IDENTIFICATION AND STRUCTURE-FUNCTION ANALYSES OF BACTERIAL C-DI-GMP RECEPTORS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Petya Violinova Krasteva

January 2011

IDENTIFICATION AND STRUCTURE-FUNCTION ANALYSES OF BACTERIAL C-DI-GMP RECEPTORS

Petya Violinova Krasteva, Ph. D.

Cornell University 2011

In recent years a novel, nucleotide-based small molecule, c-di-GMP, has emerged in the spotlight of scientific investigation as a second messenger unique to the bacterial world. The discovery that its intracellular levels strictly regulate cell adhesion and persistence of bacterial biofilms on one hand, and motility and virulence of planktonic cells on the other, has related this RNA molecule to a variety of disease states including both chronic and acute bacterial infections.

Interestingly, intracellular signaling mechanisms involving c-di-GMP appear to be spatially restricted, yet cellular targets for this nucleotide remain mostly unknown. Here we set out to identify and provide comprehensive structure-function analyses of putative or known c-di-GMP receptors. By using structural biology methods we would first determine the atomic resolution structures and conformational states of appropriate targets and then use these molecular blueprints to guide our research into their mechanistic role in the big picture of intracellular signaling networks.

We identified VpsT of *V.cholerae* as a novel c-di-GMP receptor and solved the crystal structures of the nucleotide-free and c-di-GMP-bound states. Our studies identified two biologically relevant dimerization interfaces and the potential formation of higher order oligomeric species assembling upon nucleotide recognition. VpsT defines a novel class of response regulators that utilize a characteristic structural feature to dimerize upon a signaling input regardless of concurrent phosphorylation

events. The relative orientation of the DNA-binding domains of VpsT favors a model in which gene regulation is likely accompanied by major changes in DNA architecture. We showed that VpsT is a master regulator of biofilm formation, integrating c-di-GMP signaling events to inversely control exopolysaccharide production and flagellar motility.

In a separate study, we provide a complete mechanistic analysis of the structure and function of LapD, a transmembrane c-di-GMP receptor in *P. fluorescens* which directly translates intracellular c-di-GMP levels in a signal for biofilm dispersal or initiation. We solved three novel crystal structures capturing distinct intermediates in the inside-out signaling process. Most importantly, our structural and functional analyses helped us identify homologous systems in a number of free-living and pathogenic species, likely controlling biofilm formation or toxin expression in a largely similar manner.

BIOGRAPHICAL SKETCH

The author was born in the summer of 1981 in the vibrant European capital of Bulgaria, Sofia City. She studied as a full-time student at *First English Language School*, while completing a degree in French and French Literature from the city *French Language School Alphonse de Lamartine* at the same time. In the year of 2000 she was accepted in the Department of Biology of *Sofia University St. Kliment Ohridski* to embark on a Bachelor's degree in Molecular Biology. After graduating with honors, she was awarded a government funded tuition fellowship to continue her studies in a Master's degree in Biochemistry, where she studied inflammatory processes in atherogenesis under the guidance of Dr. Margarita Apostolova at the *Bulgarian Academy of Sciences*. In 2006, she came to *Cornell University* as a graduate student in the field of Biochemistry, Molecular, and Cell Biology, where she joined the research group of Dr. Holger Sondermann to pursue her doctoral dissertation. During this fruitful experience she devoted her efforts to the structural and functional characterization of bacterial c-di-GMP sensor proteins and their role as molecular effectors in biofilm formation and pathogenicity of free-living and disease-causing microorganisms.

In loving memory of Stefan Babakolev

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor Dr. Holger Sondermann for his continuous encouragement, guidance, and support, while allowing me the freedom to work independently and to my very own pace. During my three years of research in his lab, Holger has been always approachable and open-minded and has helped me adapt my research focus and practical approaches in a truly constructive manner. I have been fortunate to learn a number of novel techniques directly from him and in that I am grateful for having such a great teacher. Holger has allowed me to pursue and develop my personal scientific interests and has encouraged me to establish a vast network of contacts and collaborations both on- and off-campus. I have thus been able to use and get valuable training on methods and instrumentation not immediately available in the lab, while collaborators from various universities have provided us with data to help us confirm our hypotheses and guide our studies.

I would also like to thank my other committee members, Dr. Anthony Hay and Dr. Brian Crane, for their valuable input and continuous support. It has been my greatest pleasure to discuss my project with them and I deeply appreciate their suggestions and practical help with borrowed reagents or shared equipment.

I am extremely grateful to everyone else who has helped me in one way or another advance in my research. First, it is important for me to acknowledge the *Thanks to Scandinavia* scholarship fund, whose generous award paved my way to Cornell in the very first place. Special thanks to Marcos V.A.S. Navarro, a former postdoctoral fellow in our lab, from whom I learned a lot of the practical aspects of protein crystallography and who accompanied me during many of my synchrotron trips to make each one of them a fun and exciting experience. In addition, it has been

great pleasure to work with him and with my colleague Debashree Chatterjee on the LapD protein project and I sincerely thank them for their respective contributions. I would like to acknowledge my former labmate and good friend Nabanita De for practical help and initial guidance on the biofilm project. I am also grateful to Yee Ling "June" Lam, an exceptional undergraduate intern, and to Di Luo, a rotation graduate student, for cloning and initial screening on some of the proteins described in this study. Many thanks to other Sondermann lab members like Shih Lin "Lynda" Goh and Laura J. Byrnes, as well as to Collins lab members Fabio Rinaldi and Duane Hoch for making my stay in the lab a truly enjoyable experience. Finally, I would like to thank all the staff of the Molecular Medicine and MBG Departments including Debbie Crane, Cindy Westmiller, Greg Mitchell, Blake Werner, Diane Colf, and Volker Vogt whose continuous help ensured that everything ran fine on a day-to-day basis.

This work would not have been possible without the outstanding contributions from our various collaborators. For this I am deeply grateful to Fitnat Yildiz, Jiunn Fong, and Nicholas Shikuma at UCSC for their continuous contribution to the VpsT project, as well as to George O'Toole and Peter Newell at Dartmouth for work on the LapD functional characterization. I also want to thank all the people who have provided access to data collection, such as Stevan Hubbard and Luis Cunha at NYU for access to multi-angle light scattering, Bill Horne and Johanna Holm at Cornell for access to isothermal titration calorimetry, as well as all the scientists at MacCHESS, APS and NSLS for maintenance and help during synchrotron data collection. I would also like to acknowledge Bret Judson, Joshua Filter and Christopher "Kit" Umbach at Cornell for training me on electron microscopy, DNA footprinting, and atomic force microscopy, respectively.

I would also like to thank all the great teachers I have had the privilege to learn from over the years. This includes Cornell professors Linda Nicholson, Tim Huffaker,

John Lis, Mariana Wolfner and Mary K. Redmond among others, as well as all the outstanding educators at Sofia University and First English Language School back in Bulgaria. It is a pleasure for me to acknowledge my rotation advisors Tom Fox and Paul Soloway, as well as my temporary colleagues and dear friends Christine Butler and Krista Pia Kauppinen, for letting me work on their exciting projects during my first months at Cornell University. In addition, I thank Tom for participating in my A-exam committee and for his continuous interest and support.

Finally, I would like to thank my parents Dobroudjanka and Violin Krastevi, my brother Alexander Krastev, and my partner Nicolàs Reyes for their unconditional love and support. I am also grateful for having some of the most wonderful people I've met as my dear friends, and would like to thank them for always being there for me regardless of how physically far life has, or will, split us apart.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1. C-di-GMP as a bacterial second messenger

Cyclic di-GMP is a monocyclic RNA dinucleotide that functions as a global regulatory molecule found uniquely in bacteria. Approximately twenty years after its discovery by Benziman and colleagues as an allosteric activator of *Gluconacetobacter xylinus* cellulose synthase, c-di-GMP has gained appreciation as a universal second messenger triggering a plethora of physiological responses (*1-4*). Examples include cell differentiation, changes in motility and surface adhesiveness, secretion of extracellular polysaccharides and proteinaceous fimbriae, host cell cytotoxicity, and virulence gene expression (*4-7*). Signaling cascades employing this small molecule show evidence for multilayer impact which includes control at the transcriptional, translational and posttranslational levels. Proteins involved in c-di-GMP mediated signal transduction are often characterized by multi-domain architecture with such modularity to allow for a variety of regulatory inputs and/or signal ramifications (*8, 9*). This complexity is in stark contrast with canonical two-component transduction systems, where upon signal generation a sensor histidine kinase phosphorylates its cognate response regulator to alter the expression of a limited number of genes (*10*).

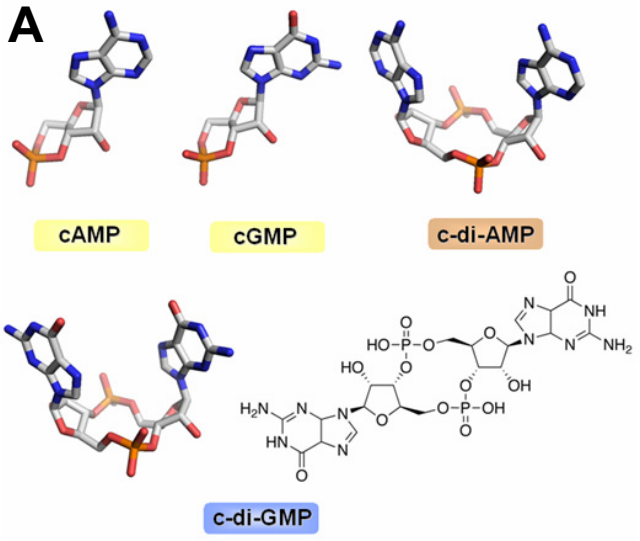
C-di-GMP is generated from two GTP molecules by GGDEF domain-containing diguanylate cyclases, whereas phosphodiesterases containing either EAL or HD-GYP protein domains provide selective signal degradation (*11-15*). C-di-GMP turnover domains are found to be among the most abundant protein domains encoded in bacterial genomes. Interestingly, such abundance correlates with the ability of

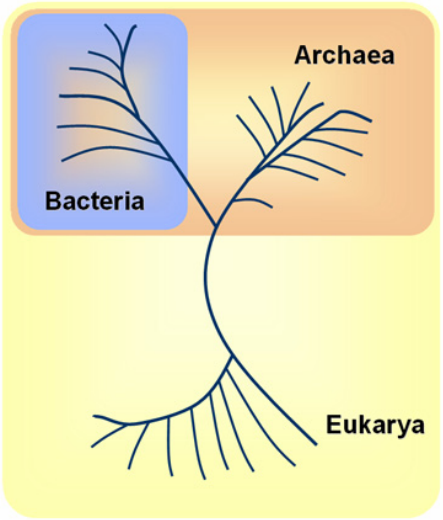
microbes to adapt to different ecological niches with opportunistic pathogens showing the largest number of c-di-GMP metabolizing enzymes per genome (*9, 16*).

In general, increased intracellular c-di-GMP levels resulting from higher diguanylate cyclase activity lead to enhanced biofilm formation and inhibit flagellar and pilus-mediated motility (*4, 18*). Conversely, low levels of the nucleotide associated with active phosphodiesterase catalysis suppress the maintenance of extracellular adhesins and promote biofilm dispersion and bacterial virulence (*19, 20*). Although such an overall effect has been demonstrated in several overexpression studies, redundancy between different diguanylate cyclases or phosphodiesterases in a given genome is only apparent (*12, 21*). Oftentimes proteins with similar domain architecture or enzymatic activity trigger distinct physiological responses (*11, 22*). This is particularly surprising if one assumes that, as a small hydrophilic molecule, c-di-GMP is freely diffusible in the cell. A number of studies argue instead that once generated, c-di-GMP is a sequestered, rather than general, diffusive signal (*2, 4*). Measurements of the intracellular levels of c-di-GMP in several bacterial species indicate concentrations in the micromolar range or lower, without taking into account probable local fluctuations. Based on the fact that most identified c-di-GMP receptors and phosphodiesterases have lower affinity constants for the nucleotide, it has been hypothesized that cellular c-di-GMP exists primarily in a protein bound form (*2, 12*).

Figure 1.1: Cyclic di-GMP as a signaling molecule.
(A) Phylogenetic distribution of prevalent RNA-based second messengers: Signal transduction networks that rely on the monocyclic nucleotide second messengers cAMP and cGMP modulate a diverse range of cellular responses in all three domains of life. Recent work has identified two novel RNA dinucleotides that function as intracellular signal amplifiers: c-di-AMP utilized by both domains of prokaryotes, and c-di-GMP found uniquely in bacteria (*17*).
(B) Cyclic di-GMP turnover domains in model opportunistic pathogens: *Vibrio cholerae*, causative agent of the acute diarrheal disease cholera, and *Pseudomonas aeruginosa*, causing persistent infections and lethality in patients with cystic fibrosis.
(C) Cyclic di-GMP signal generation, recognition and degradation: While protein domains catalyzing c-di-GMP synthesis and hydrolysis have been fairly well characterized, nucleotide sensors decoding the signal to inversely regulate bacterial quorum sensing and biofilm formation are mostly unknown.

