). Anti-Sigma Factors in *E. coli*: Common Regulatory Mechanisms Controlling Sigma Factors Availability. *Current Genomics*, 14(6):378–387.

Trotochaud, A. E. and Wassarman, K. M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. *Nature Structural and Molecular Biology*, 12(4):7.

Utada, A. S., Bennett, R. R., Fong, J. C. N., Gibiansky, M. L., Yildiz, F. H., Golestanian, R., and Wong, G. C. L. (2014). *Vibrio cholerae* use pili and flagella synergistically to effect motility switching and conditional surface attachment. *Nature Communications*, 5(1).

Valiente, E., Davies, C., Mills, D. C., Getino, M., Ritchie, J. M., and Wren, B. W. (2018). *Vibrio cholerae* accessory colonisation factor AcfC: a chemotactic protein with a role in hyperinfectivity. *Scientific Reports*, 8(1).

Van der Henst, C., Scrignari, T., Maclachlan, C., and Blokesch, M. (2016). An intracellular replication niche for *Vibrio cholerae* in the amoeba *Acanthamoeba castellanii*. *The ISME Journal*, 10(4):897–910.

Visweswariah, S. S. and Busby, S. J. (2015). Evolution of bacterial transcription factors: how proteins take on new tasks, but do not always stop doing the old ones. *Trends in Microbiology*, 23(8):463–467.

Vora, G. J., Meador, C. E., Bird, M. M., Bopp, C. A., Andreadis, J. D., and Stenger, D. A. (2005). Microarray-based detection of genetic heterogeneity, antimicrobial resistance, and the viable but nonculturable state in human pathogenic *Vibrio spp. Proceedings of the National Academy of Sciences*, 102(52):19109–19114.

Wade, J. T. and Grainger, D. C. (2014). Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nature reviews. Microbiology*, 12(9):647.

Wade, J. T., Struhl, K., Busby, S. J. W., and Grainger, D. C. (2007). Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization. *Molecular Microbiology*, 65(1):21–26.

Wang, H., Ayala, J. C., Benitez, J. A., and Silva, A. J. (2014). The LuxR-Type Regulator VpsT Negatively Controls the Transcription of *rpoS*, Encoding the General Stress Response Regulator, in *Vibrio cholerae* Biofilms. *Journal of Bacteriology*, 196(5):1020–1030.

Wang, H., Ayala, J. C., Silva, A. J., and Benitez, J. A. (2012). The Histone-Like Nucleoid Structuring Protein (H-NS) Is a Repressor of *Vibrio cholerae* Exopolysaccharide Biosynthesis (*vps*) Genes. *Applied and Environmental Microbiology*, 78(7):2482–2488.

Warman, E. A., Forrest, D., Guest, T., Haycocks, J. R. J., Wade, J. T., and Grainger, D. C. (2021). Widespread divergent transcription from bacterial and archaeal promoters is a consequence of DNA-sequence symmetry. *Nature Microbiology*.

Warman, E. A., Singh, S. S., Gubieda, A. G., and Grainger, D. C. (2020). A non-canonical promoter element drives spurious transcription of horizontally acquired bacterial genes. *Nucleic Acids Research*, 48(9):4891–4901.

Waters, C. M., Lu, W., Rabinowitz, J. D., and Bassler, B. L. (2008). Quorum Sensing Controls Biofilm Formation in *Vibrio cholerae* through Modulation of Cyclic Di-GMP Levels and Repression of *vpsT*. *Journal of Bacteriology*, 190(7):2527–2536.

Waters, L. S. and Storz, G. (2009). Regulatory RNAs in Bacteria. *Cell*, page 14.

Watnick, P. I., Lauriano, C. M., Klose, K. E., Croal, L., and Kolter, R. (2001). The absence of a flagellum leads to altered colony morphology, biofilm development and virulence in *Vibrio cholerae* O139. *Molecular Microbiology*, 39(2):223–235.

Weill, F.-X., Domman, D., Njamkepo, E., Tarr, C., Rauzier, J., Fawal, N., Keddy, K. H., Salje, H., Moore, S., Mukhopadhyay, A. K., Bercion, R., Luquero, F. J., Ngandjio, A., Dosso, M., Monakhova, E., Garin, B., Bouchier, C., Pazzani, C., Mutreja, A., Grunow, R., Sidikou, F., Bonte, L., Breurec, S., Damian, M., Njanpop-Lafourcade, B.-M., Sapriel, G., Page, A.-L., Hamze, M., Henkens, M., Chowdhury, G., Mengel, M., Koeck, J.-L., Fournier, J.-M., Dougan, G., Grimont, P. A. D., Parkhill, J., Holt, K. E., Piarroux, R., Ramamurthy, T., Quilici, M.-L., and Thomson, N. R. (2017). Genomic history of the seventh pandemic of cholera in Africa. *Science*, 358(6364):785–789.