A

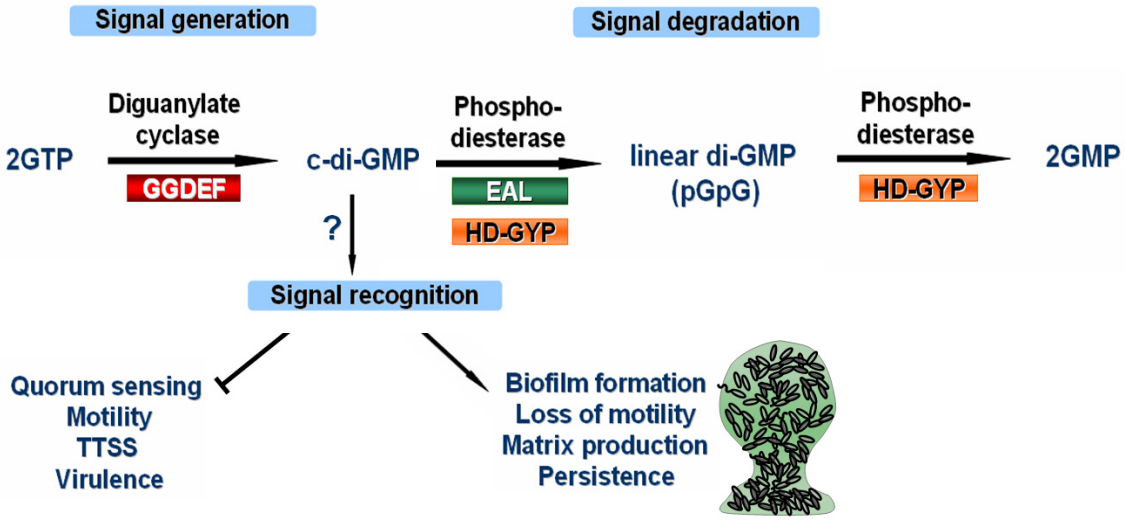




B

Model organism	Mode of infection	GGDEF domains	EAL domains	HD-GYP domains
Vibrio cholerae	Acute (Diarrheal Cholera)	41	22	9
Pseudomonas aeruginosa	Chronic (Cystic Fibrosis)	33	21	3

C



1.2. Bacterial cyclic-di-GMP receptors

Whether functioning as intermediaries in nucleotide signal relay or as final effectors in signaling cascades, bacterial c-di-GMP receptors remain largely unknown. Initial *in silico* prediction later corroborated by experimental evidence identified PilZ domains as the first known c-di-GMP targets (*23-28*). The domains are expressed either alone or as fusions with other modules, including but not restricted to EAL, GGDEF, HD-GYP, PAS and helix-turn-helix motifs (*23*). A PilZ domain was also identified as the c-di-GMP binding module in *G. xylinus* cellulose synthase, where c-di-GMP regulator function was first reported (*23*). Phylogenetic and structural analyses showed that PilZ domains have low sequence conservation apart from a few interspersed residues responsible for nucleotide docking (*25*). Most importantly, PilZ domains could not be the only c-di-GMP targets as some species utilizing c-di-GMP mediated signaling do not encode for PilZ domain-containing proteins (*23*).

Another c-di-GMP binding motif was identified on some diguanylate cyclases, where binding of the nucleotide to the so-called I-site leads to inhibition of enzymatic activity (*29-31*). A similar I-site motif (RxxD) is responsible for c-di-GMP binding in two other c-di-GMP receptors, PleD of *P. aeruginosa* and CdgG of *V. cholerae*, likely a part of degenerate diguanylate cyclase domains based on sequence alignment and secondary structure prediction (*32, 33*). Degenerate c-di-GMP specific phospho-diesterase domains have also been reported to function as sensors for the nucleotide, as in the case with the *P. aeruginosa* FimX protein where the EAL domain module functions as a high-affinity nucleotide sensor to regulate biofilm formation and pilus-mediated motility in a yet unknown mechanism (*34*). This, together with the many catalytically inactive GGDEF, EAL and HD-GYP domains identified, raises the question to what extent divergent c-di-GMP turnover domains have preserved signal transduction function based on nucleotide-dependent protein-protein interactions.

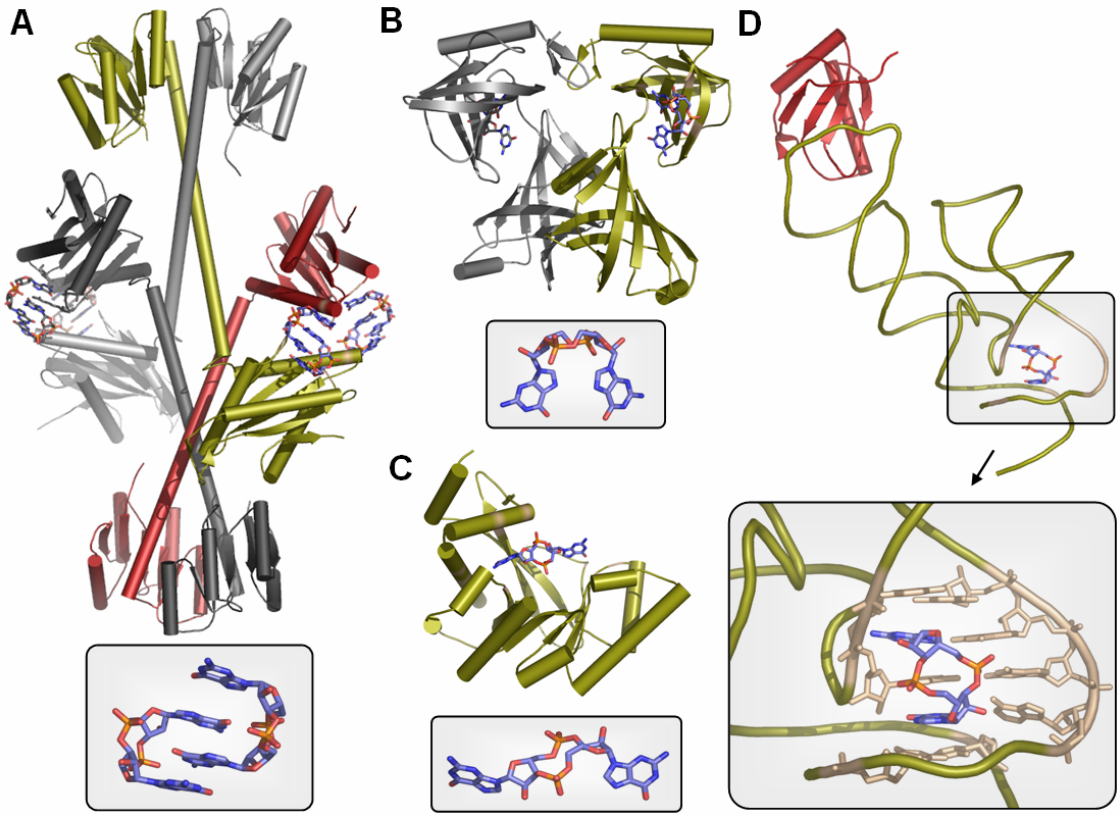


Figure 1.2: Diversity of cyclic di-GMP sensors and adopted nucleotide conformations.
(A) Crystal structure of the diguanylate cyclase WspR, regulating biofilm formation in *Pseudomonas aeruginosa* (pdb entry: 3BRE) (*30*). C-di-GMP binds to an allosteric inhibitory "I" site to regulate its own catalytic synthesis. The nucleotide forms intercalated dimers (box), each of which recognizes the characteristic I-site RxxD motif of a GGDEF domain, plus additional residues from the GGDEF domain of a symmetry-related protomer. The latter nucleotide-mediated interprotomer interactions are likely to stabilize a catalytically inactive quaternary assembly (*30*).
(B) Crystal structure of the *Vibrio cholerae* protein VCA0042 containing a PilZ domain in complex with c-di-GMP (pdb entry: 2RDE). Each protein subunit binds a single c-di-GMP molecule, whose distinct conformation is shown in the box below (*25*). Interestingly, although capable of c-di-GMP recognition, PilZ domain-containing proteins are not essential for *Vibrio cholerae* biofilm formation (*33*).
(C) Cyclic di-GMP binding to the EAL domain of FimX reveals alternative nucleotide conformation (box) and confirms the capacity of degenerate c-di-GMP turnover domains to serve as sensors for the nucleotide (pdb entry: 3HV8) (*34*).
(D) C-di-GMP base-stacking propensity is a mechanism for nucleotide recognition by a *Vibrio cholerae* riboswitch (pdb entry: 3IWN) (*35*).

Contributing to the complexity of identified c-di-GMP receptors, two other protein c-di-GMP targets have been recently identified: the transcription factor FleQ of *Pseudomonas aeruginosa*, and the catabolite activator-like protein Clp of *Xanthomonas campestris* (*37-39*). FleQ is an enhancer-binding protein with an N-terminal receiver domain, an AAA σ^{54} interaction domain, and a C-terminal helix-turn-helix DNA-binding motif (*36, 37*). Cyclic di-GMP recognition is shown to occur independently of the N-terminal receiver domain, but other than that nothing is known about the mode of protein-nucleotide interaction (*37*). Although the crystal structure of Clp has been recently solved, data regarding c-di-GMP recognition is limited and relies on merely computational models. While Clp contains a relatively conserved N-terminal cyclic nucleotide monophosphate binding domain, responsible for cAMP recognition in its *E. coli* homolog, c-di-GMP is proposed to bind Clp at an allosteric site, formed at the interface between the N-terminal domain and the C-terminal DNA-binding module (*39*).

Although limited in number, identified c-di-GMP targets are characterized by an obvious diversity and multilevel impact. The recent identification of a widespread c-di-GMP responsive riboswitch class in messenger RNAs adds another layer of complexity to the mechanisms of c-di-GMP function (*40*). Structural information regarding the precise mode of nucleotide recognition is limited to some highly divergent PilZ domains (*25, 27, 28*), the catalytically inactive FimX of Pseudomonas aeruginosa (*34*), some I-site containing diguanylate cyclases (*29, 30*), and riboswitches from *Vibrio cholerae* (*35, 41*). Whether there are a limited number of universal c-di-GMP binding motifs, or individual receptors have evolved specific modes of recognition to ensure signal isolation, remains to be further investigated. In any case, c-di-GMP effectors represent a crucial part of bacterial signaling circuits and therefore an important target for therapeutic design and development.

1.3. Clinical significance of c-di-GMP mediated processes

Opportunistic bacterial pathogens are causative agents for a variety of human diseases or disease accompanying infections. Their ability to sense and respond to different microenvironments, particularly during transition from free-living to indwelling pathogenic lifestyle, is largely dependent on a variety of adaptational strategies. Examples include phenotypic variation, biofilm formation, resistance to antibiotic treatments and virulence gene expression (*4, 6, 20, 42-49*). A number of studies have suggested that these are interlinked phenotypes, largely dependent on c-di-GMP mediated signaling phenomena (*4*).

Bacterial biofilm formation describes the process of surface attachment of planktonic cells to form sessile multicellular communities, where an extracellular matrix of exopolysaccharides, proteinaceous fimbriae and DNA provides protection from environmental insults and a medium for intercellular communication. On the cellular level, functional differentiation, genomic rearrangements, changes in motility, cell adhesion, and secretion are among the many processes accompanying the transition to a biofilm life style. Recent advances have identified biofilm formation as a multiphase process of collaborative group behavior, where distinct expression profiles reflect the progression through the biofilm cycle. This suggests that biofilm formation is a highly specialized developmental process, subject to strict temporal and spatial regulation (*5, 48-50*).

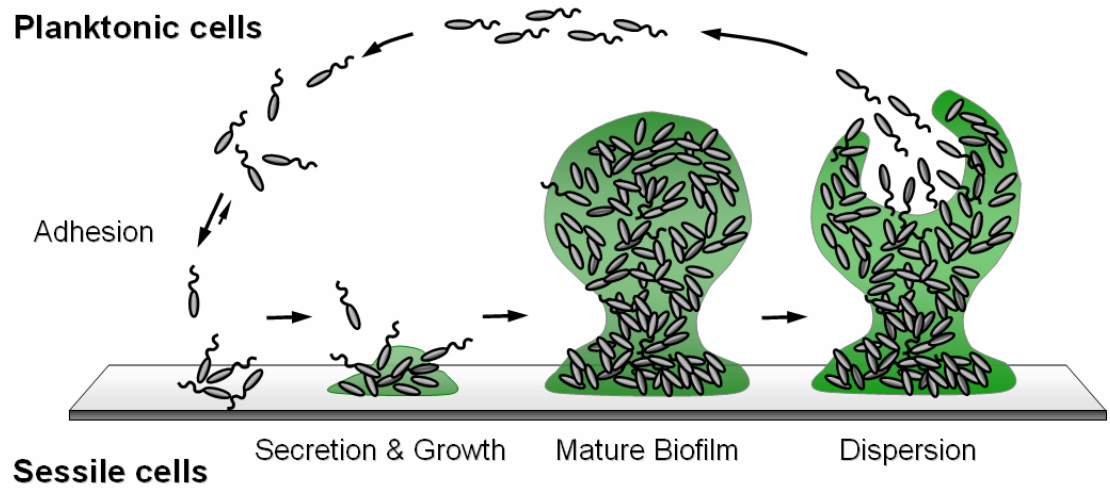


Figure 1.3: Overview of bacterial biofilm formation. Bacteria can reversibly attach to surfaces upon which loss of motility and synthesis of cellular adhesins provides strong anchorage and protective environment. Mature biofilms are developed within a certain time frame and are characterized by colony variance, complex architecture, and collaborative group behavior between functionally differentiated members. Finally, dispersion processes governed by the secretion of glycolipid surfactants, and ramnolipids in particular, lead to the release of highly motile planktonic cells, capable to colonize new surfaces (*5, 51, 52*).

Widespread in the environment, biofilms can develop on a variety of biotic and abiotic surfaces. Biofilm formation on indwelling prosthetic devices and infected luminal organs has been recognized as a major cause for chronic bacterial diseases (*45*). Examples include infections of the inner ear (otitis media), heart (bacterial endocarditis), and lungs (Legionnaire's disease and other pneumonias) of otherwise healthy individuals, as well as opportunistic infections contributing to the disease pattern of immunocompromised patients, such as in cases of AIDS, cancer, or cystic fibrosis (*7, 44, 53*). A hallmark of the latter genetic disorder is that patients succumb at a very young age to persistent lung infections by *Pseudomonas aeruginosa*, an organism commonly found in nature. It is now becoming increasingly evident that under c-di-GMP control, the lower respiratory tract of these individuals becomes

colonized by dense bacterial biofilms, characterized by partial or complete oxygen depletion and induced expression of antibiotic resistances (*48, 54*).

The other end of the spectrum of bacterial pathogenicity is represented by acute infections, characterized by short spikes in bacterial virulence. Interestingly enough, a single microorganism has oftentimes the ability to choose between the two modes, chronic and acute, of in-host propagation (*6*). *Pseudomonas aeruginosa* for example, is in some cases capable to cause acute pneumonia, breaking down host defenses and disseminating in the bloodstream within hours or days post infection. This is in stark contrast with cases of cystic fibrosis where the pathogen rarely if ever reaches the bloodstream, indicating that the acute versus chronic modes of infection are indeed quite distinct (*6*). As opposed to biofilms, systemic bacteremia during acute-phase pneumonia is caused by highly motile bacteria, which release a variety of extracellular toxins and quorum sensing molecules (*55, 56*). A number of studies have suggested c-di-GMP dependent reciprocal regulation of the biofilm forming and virulence phenotypes but the exact mechanisms through which pathogens choose between possible lifestyles remain largely unknown.

Studies on *Vibrio cholerae*, an organism indigenous to aquatic biotopes and causative agent of the diarrheal disease cholera, have helped to shed light on the mechanisms controlling the switch between biofilm formation and acute-phase bacterial virulence. Biofilm formation by this facultative pathogen has been shown to increase its resistance to osmotic and oxidative stress, protozoan grazing and acidification (*57*). It has therefore been suggested that *Vibrio cholerae* exists in the environment and enters the human host primarily in the form of biofilms, which secure passage through the acidic environment of the stomach *en route* to the intestinal tract (*58*). Once in the small intestine, however, the pathogen needs to actively repress its biofilm forming capacity in order for expression of virulence

determinants to occur (*20, 46*). On the molecular level this is achieved by the catalytic reduction of intracellular c-di-GMP levels through the specific function of one or more phosphodiesterases (*59*). Although the trigger for this catalytic activity remains unknown, some of the downstream signaling targets are now beginning to emerge. As discussed in the next chapter, we recently identified VpsT, a master transcriptional regulator of *Vibrio cholerae* biofilm formation, as a novel c-di-GMP binding receptor (60). This raises the possibility that VpsT and other transcription factors directly respond to changes in cellular c-di-GMP concentration to cause transition between different lifestyles through large-scale shifts in gene expression. In addition, a recent study identified a riboswitch class in mRNA which senses reduction in c-di-GMP levels to allow for expression of one of the main *Vibrio cholerae* virulence determinants (*40*).

In conclusion, opportunistic pathogens are able to rapidly adapt their physiology and virulence potential to the specific microconditions they encounter in the host, regardless of their environmental reservoir. On the molecular level this is achieved through various intracellular signaling cascades employing c-di-GMP as a second messenger. The overall scheme where c-di-GMP reciprocally regulates cell adhesion and persistence of biofilm forming communities on one hand, and motility and virulence of planktonic cells on the other, places c-di-GMP signaling mechanisms as an important target in the fight against both chronic and acute bacterial diseases.

1.4. Model organisms for the study of c-di-GMP mediated signal transduction

According to recent World Health Organization reports, infectious diseases of bacterial origin contribute to some of the leading mortality causes worldwide. These include gastroenteric diseases, obstructive pulmonary infections, and AIDS-derived opportunistic infections among others (*61*). Bacterial biofilm formation represents an

important obstacle to the treatment and prevention of infectious diseases, as it has been estimated that 60-80% of chronic infections can be attributed to biofilm persistence. The elucidation of the underlying signaling phenomena is therefore of immediate clinical significance, whereas the choice of appropriate model systems is of primary consideration.

Table 1.1: The leading causes of death worldwide by broad income group, 2004 (*61*).

Causes of death	Deaths in millions	% of deaths
Coronary heart disease	7.20	12.2
Stroke and other cerebrovascular diseases	5.71	9.7
Lower respiratory infections	4.18	7.1
Chronic obstructive pulmonary disease	3.02	5.1
Diarrheal diseases	2.16	3.7
HIV / AIDS	2.04	3.5
Tuberculosis	1.46	2.5
Trachea, bronchus, and lung cancers	1.32	2.3
Road traffic accidents	1.27	2.2
Prematurity and low birth weight	1.18	2.0

Bacterial biofilms can be classified according to their member species as hetero- or homobiofilms. While in the former case the collaborative community is comprised of different microbial species contributing specific metabolic and structural features, the latter indicates sufficiency of a single species to provide and coordinate all functions necessary for biofilm initiation, growth, and propagation (*62, 63*).

We chose *Pseudomonas aeruginosa*, its related species *Pseudomonas fluorescens*, and the facultative enteric pathogen *Vibrio cholerae* as the primary model organisms for our studies (*30, 60, 64*). All three species opt for the formation of homobiofilms in their natural environmental habitats, indicating that all components necessary for this process are encoded in a fully functional form by their individual genomes. The latter have been fully sequenced, and in the cases of *Vibrio cholerae*

and *Pseudomonas aeruginosa* sequences from clinical isolates are readily available.

Vibrio cholerae represents an important model system to study the dynamic interplay between biofilm formation and acute-phase bacterial virulence. It is causative agent of the potentially lethal diarrheal disease cholera, seasonal outbreaks of which continue to scourge the world's population. This places *Vibrio cholerae* as a human pathogen of primary clinical significance and makes it a preferred model organism for our structural and functional studies (*60*). *Pseudomonas aeruginosa* is another important human pathogen, which colonizes mainly immunocompromised individuals such as burnt victims and patients with cystic fibrosis. It provides an important model system for studying the role of biofilm formation in chronic bacterial infections, and in particular in the case of obstructive pulmonary diseases. Many of the intracellular signaling events involved in *Pseudomonas aeruginosa* c-di-GMP mediated pathogenicity are replicated in its closely related species *Pseudomonas fluorescens*. However, rather than being pathogenic, biofilm formation by this organism contributes to its beneficial biocontrol properties, underscoring the various aspects of this signaling process. More importantly, our studies of a *Pseudomonas fluorescens* c-di-GMP receptor helped us identify a number of homologous systems, likely regulating biofilm formation and toxin secretion in various pathogenic species (*64*).

1.5. Protein targets of interest

Using the Gram-negative species *Pseudomonas aeruginosa*, *Pseudomonas fluorescens*, and *Vibrio cholerae* as model systems, we set out to decipher signal transduction pathways controlling bacterial pathogenesis, especially in the context of c-di-GMP signal recognition. Based on recent reports in the literature (*32, 33, 37, 42, 54, 57, 65-67*), as well as on personal communication with our collaborators, we

chose several putative and known c-di-GMP targets as a focus of our structural and functional studies. Overall, they can be divided in two classes based on the level of c-di-GMP signaling impact. Whereas the first class represents several transcription factors that function as global regulators at the gene expression level, the second contains proteins with degenerate c-di-GMP turnover domains that have shown strong effects on biofilm formation or virulence upon knockout, overexpression, or mutational studies. In the current model, these proteins function on the post-translational level to ensure c-di-GMP signal relay through direct protein-protein interactions.

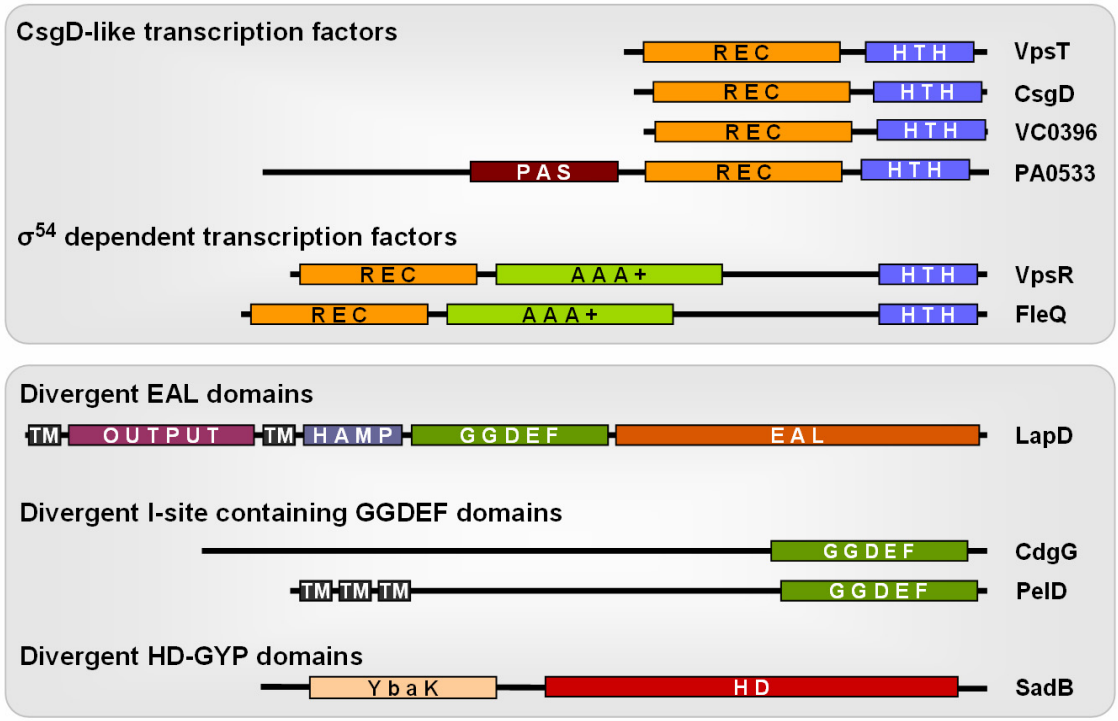


Figure 1.4: Conserved domain organization of putative and known c-di-GMP receptors. Above, CsgD-like and AAA+ σ^{54}-interaction domain-containing transcription factors involved in c-di-GMP mediated signal transduction. Below, degenerate c-di-GMP turnover domains as another class of putative c-di-GMP sensors, with divergent GGDEF, EAL, or HD-GYP domains likely involved in c-di-GMP signal recognition (*68, 69*).

A member of the first class of putative c-di-GMP targets is VpsT of *Vibrio cholerae*. VpsT was first identified as a transcription regulator required for the expression of *Vibrio* exopolysaccharide biosynthesis genes (clusters *vpsA-K* and *vpsL-Q* on the large *Vibrio* chromosome), which are required for the development of biofilms and rugose colony morphology (*42, 57*). VpsT shares high percent sequence identity with CsgD of *Escherichia coli* and *Salmonella typhimurium*, required itself for the secretion of the two major extracellular matrix components, exopolysaccharides and proteinaceous fimbriae (curli) (*43, 70*). Based on sequence homology CsgD is classified as a member of the FixJ/LuxR family of prokaryotic response regulators. The latter are typically part of archetypal two-component transduction systems, using phosphoryl transfer from upstream kinases as input signals for transcription regulation (*71*). Interestingly, CsgD contains a divergent receiver domain, where only half of the amino acids involved in phosphoryl transfer are conserved (Fig. 2.13B) and no cognate histidine kinase has been identified to date (*42, 60, 72*). Recent work has shown that CsgD can autophosphorylate in the presence of acetyl phosphate *in vitro*, but no such role for this high-energy phosphodonor has been identified *in vivo* (*73*).

As mentioned above, physiological responses following c-di-GMP signal generation appear to be spatially restricted, regardless of the high solubility and small size of this intracellular second messenger (*4, 74-76*). Regulation of CsgD represents an important example of c-di-GMP signal isolation: Although expression of the *Salmonella typhimurium csgD* gene responds to elevated c-di-GMP levels, it is not influenced by knockout of the *adrA* gene, whose protein product is a diguanylate cyclase contributing for over half of the cellular c-di-GMP pool. Rather, *csgD* expression is controlled by two additional diguanylate cyclases which by themselves

have no significant effect on the global c-di-GMP concentration (*74*).

Similarly to CsgD, VpsT of *Vibrio cholerae* regulates the expression of extracellular matrix components, has a divergent receiver domain, and lacks an associated sensor kinase (*42*). In addition, regulation by VpsT has been previously linked to c-di-GMP mediated signal transduction, and *vpsT* gene expression is under positive control by its own protein product (*42, 57*).

Based on the facts that VpsT stimulates its own transcription (*42*), that *csgD* expression in *Escherichia coli* and *Salmonella typhimurium* is controlled by the function of specific diguanylate cyclases (*74, 76*), and that small molecule binding has been previously postulated to enhance CsgD activity *in vivo* (*71*), we hypothesized that CsgD-like transcription regulators function as direct sensors for intracellular c-di-GMP concentration and act downstream of diguanylate cyclases as endpoint signal effectors.

VC0396 is another protein encoded by the *Vibrio* genome, which shares high percent sequence identity with VpsT (*68, 70*). While its role in biofilm formation remains to be experimentally determined, the fact that knockout of VpsT is sufficient for rugose to smooth morphotype conversion argues against functional redundancy between the two transcription regulators (*42*). Whether VpsT and VC0396 regulate distinct sets of genes, or they function at different stages of biofilm formation, represents yet another interesting question and underscores the rapid evolution and functional complexity of biofilm-related signaling pathways.

Based on sequence homology to VpsT, we identified PA0533 of *Pseudomonas aeruginosa* as another potential c-di-GMP target (*68, 70*). Interestingly, the latter contains an additional PAS domain N-terminal to the putative c-di-GMP binding receiver domain. Considering that among PAS domains are heme-dependent oxygen sensors and that *Pseudomonas aeruginosa* biofilms in cystic fibrosis lungs are

characterized by reduced or absent oxygen tension (*54, 77*), similar domain organization might serve to provide an important additional signaling input.

The genome of *Pseudomonas aeruginosa* also encodes for several proteins that have been experimentally confirmed to function as c-di-GMP binding receptors (*32, 34, 37*). Among these is the σ^{54}-dependent FleQ, identified as the first transcription factor to inversely regulate motility and extracellular matrix production in direct response to nucleotide recognition (*37*). When bound to DNA FleQ acts as a master activator of flagellar gene expression, as well as a repressor for the *pel* exopolysaccharide biosynthesis operon. This FleQ-DNA interaction is disrupted upon c-di-GMP recognition, which in turn leads to motility inhibition and biofilm formation (*37*). Interestingly, when we searched for FleQ homologs in other bacterial species (*68*), we stumbled upon VpsR of *Vibrio cholerae*, a positive regulator of *vps* synthesis and biofilm formation with an overall similar regulon as VpsT (*57*). Whether VpsR is another c-di-GMP target, or it relies on VpsT and/or VC0396 to introduce c-di-GMP sensitivity to the signaling network remains to be further investigated. Apart from the fact that the FleQ • c-di-GMP complex formation appears independent of the protein's N-terminal receiver domain (*37*), c-di-GMP binding and concurrent conformational changes have remained largely uncharacterized. This, together with the biological significance of *Pseudomonas aeruginosa* FleQ and *Vibrio cholerae* VpsR as the pathogens' global biofilm regulators, led us to pursue their structural and functional characterization in the light of c-di-GMP signal recognition.

A separate class of c-di-GMP sensors comprises proteins with divergent, catalytically inactive c-di-GMP turnover domains. Examples include *Pseudomonas* FimX and LapD proteins where degenerate EAL domains function as the nucleotide sensor modules (*19, 34*), as well as *Vibrio cholerae* CdgG and *Pseudomonas aeruginosa* PelD, where c-di-GMP binds to the so-called I-site of inactive diguanylate

cyclase domains (*32, 33*). The role of degenerate HD-GYP domains as potential c-di-GMP receptors remains to be experimentally determined. A putative candidate is SadB of *Pseudomonas aeruginosa*, which functions at the initial stages of bacterial attachment to surfaces and is thus crucial for biofilm formation by the pathogen (*65*). SadB functions downstream of two specific c-di-GMP turnover enzymes, a diguanylate cyclase and a phosphodiesterase, which inversely control its effects on biofilm formation and swarming motility (*66, 67*). This, together with the fact that SadB contains a domain homologous to phosphodiesterase HD-GYP modules, presents the possibility that the latter can also serve as nucleotide sensors in various c-di-GMP signaling pathways, rather than simply as enzymes providing selective signal degradation.