WHO (2018). Cholera 2017.pdf. *Weekly Epidemiological Record*, 38:489–500.

Wierzba, T. F. (2019). Oral cholera vaccines and their impact on the global burden of disease. *Human Vaccines and Immunotherapeutics*, 15(6):1294–1301.

Wigneshweraraj, S., Bose, D., Burrows, P. C., Joly, N., Schumacher, J., Rappas, M., Pape, T., Zhang, X., Stockley, P., Severinov, K., and Buck, M. (2008). Modus operandi of the bacterial RNA polymerase containing the σ ^{54} promoter-specificity factor. *Molecular Microbiology*, 68(3):538–546.

Winkelman, J. T. and Gourse, R. L. (2017). Open complex DNA scrunching: A key to transcription start site selection and promoter escape. *BioEssays*, 39(2):1600193.

Wu, B., Liang, W., Yan, M., Li, J., Zhao, H., Cui, L., Zhu, F., Zhu, J., and Kan, B. (2020). Quorum sensing regulation confronts the development of a viable but non□culturable state in *Vibrio cholerae*. *Environmental Microbiology*, 22(10):4314–4322.

Xu, Q., Dziejman, M., and Mekalanos, J. J. (2003). Determination of the transcriptome of *Vibrio cholerae* during intraintestinal growth and midexponential phase in vitro. *Proceedings of the National Academy of Sciences*, 100(3):1286–1291.

Xu, T., Cao, H., Zhu, W., Wang, M., Du, Y., Yin, Z., Chen, M., Liu, Y., Yang, B., and Liu, B. (2018). RNA-seq-based monitoring of gene expression changes of viable but non-culturable state of *Vibrio cholerae* induced by cold seawater: The VBNC state of *Vibrio cholerae*. *Environmental Microbiology Reports*, 10(5):594–604.

Yaniv, M. (2011). The 50th Anniversary of the Publication of the Operon Theory in the Journal of Molecular Biology: Past, Present and Future. *Journal of Molecular Biology*, 409(1):1–6.

Yildiz, F. H., Dolganov, N. A., and Schoolnik, G. K. (2001). VpsR, a Member of the Response Regulators of the Two-Component Regulatory Systems, Is Required for Expression of vps Biosynthesis Genes and EPSETr-Associated Phenotypes in *Vibrio cholerae* O1 El Tor. *Journal of Bacteriology*, 183(5):1716–1726.

Yildiz, F. H., Liu, X. S., Heydorn, A., and Schoolnik, G. K. (2004). Molecular analysis of rugosity in a *Vibrio cholerae* O1 El Tor phase variant: Regulation of rugosity. *Molecular Microbiology*, 53(2):497–515.

Yildiz, F. H. and Schoolnik, G. K. (1998). Role of *rpoS* in Stress Survival and Virulence of *Vibrio cholerae*. *Journal of Bacteriology*, 180(4):773–784.

Yildiz, F. H. and Schoolnik, G. K. (1999). *Vibrio cholerae* O1 El Tor: identification of a gene cluster required for the rugose colony type, exopolysaccharide production,

chlorine resistance, and biofilm formation. *Proceedings of the National Academy of Sciences*, 96(7):4028–4033.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). Chipseeker: an r/bioconductor package for chip peak annotation, comparison and visualization. *Bioinformatics*, 31(14):2382–2383.

Zago, V., Zambon, M., Civettini, M., Zaltum, O., and Manfrin, A. (2017). Virulence-associated factors in *Vibrio cholerae* non-O1/non-O139 and *V. mimicus strains* isolated in ornamental fish species. *Journal of Fish Diseases*.

Zamorano-Sánchez, D., Fong, J. C. N., Kilic, S., Erill, I., and Yildiz, F. H. (2015). Identification and Characterization of VpsR and VpsT Binding Sites in *Vibrio cholerae*. *Journal of Bacteriology*, 197(7):1221–1235.

Zamorano-Sánchez, D., Xian, W., Lee, C. K., Salinas, M., Thongsomboon, W., Cegelski, L., Wong, G. C. L., and Yildiz, F. H. (2019). Functional Specialization in *Vibrio cholerae* Diguanylate Cyclases: Distinct Modes of Motility Suppression and c-di-GMP Production. *mBio*, 10(2).

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9):R137.

Zhou, H., Zheng, C., Su, J., Chen, B., Fu, Y., Xie, Y., Tang, Q., Chou, S.-H., and He, J. (2016). Characterization of a natural triple-tandem c-di-GMP riboswitch and application of the riboswitch-based dual-fluorescence reporter. *Scientific Reports*, 6(1).

Zuo, Y., Wang, Y., and Steitz, T. (2013). The Mechanism of *E. coli* RNA Polymerase Regulation by ppGpp Is Suggested by the Structure of their Complex. *Molecular Cell*, 50(3):430–436.