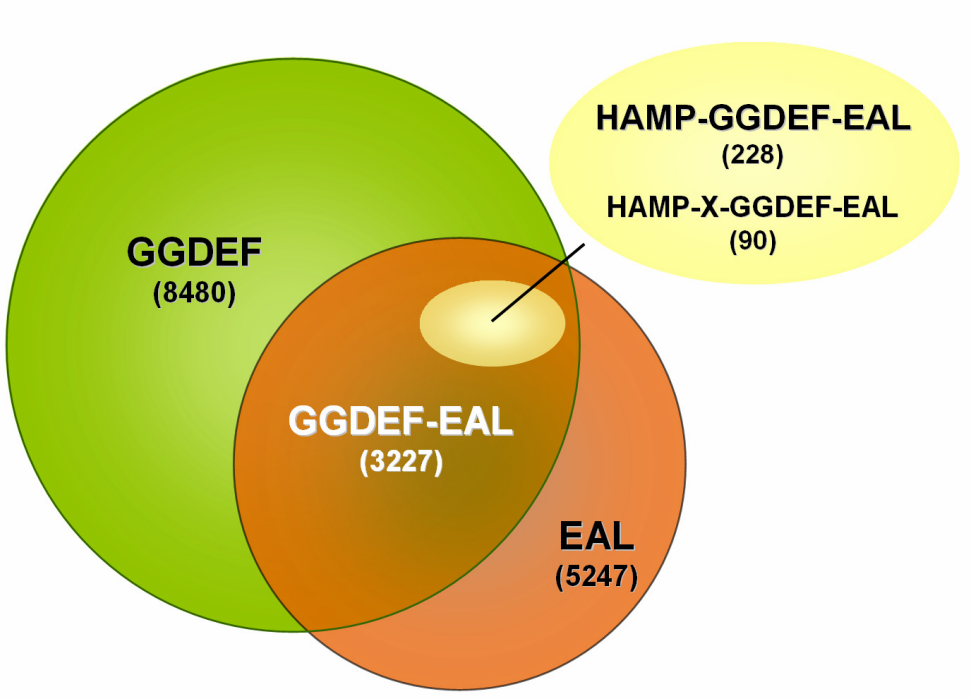


Figure 1.5: GGDEF and EAL domain abundance among bacterial genomes. Data retrieved from the SMART research database *(78, 79)*.

Catalytically active and degenerate c-di-GMP turnover domains are encoded

in large numbers in bacterial genomes (Fig. 1.5) (*9, 78, 79*). Oftentimes, these are part of multidomain proteins, allowing for a variety of regulatory inputs and signal ramifications. Interestingly, almost a fourth of all GGDEF and EAL domain-containing proteins possess both types of domains in the same polypeptide chain where no, single, or dual catalytic activity can be preserved among protein family members. About 10% of these contain an additional HAMP domain N-terminal to the GGDEF-EAL domain tandem where HAMP domains function as juxtamembrane signal relay modules found in bacterial transmembrane receptors (*80*). One such receptor with a conserved HAMP-GGDEF-EAL domain organization is the previously mentioned LapD of *Pseudomonas fluorescens*, which controls cell adhesion and biofilm formation in a c-di-GMP dependent manner (*19*). Given the direct role for LapD as a bacterial c-di-GMP receptor, as well as the large abundance of proteins with similar domain architecture– especially in the context of related pathogenic species– we set out to decipher the structure and function of LapD homologs as c-di-GMP sensors in bacterial adaptation and virulence.

In the following chapters we will present detailed structural and functional analyses of VpsT of *Vibrio cholerae* and LapD of *Pseudomonas fluorescens* as members of the two major classes of likely c-di-GMP sensors. In addition, we will outline current advances toward the structural studies of FleQ of *Pseudomonas aeruginosa* and propose an array of future experiments to help further characterize proteins responsible for bacterial c-di-GMP signal recognition.

REFERENCES

1. Ross, P., R. Mayer, et al. (1991). "Cellulose biosynthesis and function in bacteria." *Microbiol Rev* **55**(1): 35-58.
2. Weinhouse, H., S. Sapir, et al. (1997). "c-di-GMP-binding protein, a new factor regulating cellulose synthesis in Acetobacter xylinum." *FEBS Lett* **416**(2): 207-11.
3. Hengge, R. (2009). "Principles of c-di-GMP signalling in bacteria." *Nat Rev Microbiol* **7**(4): 263-73.
4. Jenal, U. and J. Malone (2006). "Mechanisms of cyclic-di-GMP signaling in bacteria." *Annu Rev Genet* **40**: 385-407.
5. O'Toole, G., H. B. Kaplan, et al. (2000). "Biofilm formation as microbial development." *Annu Rev Microbiol* **54**: 49-79.
6. Furukawa, S., S. L. Kuchma, et al. (2006). "Keeping their options open: acute versus persistent infections." *J Bacteriol* **188**(4): 1211-7.
7. Parsek, M. R. and P. K. Singh (2003). "Bacterial biofilms: an emerging link to disease pathogenesis." *Annu Rev Microbiol* **57**: 677-701.
8. Galperin, M. Y. (2006). "Structural classification of bacterial response regulators: diversity of output domains and domain combinations." *J Bacteriol* **188**(12): 4169-82.
9. Galperin, M. Y., A. N. Nikolskaya, et al. (2001). "Novel domains of the prokaryotic two-component signal transduction systems." *FEMS Microbiol Lett* **203**(1): 11-21.
10. Albright, L. M., E. Huala, et al. (1989). "Prokaryotic signal transduction mediated by sensor and regulator protein pairs." *Annu Rev Genet* **23**: 311-36.
11. Tal, R., H. C. Wong, et al. (1998). "Three cdg operons control cellular

turnover of cyclic di-GMP in Acetobacter xylinum: genetic organization and occurrence of conserved domains in isoenzymes." *J Bacteriol* **180**(17): 4416-25.

12. Simm, R., M. Morr, et al. (2004). "GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility." *Mol Microbiol* **53**(4): 1123-34.
13. Ryan, R. P., Y. Fouhy, et al. (2006). "Cell-cell signaling in Xanthomonas campestris involves an HD-GYP domain protein that functions in cyclic di-GMP turnover." *Proc Natl Acad Sci U S A* **103**(17): 6712-7.
14. D'Argenio, D. A. and S. I. Miller (2004). "Cyclic di-GMP as a bacterial second messenger." *Microbiology* **150**(Pt 8): 2497-502.
15. Schirmer, T. and U. Jenal (2009). "Structural and mechanistic determinants of c-di-GMP signalling." *Nat Rev Microbiol* **7**(10): 724-35.
16. Romling, U., M. Gomelsky, et al. (2005). "C-di-GMP: the dawning of a novel bacterial signalling system." *Mol Microbiol* **57**(3): 629-39.
17. Pesavento, C. and R. Hengge (2009). "Bacterial nucleotide-based second messengers." *Curr Opin Microbiol* **12**(2): 170-6.
18. Wolfe, A. J. and K. L. Visick (2008). "Get the message out: cyclic-Di-GMP regulates multiple levels of flagellum-based motility." *J Bacteriol* **190**(2): 463-75.
19. Newell, P. D., R. D. Monds, et al. (2009). "LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by Pseudomonas fluorescens Pf0-1." *Proc Natl Acad Sci U S A* **106**(9): 3461-6.
20. Tischler, A. D. and A. Camilli (2005). "Cyclic diguanylate regulates Vibrio cholerae virulence gene expression." *Infect Immun* **73**(9): 5873-82.
21. Kim, Y. K. and L. L. McCarter (2007). "ScrG, a GGDEF-EAL protein,

participates in regulating swarming and sticking in Vibrio parahaemolyticus." *J Bacteriol* **189**(11): 4094-107.

22. Kulasakara, H., V. Lee, et al. (2006). "Analysis of Pseudomonas aeruginosa diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence." *Proc Natl Acad Sci U S A* **103**(8): 2839-44.
23. Amikam, D. and M. Y. Galperin (2006). "PilZ domain is part of the bacterial c-di-GMP binding protein." *Bioinformatics* **22**(1): 3-6.
24. Ryjenkov, D. A., R. Simm, et al. (2006). "The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria." *J Biol Chem* **281**(41): 30310-4.
25. Benach, J., S. S. Swaminathan, et al. (2007). "The structural basis of cyclic diguanylate signal transduction by PilZ domains." *Embo J* **26**(24): 5153-66.
26. Pratt, J. T., R. Tamayo, et al. (2007). "PilZ domain proteins bind cyclic diguanylate and regulate diverse processes in Vibrio cholerae." *J Biol Chem* **282**(17): 12860-70.
27. Ramelot, T. A., A. Yee, et al. (2007). "NMR structure and binding studies confirm that PA4608 from Pseudomonas aeruginosa is a PilZ domain and a c-di-GMP binding protein." *Proteins* **66**(2): 266-71.
28. Ko, J., K. S. Ryu, et al. (2010). "Structure of PP4397 reveals the molecular basis for different c-di-GMP binding modes by Pilz domain proteins." *J Mol Biol* **398**(1): 97-110.
29. Chan, C., R. Paul, et al. (2004). "Structural basis of activity and allosteric control of diguanylate cyclase." *Proc Natl Acad Sci U S A* **101**(49): 17084-9.
30. De, N., M. Pirruccello, P. V. Krasteva et al. (2008). "Phosphorylation-independent regulation of the diguanylate cyclase WspR." *PLoS Biol* **6**(3): e67.

31. Christen, B., M. Christen, et al. (2006). "Allosteric control of cyclic di-GMP signaling." *J Biol Chem* **281**(42): 32015-24.

32. Lee, V. T., J. M. Matewish, et al. (2007). "A cyclic-di-GMP receptor required for bacterial exopolysaccharide production." *Mol Microbiol* **65**(6): 1474-84.

33. Beyhan, S., L. S. Odell, et al. (2008). "Identification and characterization of cyclic diguanylate signaling systems controlling rugosity in Vibrio cholerae." *J Bacteriol* **190**(22): 7392-405.

34. Navarro, M. V., N. De, et al. (2009). "Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX." *Structure* **17**(8): 1104-16.

35. Kulshina, N., N. J. Baird, et al. (2009). "Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch." *Nat Struct Mol Biol* **16**(12): 1212-7.

36. Arora, S. K., B. W. Ritchings, et al. (1997). "A transcriptional activator, FleQ, regulates mucin adhesion and flagellar gene expression in Pseudomonas aeruginosa in a cascade manner." *J Bacteriol* **179**(17): 5574-81.

37. Hickman, J. W. and C. S. Harwood (2008). "Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor." *Mol Microbiol* **69**(2): 376-89.

38. Leduc, J. L. and G. P. Roberts (2009). "Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of Xanthomonas axonopodis pv. citri." *J Bacteriol* **191**(22): 7121-2.

39. Chin, K. H., Y. C. Lee, et al. (2010) "The cAMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in Xanthomonas campestris." *J Mol Biol* **396**(3): 646-62.

40. Sudarsan, N., E. R. Lee, et al. (2008). "Riboswitches in eubacteria sense the second messenger cyclic di-GMP." *Science* **321**(5887): 411-3.

41. Smith, K. D., S. V. Lipchock, et al. (2009). "Structural basis of ligand binding by a c-di-GMP riboswitch." *Nat Struct Mol Biol* **16**(12): 1218-23.

42. Casper-Lindley, C. and F. H. Yildiz (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." *J Bacteriol* **186**(5): 1574-8.

43 Romling, U. (2005). "Characterization of the rdar morphotype, a multicellular behaviour in Enterobacteriaceae." *Cell Mol Life Sci* **62**(11): 1234-46.

44. Hall-Stoodley, L., J. W. Costerton, et al. (2004). "Bacterial biofilms: from the natural environment to infectious diseases." *Nat Rev Microbiol* **2**(2): 95-108.

45. del Pozo, J. L. and R. Patel (2007). "The challenge of treating biofilm-associated bacterial infections." *Clin Pharmacol Ther* **82**(2): 204-9.

46. Tamayo, R., S. Schild, et al. (2008). "Role of cyclic Di-GMP during el tor biotype Vibrio cholerae infection: characterization of the in vivo-induced cyclic di-GMP phosphodiesterase CdpA." *Infect Immun* **76**(4): 1617-27.

47. Hoffman, L. R., D. A. D'Argenio, et al. (2005). "Aminoglycoside antibiotics induce bacterial biofilm formation." *Nature* **436**(7054): 1171-5.

48. Whiteley, M., M. G. Bangera, et al. (2001). "Gene expression in Pseudomonas aeruginosa biofilms." *Nature* **413**(6858): 860-4.

49. Mulcahy, H., L. Charron-Mazenod, et al. (2008). "Extracellular DNA chelates cations and induces antibiotic resistance in Pseudomonas aeruginosa biofilms." *PLoS Pathog* **4**(11): e1000213.

50. Barken, K. B., S. J. Pamp, et al. (2008). "Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in Pseudomonas aeruginosa biofilms." *Environ Microbiol* **10**(9): 2331-43.

51 Boles, B. R., M. Thoendel, et al. (2005). "Rhamnolipids mediate detachment of Pseudomonas aeruginosa from biofilms." *Mol Microbiol* **57**(5): 1210-23.

52. Lequette, Y. and E. P. Greenberg (2005). "Timing and localization of rhamnolipid synthesis gene expression in Pseudomonas aeruginosa biofilms." *J Bacteriol* **187**(1): 37-44.

53. Singh, P. K., A. L. Schaefer, et al. (2000). "Quorum-sensing signals indicate that cystic fibrosis lungs are infected with bacterial biofilms." *Nature* **407**(6805): 762-4.

54. Platt, M. D., M. J. Schurr, et al. (2008). "Proteomic, microarray, and signature-tagged mutagenesis analyses of anaerobic Pseudomonas aeruginosa at pH 6.5, likely representing chronic, late-stage cystic fibrosis airway conditions." *J Bacteriol* **190**(8): 2739-58.

55. Vance, R. E., A. Rietsch, et al. (2005). "Role of the type III secreted exoenzymes S, T, and Y in systemic spread of Pseudomonas aeruginosa PAO1 in vivo." *Infect Immun* **73**(3): 1706-13.

56. Mougous, J. D., M. E. Cuff, et al. (2006). "A virulence locus of Pseudomonas aeruginosa encodes a protein secretion apparatus." *Science* **312**(5779): 1526-30.

57. Beyhan, S., K. Bilecen, et al. (2007). "Regulation of rugosity and biofilm formation in Vibrio cholerae: comparison of VpsT and VpsR regulons and epistasis analysis of vpsT, vpsR, and hapR." *J Bacteriol* **189**(2): 388-402.

58. Faruque, S. M., K. Biswas, et al. (2006). "Transmissibility of cholera: in vivo-formed biofilms and their relationship to infectivity and persistence in the environment." *Proc Natl Acad Sci U S A* **103**(16): 6350-5.

59. Tamayo, R., A. D. Tischler, et al. (2005). "The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase." *J Biol Chem* **280**(39): 33324-30.

60. Krasteva, P. V., J. C. Fong, et al. (2010). "Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP." *Science* **327**(5967): 866-8.

61. World Health Organization (2004). “The ten leading causes of death by broad income group.” *Factsheet* **310**

62. Davey, M. E. and A. O'Toole G (2000). "Microbial biofilms: from ecology to molecular genetics." *Microbiol Mol Biol Rev* **64**(4): 847-67.

63. Boles, B. R., M. Thoendel, et al. (2004). "Self-generated diversity produces "insurance effects" in biofilm communities." *Proc Natl Acad Sci U S A* **101**(47): 16630-5.

64. Navarro, M. V. A. S., P. D. Newell, P. V. Krasteva et al. “Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis” (*submitted*)

65. Caiazza, N. C., J. H. Merritt, et al. (2007). "Inverse regulation of biofilm formation and swarming motility by Pseudomonas aeruginosa PA14." *J Bacteriol* **189**(9): 3603-12.

66. Kuchma, S. L., K. M. Brothers, et al. (2007). "BifA, a cyclic-Di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by Pseudomonas aeruginosa PA14." *J Bacteriol* **189**(22): 8165-78.

67. Merritt, J. H., K. M. Brothers, et al. (2007). "SadC reciprocally influences biofilm formation and swarming motility via modulation of exopolysaccharide production and flagellar function." *J Bacteriol* **189**(22): 8154-64.

68. Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." *J Mol Biol* **215**(3): 403-10.

69. Marchler-Bauer, A., J. B. Anderson, et al. (2009). "CDD: specific functional annotation with the Conserved Domain Database." *Nucleic Acids Res*

37(Database issue): D205-10.

70. Larkin, M. A., G. Blackshields, et al. (2007). "Clustal W and Clustal X version 2.0." *Bioinformatics* **23**(21): 2947-8.

71. Chirwa, N. T. and M. B. Herrington (2003). "CsgD, a regulator of curli and cellulose synthesis, also regulates serine hydroxymethyltransferase synthesis in Escherichia coli K-12." *Microbiology* **149**(Pt 2): 525-35.

72. Gao, R. and A. M. Stock (2009). "Biological insights from structures of two-component proteins." *Annu Rev Microbiol* **63**: 133-54.

73. Zakikhany, K., C. R. Harrington, et al. (2010) "Unphosphorylated CsgD controls biofilm formation in Salmonella enterica serovar Typhimurium." *Mol Microbiol* **77**(3): 771-86.

74. Kader, A., R. Simm, et al. (2006). "Hierarchical involvement of various GGDEF domain proteins in rdar morphotype development of Salmonella enterica serovar Typhimurium." *Mol Microbiol* **60**(3): 602-16.

75. Christen, M., H. D. Kulasekara, et al. (2010) "Asymmetrical distribution of the second messenger c-di-GMP upon bacterial cell division." *Science* **328**(5983): 1295-7.

76. Weber, H., C. Pesavento, et al. (2006). "Cyclic-di-GMP-mediated signalling within the sigma network of Escherichia coli." *Mol Microbiol* **62**(4): 1014-34.

77. Taylor, B. L. and I. B. Zhulin (1999). "PAS domains: internal sensors of oxygen, redox potential, and light." *Microbiol Mol Biol Rev* **63**(2): 479-506.

78. Schultz, J., F. Milpetz, et al. (1998). "SMART, a simple modular architecture research tool: identification of signaling domains." *Proc Natl Acad Sci U S A* **95**(11): 5857-64.

79. Letunic, I., T. Doerks, et al. (2009). "SMART 6: recent updates and new developments." *Nucleic Acids Res* **37**(Database issue): D229-32.

80. Parkinson, J. S. (2010) "Signaling Mechanisms of HAMP Domains in Chemoreceptors and Sensor Kinases." *Annu Rev Microbiol.***64**: 101-22

CHAPTER 2

Preface

The biofilm-forming capacity of *Vibrio cholerae* in natural aquatic habitats is well documented and predicted to be a survival strategy for the organism outside of the host body. It has also been recently reported that the average infectivity of the biofilm forming rugose variants of *Vibrio cholerae* is significantly higher than that of planktonic cells and that the biofilm life form secures survival of the pathogen in the acidic environment of the stomach *en route* to the intestinal tract. Biofilm formation and other processes contributing to pathogenicity are regulated by the bacterially unique second messenger cyclic di-GMP (c-di-GMP). While enzymes controlling cellular c-di-GMP levels have been identified in large numbers in the *Vibrio cholerae* genome, little is known regarding signal transmission and the targets of this signaling molecule.

In the following work, "*Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP" by Petya Krasteva *et al.*, we identify VpsT as a novel receptor for c-di-GMP signal transduction. By combining structural and functional approaches, we present several novel findings elucidating the molecular mechanism by which this transcriptional regulator controls a switch in bacterial life style through transition from free-swimming, motile existence to the sessile collaborative group behavior of biofilm formation. Our work identifies VpsT as a signal-integrating master regulator that inversely controls extracellular matrix production and flagellar motility. Interestingly, VpsT is not regulated by phosphorylation as established for canonical response receiver proteins, but employs a

novel structural feature and small molecule binding for regulation. The structural analysis defines a new class of response receiver domains that includes other phosphorylation incompetent VpsT family proteins, the homologous CsgD from *Salmonella* and *Escherichia coli,* as well as canonical phosphorylation-dependent response regulators. In addition, the relative orientation of the DNA binding domains of VpsT in the functionally relevant oligomeric assemblies indicates that changes in DNA architecture are likely to accompany VpsT-dependent transcriptional regulation.

The following work would not have been possible without the outstanding contributions from members of the Yildiz Lab at UC Santa Cruz and the Sondermann Lab at Cornell University. I am deeply grateful to Fitnat Yildiz and my mentor, Dr. Holger Sondermann, for help with the experimental design, data analysis, and manuscript writing; to Jiunn C. N. Fong and Nicholas J. Shikuma for conducting the cell-based experiments; to Sinem Beyhan for providing certain strains and reagents; and to Marcos V. A. S. Navarro for providing access to synchrotron data collection and for general advice in crystallographic data analysis.

Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP *

Microorganisms can switch from a planktonic, free-swimming life style to a sessile, colonial state, called a biofilm, conferring resistance to environmental stress. Conversion between the motile and biofilm life style has been attributed to increased levels of the prokaryotic second messenger cyclic di-guanosine monophosphate (c-di-GMP), yet the signaling mechanisms mediating such a global switch are poorly understood. Here we show that the transcriptional regulator VpsT from *Vibrio cholerae* directly senses c-di-GMP to inversely control extracellular matrix production and motility, identifying VpsT as a master regulator for biofilm formation. Rather than being regulated by phosphorylation, VpsT undergoes a change in oligomerization upon c-di-GMP binding.

In *Vibrio cholerae*, biofilm formation is facilitated by colonial morphotype variation (*1-4*). Rugose variants produce increased levels of extracellular matrix via the expression of *Vibrio* polysaccharide (*vps*) genes and genes encoding matrix proteins. *vps* expression is under the control of two positive transcriptional regulators, VpsT and VpsR (*5, 6*). VpsT is a member of the FixJ/LuxR/CsgD family of response regulators, typically effectors in two-component signal transduction systems that use

---

* Reproduced with permission from [Petya V. Krasteva, Jiunn C.N. Fong, Nicholas J. Shikuma, Sinem Beyhan, Marcos V.A.S. Navarro, Fitnat H. Yildiz, and Holger Sondermann (2010) *Science 327*, 866-868] © 2010 Krasteva *et al.*

Author contributions are as follows: P.V.K., F.H.Y. and H.S. designed research; P.V.K., J.C.N.F., and N.J.S. performed research; S.B. provided certain strains and reagents; P.V.K., J.C.N.F., N.J.S., M.V.A.S.N., F.H.Y., and H.S. analyzed data; and P.V.K., F.H.Y. and H.S. wrote the paper.

phosphoryl transfer from upstream kinases to modulate response regulator protein activity (*7-9*). Although the putative phosphorylation site is conserved in VpsT's receiver domain, other residues crucial for phosphotransfer-dependent signaling are not and no cognate kinase has been identified to date (Fig. 2.4). Regulation by VpsT and VpsR has been linked to signal transduction utilizing the bacterial second messenger c-di-GMP (*10, 11*) (Fig. 2.5), yet little is known about the direct targets of the nucleotide. A riboswitch has been identified as a c-di-GMP-target regulating gene expression of a small number of genes, but is unlikely to account for the global change in transcriptional profile required for biofilm formation (*12*). Neither do PilZ domain-containing proteins, potential c-di-GMP effectors, affect rugosity since a *V. cholerae* strain lacking all five PilZ domain-containing proteins retains its colony morphology and ability to overproduce *vps* gene products (*13*).

VpsT consists of an N-terminal receiver (REC) and a C-terminal helix-turn-helix (HTH) domain, with the latter mediating DNA binding (Fig. 2.1A; see Supplemental Information for details). Unlike other REC domains, the canonical $(\alpha/\beta)_5$-fold in VpsT is extended by an additional helix at its C-terminus (Fig. 2.1A; helix $\alpha6$). The HTH domain buttresses against an interface formed by helices 3 and 4 of the N-terminal regulatory domain. There are two non-overlapping dimerization interfaces between non-crystallographic VpsT protomers (chain A-chain B and chain A-chain B^{sym}; Fig. 2.1B). The c-di-GMP-independent interface involves interactions mediated by a methionine residue (M^{17}) located at the beginning of $\alpha1$ and a binding pocket that extends into the putative phosphorylation site of the REC domain (Fig. 2.6A). The second interface involves $\alpha6$ of the REC domain, in contrast to canonical response regulators such as CheY and PhoB that utilize a surface formed by $\alpha4$-$\beta5$-$\alpha5$ for dimerization (*9*). The binding of two intercalated c-di-GMP molecules to the base of $\alpha6$ stabilizes VpsT dimers utilizing this interface (Fig. 2.1 and 2.6B).

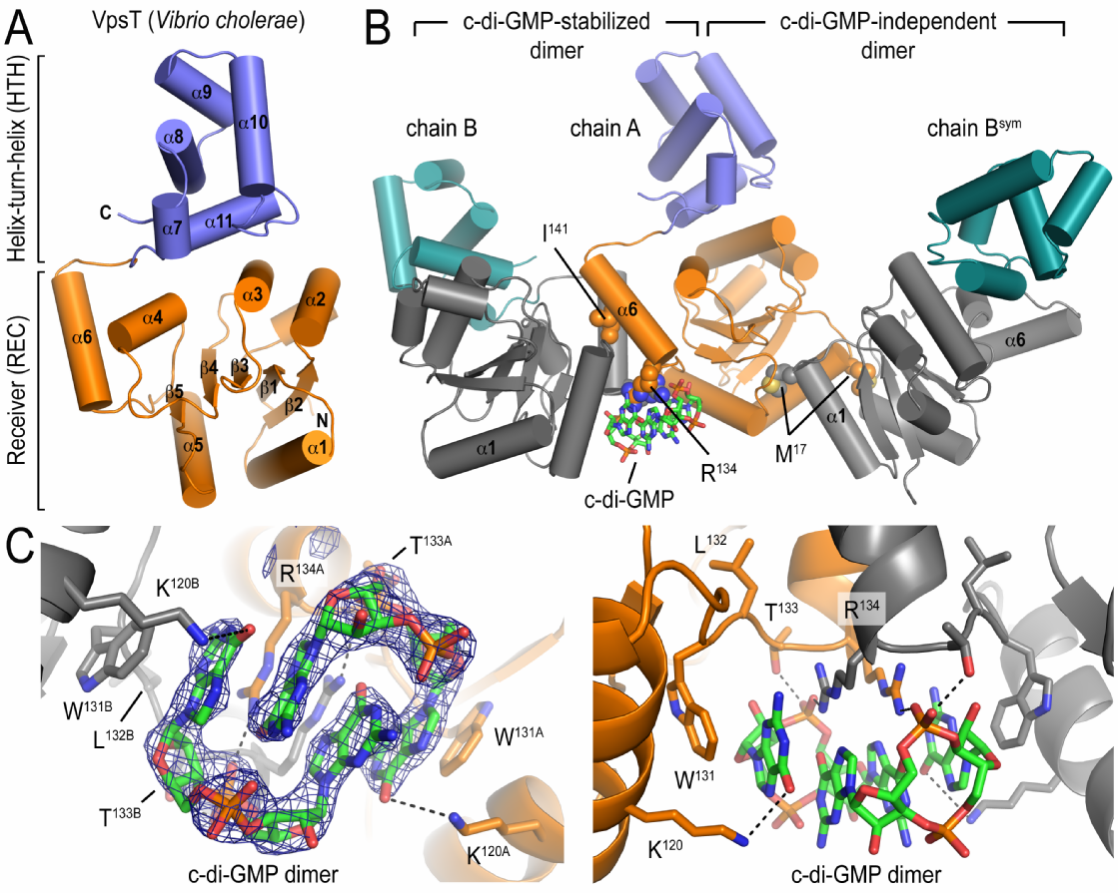


Figure 2.1: Crystal structure of VpsT. (A) Structure of a VpsT protomer. **(B)** Crystal structure of a crystallographic trimer representing two potentially relevant, non-overlapping dimerization interfaces. Cyclic di-GMP molecules are shown as sticks, key residues mediating ligand binding and interprotomer interactions are shown as spheres. **(C)** Close-up view of the nucleotide binding pocket with residues involved in coordinating the ligand shown as sticks. A (|Fo|-|Fc|) electron density map contoured at 3.6σ is shown as calculated from a model prior to inclusion of c-di-GMP.

The binding motif for c-di-GMP in VpsT consists of a 4-residue-long, conserved W[F/L/M][T/S]R sequence (Fig. 2.4). The side chains of the tryptophan and arginine form π-stacking interactions with the purine rings of the nucleotide (Fig. 2.1C). While the hydrophobic residue in the second position plays a structural role being buried in the REC domain, the threonine residue at position 3 forms a hydrogen bond with the phosphate moiety of c-di-GMP. A subclass of VpsT/CsgD homologs

exists with a proline substitution in position 3 (W[F/L/M]PR). Although CsgD is also functionally linked to c-di-GMP signaling in *E. coli* and *Salmonella* (*14, 15*), its binding pocket appears to be distinct from that of VpsT since it displays a highly conserved YF[T/S]Q motif that is unlikely to accommodate c-di-GMP (Fig. 2.6B).

The apparent affinity of VpsT for c-di-GMP, determined by isothermal titration calorimetry, is 3.2 μM with 1:1 stoichiometry, consistent with a dimer of c-di-GMP binding to a dimer of VpsT (Fig. 2.7A). Single-point mutations in the conserved c-di-GMP binding motif ($VpsT^{R134A}$, $VpsT^{W131F}$ or $VpsT^{T133V}$) or in the isoleucine in α6 of the c-di-GMP-stabilized REC dimerization interface ($VpsT^{I141E}$) abolished c-di-GMP binding, indicating that dimeric REC domains are required for binding (Fig. 2.7B). Conversely, mutation of a key residue in the nucleotide-independent interface ($VpsT^{M17D}$) had no effect on c-di-GMP binding. Based on static multi-angle light scattering, $VpsT^{M17D}$ exists as a monomeric species in the absence of c-di-GMP, whereas intermediate molecular weights for the wild-type VpsT and the mutants $VpsT^{R134A}$ and $VpsT^{I141E}$ indicated fast exchange between monomers and dimers, presumably through the c-di-GMP-independent interface (Fig. 2.8 and Table S2). Addition of c-di-GMP increases the molecular weight of $VpsT^{M17D}$ and wild-type VpsT (Fig. 2.8 and 2.9) while the oligomeric state of $VpsT^{R134A}$ and $VpsT^{I141E}$ is insensitive to the nucleotide.

The role of c-di-GMP recognition and the relevance of the two dimer interfaces in DNA-binding and VpsT-regulated gene expression was assessed by using c-di-GMP binding (R^{134}) and dimerization (I^{141} or M^{17}) mutants (Fig. 2.2). In electromobility shift assays we used regulatory sequences upstream of *vpsL*, a gene under positive control of VpsT (Fig. 2.2A) (*6*). DNA mobility shifts were observed only for the wild-type (wt) and $VpsT^{M17D}$ forms, where the effect was protein-specific and c-di-GMP-dependent. In addition, nucleotide-dependent DNA binding of VpsT

was observed to multiple and relatively remote sites in the regulatory region of *vpsL*.

To evaluate the functional importance of VpsT oligomers and c-di-GMP binding in cells, we measured transcription of *vps* genes by using a chromosomal *vpsL*p-*lacZ* transcriptional fusion in the Δ*vpsT* strain harboring wild-type VpsT, VpsT point mutants ($VpsT^{M17D}$, $VpsT^{R134A}$ or $VpsT^{I141E}$) or the insert-less expression vector (pBAD) (Fig. 2.2B). The presence of wild-type VpsT and $VpsT^{M17D}$ resulted in increased *vpsL* expression, similar to the wild-type rugose strain carrying vector only, while Δ*vpsT* strains with $VpsT^{R134A}$, $VpsT^{I141E}$ or the empty vector did not exhibit such an increase. These data confirm that c-di-GMP-mediated oligomerization is critical for VpsT function. Mutations in the putative phosphorylation site designed to produce a constitutively inactive or active state, $VpsT^{D60A}$ or $VpsT^{D60E}$, respectively, did not alter the efficiency of VpsT significantly. Hence, regulation of gene expression is presumably independent of phosphorylation of VpsT (see also Fig. 2.10).

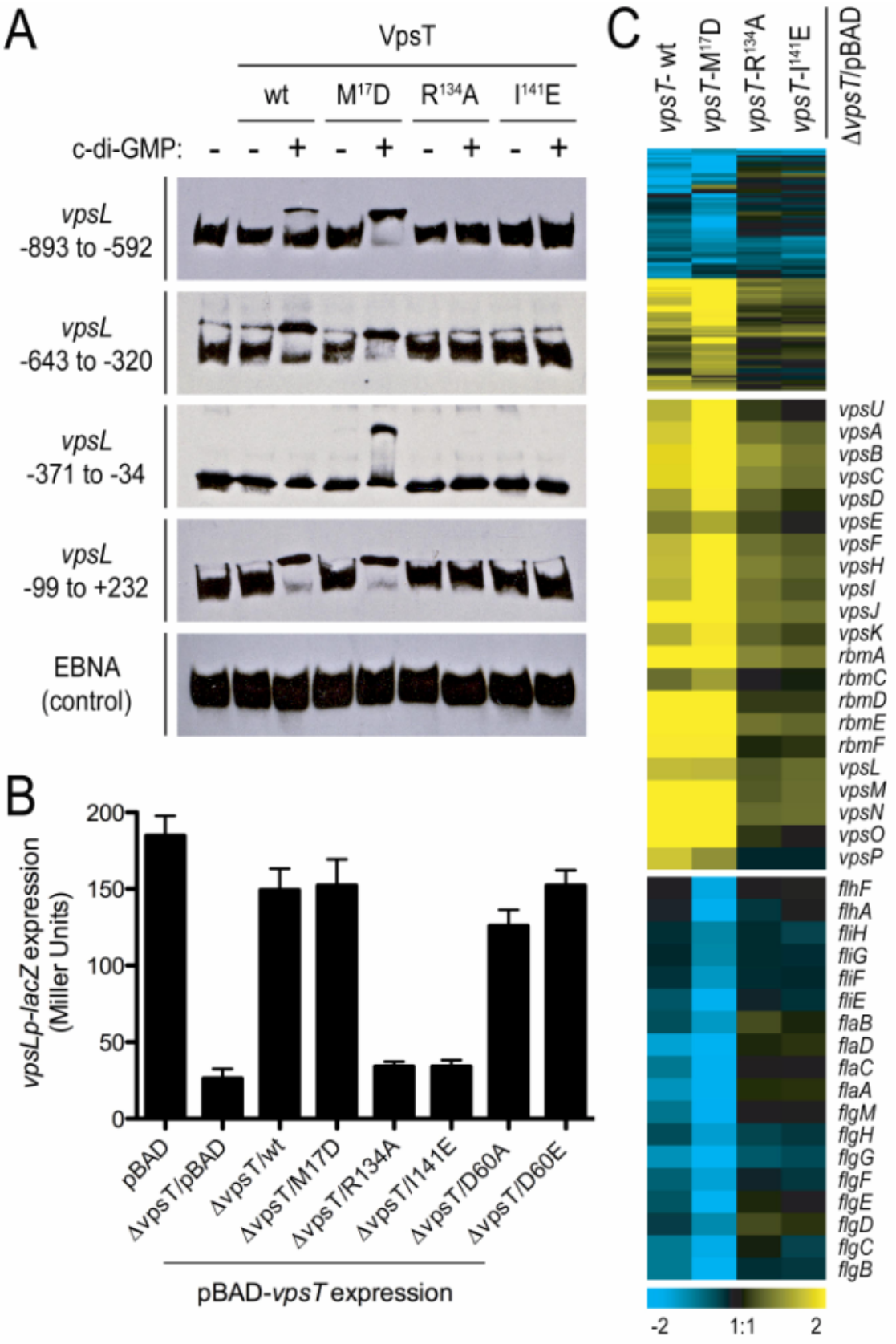
Next, we determined the gene regulatory potential as a function of c-di-GMP binding and oligomerization through whole genome expression profiling by comparing a Δ*vpsT* strain harboring either wild-type VpsT or VpsT point mutants ($VpsT^{M17D}$, $VpsT^{R134A}$ or $VpsT^{I141E}$) to that of cells harboring the pBAD vector alone (Fig. 2.2C; Table S3). Genes located in the *vps*-I and *vps*-II clusters, as well as the *vps* intergenic region were strongly induced upon expression of wild-type VpsT and $VpsT^{M17D}$, and significantly less so in the strains expressing $VpsT^{R134A}$ or $VpsT^{I141E}$. We also observed that the expression of several genes encoding flagellar proteins was decreased in cells expressing wild-type VpsT and $VpsT^{M17D}$ but not in cells expressing $VpsT^{R134A}$ or $VpsT^{I141E}$, suggesting that VpsT inversely regulates motility and matrix production in a c-di-GMP-dependent manner (Fig. 2.2C and 2.10). These results were corroborated in motility assays, in which a Δ*vpsT* strain or strains

expressing c-di-GMP-binding mutants showed increased migration on soft agar plates compared to rugose strains that express VpsT forms that are competent of c-di-GMP-dependent dimerization (Fig. 2.3A).

The strain harboring $VpsT^{M17D}$ had a similar expression profile to the strains harboring wild-type VpsT however with increased magnitude, indicating that c-di-GMP-independent dimerization could be inhibitory or regulatory (Fig. 2.2A and 2.2C). In contrast, the c-di-GMP-dependent interaction between two VpsT monomers is sufficient and necessary for DNA recognition and transcriptional regulation.

The corrugated appearance of rugose colonies can be attributed largely to increased levels of exopolysaccharides, which are induced by VpsT (*6*). As a consequence, *V. cholerae* mutants lacking *vpsT* produce smooth and flat colonies (Fig. 2.3B). To elucidate phenotypic consequences of mutations abolishing c-di-GMP binding and/or dimerization of VpsT, we compared the colony morphology of a Δ*vpsT* strain harboring wild-type VpsT or one of the point mutants described above. Expression of wild-type VpsT and $VpsT^{M17D}$ resulted in smooth-to-rugose conversion, where spot corrugation was greater in a Δ*vpsT* strain harboring $VpsT^{M17D}$ compared to a strain with wild-type VpsT. Introduction of $VpsT^{R134A}$, $VpsT^{I141E}$ or a double-mutant $VpsT^{M17D/R134A}$ failed to promote the smooth-to-rugose switch, but led to a distinct phenotype, characterized by increased spot diameter and weak corrugation with a notable radial pattern (Fig. 2.3B and 2.11).

Figure 2.2: Transcriptional regulation by VpsT.
(A) Electromobility shift assays with purified proteins and biotin labeled fragments tiling the *vpsL* promoter region. Numbers indicate position relative to the open reading frame start.
(B) *vpsL* gene expression in different genetic backgrounds harboring a single-copy chromosomal *vpsLp-lacZ* fusion. Data are mean of 8 replicates -/+ SD (Data courtesy of Jiunn C.N. Fong, Nicholas J. Shikuma, and Fitnat H. Yildiz).
(C) Whole genome expression profiling. Top, compact heatmap in a log2-based pseudocolor scale (yellow, induced; blue, repressed) comparing a total of 108 differentially expressed genes in a Δ*vpsT* strain expressing wild-type (wt) or mutated VpsT versions compared to the vector control. Midpanel, expression profiles of genes located in and between the *vps*-I and *vps*-II clusters; bottom, expression profiles of flagellar biosynthesis genes (Data courtesy of Jiunn C.N. Fong, Nicholas J. Shikuma, and Fitnat H. Yildiz).


A
VpsT
wt
M17D
R134A
I141E
c-di-GMP: - - + - + - + - +
vpsL -893 to -592
vpsL -643 to -320
vpsL -371 to -34
vpsL -99 to +232
EBNA (control)
B
vpsLp-lacZ expression (Miller Units)
200
150
100
50
0
pBAD
ΔvpsT/pBAD
ΔvpsT/wt
ΔvpsT/M17D
ΔvpsT/R134A
ΔvpsT/I141E
ΔvpsT/D60A
ΔvpsT/D60E
pBAD-vpsT expression
C
vpsT- wt
vpsT-M17D
vpsT-R134A
vpsT-I141E
ΔvpsT/pBAD
vpsU
vpsA
vpsB
vpsC
vpsD
vpsE
vpsF
vpsH
vpsI
vpsJ
vpsK
rbmA
rbmC
rbmD
rbmE
rbmF
vpsL
vpsM
vpsN
vpsO
vpsP
flhF
flhA
fliH
fliG
fliF
fliE
flaB
flaD
flaC
flaA
flgM
flgH
flgG
flgF
flgE
flgD
flgC
flgB
-2
1:1
2

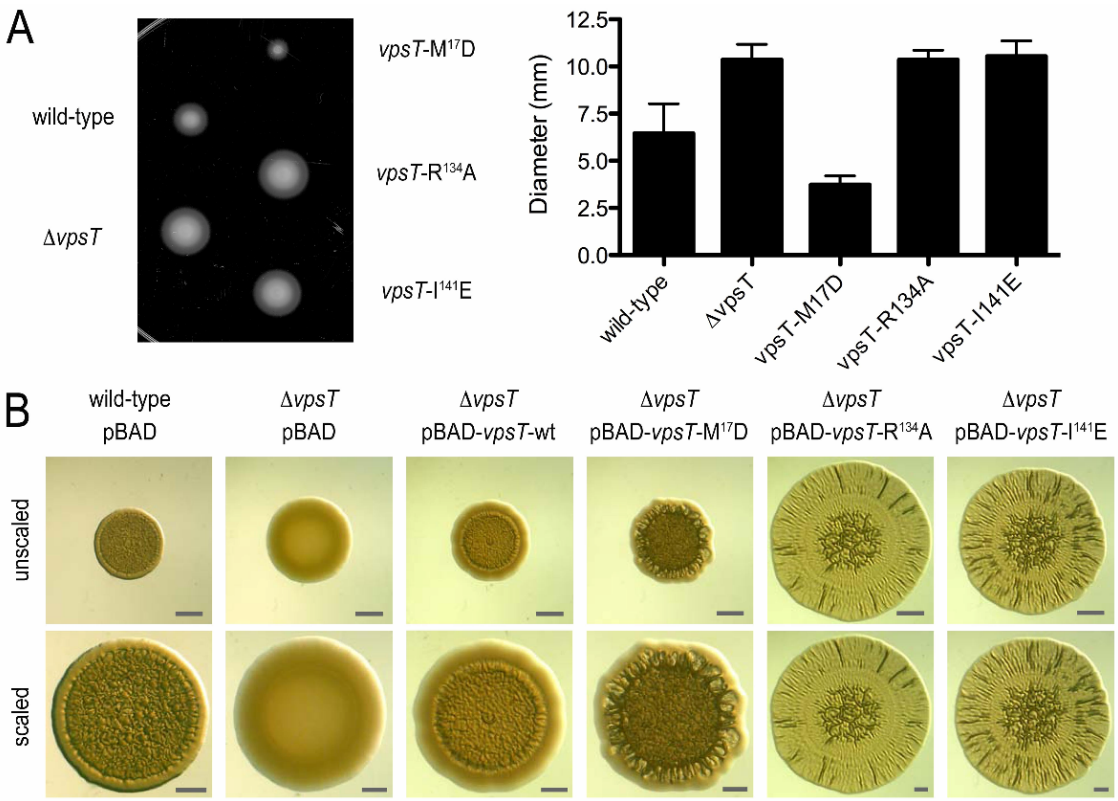


Figure 2.3: Functional characterization of rugose wild-type, Δ*vpsT* strains, and Δ*vpsT* strains expressing wild-type or mutant forms of VpsT.
(A) Motility phenotypes on semisolid LB agar plates. For strains expressing mutants of VpsT, single chromosomal insertion mutants are shown. The graph shows the mean migration zone diameter of each strain. Data are mean of 11 replicates -/+ SD (Data courtesy of Jiunn C.N. Fong, Nicholas J. Shikuma, and Fitnat H. Yildiz).
(B) Spot morphologies. A wild-type rugose strain carrying the vector (pBAD) and Δ*vpsT* strains carrying the vector or plasmids containing wild-type or mutant *vpsT* are shown (top, unscaled; bottom, scaled to similar diameter; bars=1 mm) (Data courtesy of Jiunn C.N. Fong, Nicholas J. Shikuma, and Fitnat H. Yildiz).

Cyclic di-GMP in the rugose variant is required for increased *vps* and VpsT gene expression (*10, 11*), suggesting that VpsT is involved in a positive feedback loop that integrates c-di-GMP to produce a robust transcriptional response. Robust matrix and biofilm formation relies on the mutual dependence of VpsT and VpsR, with VpsT introducing c-di-GMP-sensitivity to the regulatory network. In contrast, the transcriptional regulator FleQ from *Pseudomonas aeruginosa*, a distant VpsR-

homolog, appears to directly sense c-di-GMP independently of a VpsT-homolog by using a distinct c-di-GMP binding motif (*16*).

Taken together, we establish VpsT as a transcriptional regulator that inversely regulates biofilm formation and motility by directly integrating c-di-GMP signaling. Cyclic di-GMP-driven dimerization is mediated by an extension of the canonical receiver domains, a structural motif that defines a wide-spread class of response regulators including CsgD and other LuxR family proteins. While some mechanisms may only pertain to close homologs of VpsT such as c-di-GMP-dependent dimerization, the general mode of action involving dimerization accompanied with changes in the relative orientation of the DNA binding domains is likely to be relevant for the large family of homologous transcription factors.

REFERENCES

1. Mizunoe, Y., S. N. Wai, et al. (1999). "Isolation and characterization of rugose form of Vibrio cholerae O139 strain MO10." *Infect Immun* **67**(2): 958-63.
2. Yildiz, F. H. and G. K. Schoolnik (1999). "Vibrio cholerae O1 El Tor: identification of a gene cluster required for the rugose colony type, exopolysaccharide production, chlorine resistance, and biofilm formation." *Proc Natl Acad Sci U S A* **96**(7): 4028-33.
3. Fong, J. C., K. Karplus, et al. (2006). "Identification and characterization of RbmA, a novel protein required for the development of rugose colony morphology and biofilm structure in Vibrio cholerae." *J Bacteriol* **188**(3): 1049-59.
4. Fong, J. C. and F. H. Yildiz (2007). "The rbmBCDEF gene cluster modulates development of rugose colony morphology and biofilm formation in Vibrio cholerae." *J Bacteriol* **189**(6): 2319-30.
5. Beyhan, S., K. Bilecen, et al. (2007). "Regulation of rugosity and biofilm formation in Vibrio cholerae: comparison of VpsT and VpsR regulons and epistasis analysis of vpsT, vpsR, and hapR." *J Bacteriol* **189**(2): 388-402.
6. Casper-Lindley, C. and F. H. Yildiz (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." *J Bacteriol* **186**(5): 1574-8.
7. Chirwa, N. T. and M. B. Herrington (2003). "CsgD, a regulator of curli and cellulose synthesis, also regulates serine hydroxymethyltransferase synthesis in Escherichia coli K-12." *Microbiology* **149**(Pt 2): 525-35.
8. Romling, U., Z. Bian, et al. (1998). "Curli fibers are highly conserved between

Salmonella typhimurium and Escherichia coli with respect to operon structure and regulation." *J Bacteriol* **180**(3): 722-31.

9. R Gao, R. and A. M. Stock (2009). "Biological insights from structures of two-component proteins." *Annu Rev Microbiol* **63**: 133-54.
10. Lim, B., S. Beyhan, et al. (2006). "Cyclic-diGMP signal transduction systems in Vibrio cholerae: modulation of rugosity and biofilm formation." *Mol Microbiol* **60**(2): 331-48.
11. Beyhan, S. and F. H. Yildiz (2007). "Smooth to rugose phase variation in Vibrio cholerae can be mediated by a single nucleotide change that targets c-di-GMP signalling pathway." *Mol Microbiol* **63**(4): 995-1007.
12. Sudarsan, N., E. R. Lee, et al. (2008). "Riboswitches in eubacteria sense the second messenger cyclic di-GMP." *Science* **321**(5887): 411-3.
13. Beyhan, S., L. S. Odell, et al. (2008). "Identification and characterization of cyclic diguanylate signaling systems controlling rugosity in Vibrio cholerae." *J Bacteriol* **190**(22): 7392-405.
14. Kader, A., R. Simm, et al. (2006). "Hierarchical involvement of various GGDEF domain proteins in rdar morphotype development of Salmonella enterica serovar Typhimurium." *Mol Microbiol* **60**(3): 602-16.
15. Weber, H., C. Pesavento, et al. (2006). "Cyclic-di-GMP-mediated signalling within the sigma network of Escherichia coli." *Mol Microbiol* **62**(4): 1014-34.
16. Hickman, J. W. and C. S. Harwood (2008). "Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor." *Mol Microbiol* **69**(2): 376-89.

APPENDIX A

Supplemental Information

2.A1. Detailed Results and Discussion

Additional structural analysis

VpsT is a 25.8 kDa, conserved protein in *Vibrio* with sequence similarity to CsgD, FixJ and other LuxR family transcriptional regulators. Full-length VpsT from *V. cholerae* was purified to homogeneity by using standard liquid chromatography (see Material and Methods for details). Crystallization trials carried out with VpsT in the absence or presence of c-di-GMP yielded isomorphous crystals with similar unit cell dimensions, space group and diffraction properties (Table 2.S1).

For the nucleotide-free protein, crystals grown with selenomethionine-substituted protein diffracted X-rays to a maximal resolution of 3.1 Å. Data sets were collected at 3 wavelengths: the selenium anomalous scattering peak wavelength, the inflection, and a remote wavelength. The structure was solved by multi-wavelength anomalous dispersion (MAD) phasing, determining the position of 28 out of 32 selenium atoms. VpsT crystallized in the space group $P4_12_12$ with 4 molecules in the asymmetric unit (Fig. 2.12). The polypeptide chain of protomers A, B and C could be traced and built with high confidence into the experimentally phased maps. The REC domain of molecule D was also well resolved. The HTH domain of molecule D was largely disordered, and only the position of two of the 4 helices could be determined.

Cyclic di-GMP-bound VpsT diffracted X-rays to a resolution of 2.8 Å. The structure was solved by molecular replacement using a protomer of the nucleotide-free structure as the search model. Four protomers could be located in the asymmetric

unit with similar properties described above for apo-VpsT. The electron density maps revealed extra density at the base of the A-B and C-D dimer interfaces that was reminiscent of two stacked c-di-GMP molecules (Fig. 2.1C and 2.12A). A similar conformation of the nucleotide has been observed at the inhibitory site of diguanylate cyclases such as PleD and WspR, and in crystals of the nucleotide (*S9, S12-14*). Within a VpsT dimer, the two-fold symmetric c-di-GMP dimer stabilizes a two-fold symmetric protein assembly, using a similar binding mode as has been observed in other c-di-GMP-protein complexes that involves π-stacking interactions in addition to hydrogen bonds between the guanidinium groups of the binding site's arginine residues and the central guanine bases of c-di-GMP (*S9, S12*).

The asymmetric unit and crystal packing of the c-di-GMP-bound state is shown in Fig. 2.12A. In both, the nucleotide-free and c-di-GMP-bound crystals, the dimers A-B and C-D are formed via isologous contacts. The two equivalent, two-fold symmetric dimers (A-B and C-D) pack via a surface at the base of protomer A and C with an interfacial area of 519 $Å^2$. Energetic calculations yield a positive value for the free energy gain for the interaction mediated by this interface, suggesting that it is not biologically relevant. There are no interactions between the REC domains of protomers A and B with protomer D, and between the REC domains of protomers B and C. Crystal packing interactions involving the HTH domains were not considered since they would interfere with DNA binding.

Since the overall crystal packing contacts were preserved in the nucleotide-bound structure, it suggests that c-di-GMP stabilizes one of the dimers that can form under certain conditions such as in the crystallization drop. In the c-di-GMP-bound structure, there are minor adjustments of the packing, especially with regard to the c-di-GMP-mediated dimerization interface, but the protomers move as rigid bodies (Fig. 2.12B and 2.12C). Superposition of protomers indicates a rigid conformation of VpsT

with regard to the relative orientation of the REC and HTH domains (rmsd between 0.5 and 0.9 Å over all atoms) independent of c-di-GMP binding, with a buried surface area between the domains of 1355 $Å^2$, suggesting that such a conformation is unlikely to be due to crystal packing forces (Fig. 2.12B).

The c-di-GMP-stabilized dimer discussed here consists of protomers A and B, while the nucleotide-independent dimer is formed by protomer A and a protomer B from an adjacent asymmetric unit. Equivalent interactions occur in the crystals between molecules C and D, and their symmetry mates. These interactions between protomers are propagated throughout the crystal lattice, suggesting that VpsT can form higher-order oligomers with two polymerization sites per dimer (Fig. 2.12A).

The c-di-GMP-bridged protein interface has hydrophobic character and is largely stabilized by nucleotide binding to the base of α6 (Fig. 2.1C and 2.6B). An isoleucine residue at the center of the interface, located in α6, was mutated to glutamate in this study as a way to destabilize the formation of this dimer. A similar dimer involving α6 has been observed in the structure of a LuxR family two-component response regulator from *Aurantimonas sp. SI85-9A1* (PDB code 3cz5; Malashkevich *et al*.; unpublished) (Fig. 2.13A). In this case, phosphate and magnesium coordinating residues, as well as residues involved in the conformational switching, are similar to the response regulators PhoB and WspR suggesting that this REC domain is under control by phosphorylation (Fig. 2.13B). A bound phosphate ion at the putative phosphorylation site observed in the crystal structure supports this hypothesis. This response regulator lacks the VpsT c-di-GMP binding motif, but the structural resemblance and crystal packing contacts suggest a common mode for dimerization in this class of LuxR family response regulators, some of which appear to be regulated by ligand binding rather than phosphorylation.

The constitutive, c-di-GMP-independent dimer interface contains polar

interactions in addition to a hydrophobic groove at the bottom, a pocket accommodating the methionine residue crucial for formation of this dimer (M^{17}; Fig. 2.6A). In addition, the putative phosphorylation site (D^{60}; Fig. 2.6A) is part of this interface, which may explain why expression of $VpsT^{M17D}$, $VpsT^{D60A}$ or $VpsT^{D60E}$ in a Δ*vpsT* strain affects motility and spot morphology in a similar manner (Fig. 2.10). While mutations in the putative phosphorylation site were designed to mimic a constitutively inactive ($VpsT^{D60A}$) or a constitutively active ($VpsT^{D60E}$) state, they had indistinguishable activities in gene expression and motility assays, suggesting that phosphorylation does not play a major role in the regulation of VpsT.

Comparing the REC domain of VpsT to canonical REC domain-containing proteins such as PhoB, which are regulated by phosphorylation, suggests that VpsT employs a distinct mechanism (Fig. 2.13B). Many important residues for phospho-induced switching in other REC domains are not conserved in VpsT. While the phopshorylation site and one of the magnesium-coordinating aspartates are present in VpsT (D^{60} and D^{14}, respectively), it contains a serine-substitution in the position of the second magnesium-coordinating residue (S^{13}) (Fig. 2.14). The lysine residue that usually forms a salt bridge with the phosphate moiety in canonical REC domains is replaced with an aspartate residue (D^{112}). Only one of the switch residues is conserved (F^{109}), whereas the crucial threonine residue in PhoB (or serine in some other REC proteins) is a cysteine in VpsT (C^{90}). The switch residue F^{109} is involved in the conserved network of residues, and is engaged in a hydrophobic packing contact with the tryptophan residue of the c-di-GMP binding motif, possibly suggesting an allosteric path connecting the two dimer interfaces similar to that of canonical REC domains (Fig. 2.14B) (*S15*).

While the dimer interfaces are predominantly hydrophobic, the DNA binding site is largely polar. To model the protein-DNA interaction we used an alignment with

the HTH•DNA complex structure of NarL (PDB entry 1zg5) (*S16*), a response regulator, which shares 38% identity and a total of 85% sequence similarity with VpsT in its DNA-binding motif. Out of 24 residues participating in the formation of the protein-DNA interface in the NarL crystal structure (*S17*), 11 are identical and 10 show conserved or semi-conserved substitutions in VpsT. Most of these residues lie in helix α10, buried in the large groove of DNA, while the rest are interspersed in helices α8, α9, and α11, as well as in the connecting loop regions. Furthermore, all ten residues participating in direct hydrogen bond contacts with the DNA phosphate backbone or bases are conserved, with seven of them being identical between the two proteins. Substitutions were observed in two of the three residues forming hydrogen bonds with nucleobases in both VpsT and CsgD, presumably conferring specificity to the recognized DNA sequence (NarL-T^{183} is a serine in VpsT and CsgD; NarL-K^{192} is a histidine in VpsT and a tyrosine in CsgD). Thus, while consensus DNA motifs for NarL and CsgD binding have been proposed (*S18-20*), VpsT recognition sites remain to be experimentally determined.

Based on this analysis, helix α10 was identified as a structural motif that binds to the major groove of DNA (Fig. 2.15). Considering the c-di-GMP-independent VpsT dimer, superpositioning of the HTH domains with a HTH•DNA complex structure of the homolog NarL suggests a binding mode in which the DNA is bent in a 90°-angle, similar to that of transcriptional regulators such as the catabolite activator protein (*S16, S21*) (Fig. 2.15A). In contrast, the DNA molecules in the c-di-GMP-mediated VpsT dimer model run anti-parallel to each other, a configuration that would introduce DNA loops (Fig. 2.15B). DNA looping has been described as a mode of action for λ repressor and as a mechanism for the regulation of the *lac*, *gal* and *ara* operons (*S22, S23*). Notably, while AraC binds multiple operator sites to stabilize DNA loops, the catabolite activator protein participates in the regulatory network of

the *ara* operon, presumably by bending the DNA (*S24*). Based on its structure, VpsT would have the capacity to introduce both types of DNA deformations, bending and looping.

Given that the distribution of VpsT binding sites on the *Vibrio* chromosomes remains unknown, we cannot rule out the possibility that both the c-di-GMP dependent and independent interfaces could serve a regulatory function on DNA, introducing distinct deformation upon binding. In DNA mobility shift studies, stronger binding to three of the four *vpsL* promoter fragments was observed with $VpsT^{M17D}$, the mutant that cannot form the c-di-GMP-independent dimer, compared to wild-type VpsT (Fig. 2.2A). Although the c-di-GMP-independent dimer appears to affect DNA binding negatively for this particular promoter, its competence for DNA binding would depend on the distribution of VpsT recognition sequences across the *V. cholerae* genome. In addition to bending and looping DNA by dimeric VpsT, it would be feasible that VpsT forms higher order complexes on DNA upon c-di-GMP binding. A plausible tetramer, based on the VpsT crystal structure, would involve the c-di-GMP-mediated dimerization of VpsT dimers (Fig. 2.15C). Alternatively, two c-di-GMP-stabilized dimers could interact via the nucleotide independent interface.

Oligomerization of VpsT in solution

The dimeric, intercalated c-di-GMP conformation observed in the VpsT crystal structure has been also shown as prevalent under similar solution conditions (*S25*), and is hence unlikely to be due to crystal packing artifacts. Moreover, isothermal titration calorimetry (ITC) experiments revealed an apparent affinity for c-di-GMP (~3 μM) with a 1:1 binding stoichiometry and large unfavorable change in enthropy ($\Delta S \sim -34$ kcal/mole). Consistent with potential cooperativity between nucleotide binding and VpsT dimerization through the corresponding interface,

mutations destabilizing either interaction abolished c-di-GMP recognition ($VpsT^{R134A}$, $VpsT^{W131F}$, $VpsT^{T133V}$ and $VpsT^{I141E}$, respectively; Fig. 2.7). Conversely, VpsT carrying a disruptive mutation in the alternative dimerization interface, $VpsT^{M17D}$, bound c-di-GMP with similar to wild-type thermodynamic parameters (Fig. 2.7; Table 2.S2).

Elution of the nucleotide-free, wild-type VpsT showed a concentration-dependent shift in the protein peak during analytical size exclusion chromatography (Fig. 2.8, box), indicating conversion between different oligomeric species. In addition, incubation with c-di-GMP caused the protein to fall out of solution unless the salt concentration was raised (up to 600 mM NaCl depending on protein concentration) or xylitol (5-10%) was present. Salt-stabilized, c-di-GMP-bound protein eluted in a yet distinct fractionation volume, consistent with oligomerization through both crystallographic interfaces. Analysis of wild-type VpsT by analytical ultracentrifugation corroborated c-di-GMP-dependent oligomerization, with trimers being observed at higher concentration (Fig. 2.9). A further increase in protein concentration increased the heterogeneity of the sample and the tendency for protein aggregation.

In order to observe the differential effect of the two interfaces on c-di-GMP-dependent regulation of VpsT in solution, we resorted to SEC-coupled static multi-angle light scattering (MALS), a technique that reports the absolute molecular weight of macromolecules in any elution volume and independent of their shape (Fig. 2.8; Table S2). Experiments were conducted in high-salt buffer conditions (400-600mM NaCl) and at relatively high protein concentrations (~10 μg/μl or 400μM, injected concentration) in order to minimize c-di-GMP-driven polymerization and maximize association differences (see Fig. 2.8, box). While $VpsT^{M17D}$ in the absence of c-di-GMP existed exclusively as a monomer, wild-type VpsT, $VpsT^{R134A}$, and $VpsT^{I141E}$

exhibited molecular weights that were intermediate between those of a monomer and a dimer with relative concentration dependence throughout the elution peak. As reported for other systems (*S26*), such behavior is likely due to rapidly interchanging oligomeric species, with the constitutive, c-di-GMP-independent interface mediating sufficient dimerization on the SEC column. Cyclic di-GMP-mediated shifts to higher molecular weights were observed for $VpsT^{M17D}$ and wild-type VpsT, confirming a role for the second, c-di-GMP-dependent interface in signal-induced VpsT oligomerization. Shifts were more pronounced when c-di-GMP was included in the SEC buffer, compared to samples that were pre-incubated with c-di-GMP but analyzed in a mobile phase lacking the nucleotide (Fig. 2.8B-D, right column). As expected, mutants incapable of nucleotide recognition, $VpsT^{R134A}$ and $VpsT^{I141E}$, showed no change in oligomerization state in the presence or absence of the nucleotide.

Taken together, our solution data indicates a role for c-di-GMP in introducing a novel interaction interface in the VpsT quaternary structure. It is important to note that while cellular salt and protein concentrations might be significantly lower than the ones used in our *in vitro* studies, VpsT oligomerization *in vivo* would be facilitated by additional factors such as DNA binding, the architecture of available VpsT binding sequences on the *Vibrio* genome, and/or association with putative interacting partners among others.

Detailed analysis of the gene expression profiles

Whole genome expression profiling comparison of a Δ*vpsT* strain harboring wild-type VpsT or VpsT point mutants ($VpsT^{M17D}$, $VpsT^{D60A}$, $VpsT^{R134A}$ or $VpsT^{I141E}$) to that of cells harboring the pBAD vector allowed us to gain an insight into the complementation capacity of each clone. The overexpression levels were determined

to be moderate, about 1.5-fold above the expression of vpsT in the rugose strain, according to qPCR (Fig. 2.16).

Expression of wild-type *vpsT* led to the differential regulation of 54 genes (≥2-fold change in expression, 3% FDR): 31 induced and 23 repressed, respectively. A complete list of differentially regulated genes is provided in Table 2.S3. Besides genes involved in biofilm matrix production, which are discussed in the main text, transcription of a set of genes, predicted to be in a three gene operon (VC1583: *sodC*, encoding superoxide dismutase, 1.7-fold change; VC1585: *katB*, encoding catalase; VC1584: *ankB*, protein of unknown function), is increased in the Δ*vpsT* strain harboring wild-type VpsT. This observation suggests that enhanced oxidative stress resistance in rugose variants and cells grown in biofilms are likely to be controlled in part by VpsT.

VpsT with a mutation in either c-di-GMP binding ($VpsT^{R134A}$, $VpsT^{W131F}$, $VpsT^{R133V}$ or $VpsT^{M17D/R134A}$) or the c-di-GMP-dependent dimerization interface ($VpsT^{I141E}$) was unable to complement phenotypes associated with lack of *vpsT* (Fig. 2.2 and 2.3; Fig. 2.11). In contrast, a VpsT version unable to undergo c-di-GMP-independent dimerization ($VpsT^{M17D}$) was able to complement and even over-complement such phenotypes (Fig. 2.2). Expression the $VpsT^{M17D}$ led to the differential expression of 84 genes (≥2-fold change in expression, 3% FDR): 45 induced and 39 repressed, respectively. A complete list of regulated genes is provided in Table 2.S3. For instance, expression of genes in the *vps*-I cluster were 2-fold greater in a Δ*vpsT* strain harboring $VpsT^{M17D}$ relative to that harboring wild-type VpsT. Similarly, expression of greater number of flagellar biogenesis genes was downregulated in the $VpsT^{M17D}$-expressing Δ*vpsT* strain relative to that harboring wild-type VpsT. This result was confirmed by using a qPCR for transcript levels of *flaA*, one of the genes showing significant repression in the microarrays (Fig. 2.10A

and B). Transcription of *vpsR*, *cdgA* (VCA0074), VCA0075, *vpvA* and *vpvB*, genes whose products positively regulate biofilm matrix production, were also higher in a Δ*vpsT* strain harboring $VpsT^{M17D}$. Taken together, our results suggest that c-di-GMP-independent dimerization of VpsT could prevent DNA binding or contribute to the fine-tuning of the transcriptional response.

Comparison with VpsT homologs and CsgD

Vibrio parahaemolyticus, the causative agent of the most common *Vibrio*-associated, seafood-borne gastroenteritis, utilizes VpsT and VpsR homologs to regulate capsular polysaccharide (Cps) production and biofilm formation (*S27*). However, the function of these proteins in Cps production differs significantly from their counterparts in *V. cholerae*. For example, the VpsR homolog CpsR is not required for basal levels of *cps* expression but appears to induce *cps* gene expression in strains predicted to have elevated levels of c-di-GMP (*S28, S29*), which may indicate a similar, c-di-GMP-dependent regulation as has been described for FleQ (*S30*), a distant homolog of VpsR/CpsR in *P. aeruginosa*. The VpsT homolog in *Vibrio parahaemolyticus*, CpsS, negatively regulates *cps* transcription as *cpsS* deletion resulted in increased Cps production and super-rugose colony formation. Although the mechanisms by which CpsR and CpsS regulate *cps* transcription are yet to be determined, the presence of a VpsT-like c-di-GMP binding motif (W[F/L/M][T/S]R) in CpsS suggests that c-di-GMP may regulate CpsS function.

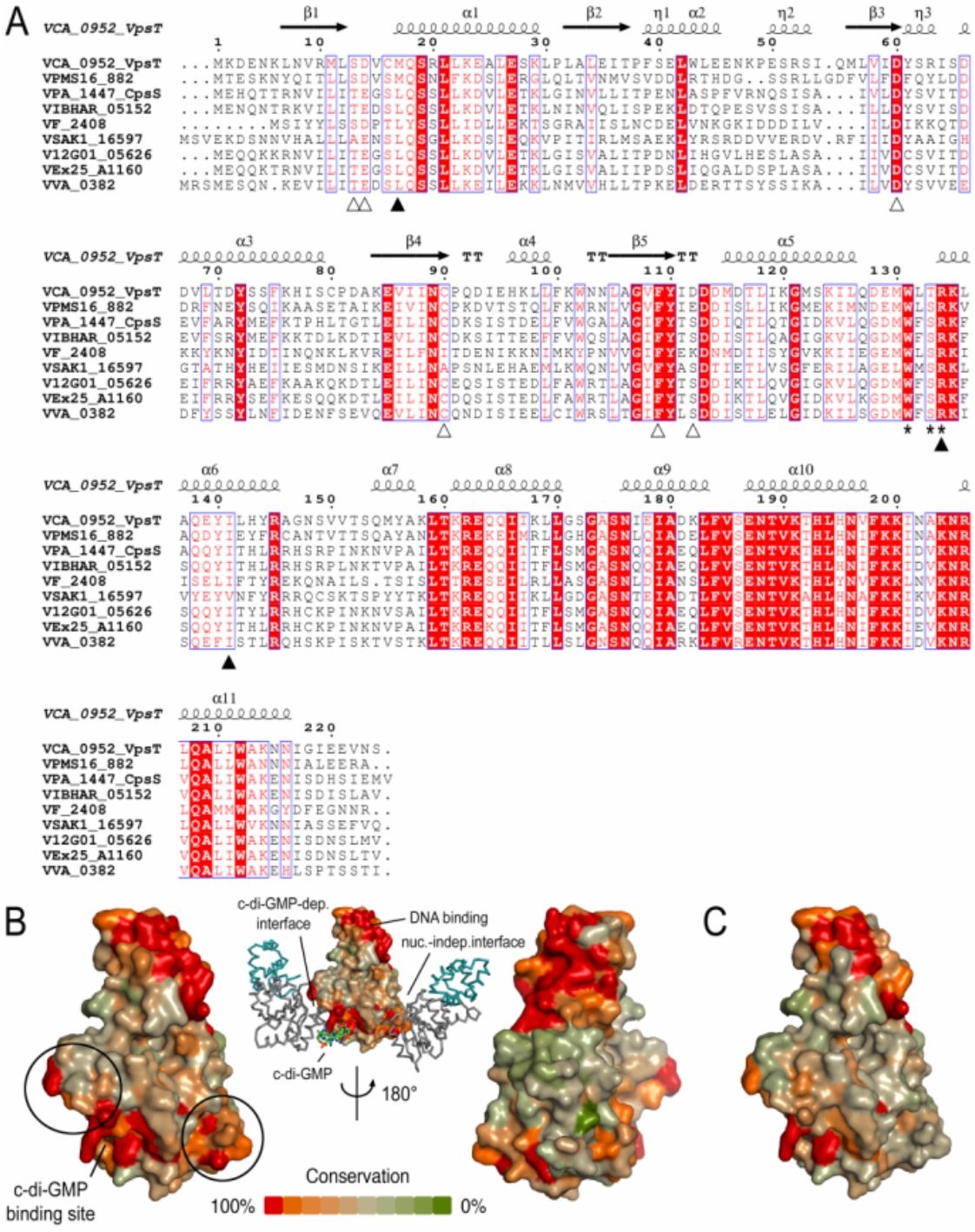
As mentioned before, VpsT is homologous to the transcriptional regulator CsgD from *E. coli* and *S. enterica serovar Typhimurium*. CsgD is required for the production of the two major extracellular matrix components, exopolysaccharides and proteinaceous fimbriae (curli), leading to the development of a unique colony morphology characterized by extensive corrugation and biofilm formation (*S31*). As a

consequence, *E. coli* and *S. enterica* mutants lacking CsgD produce flat and smooth colonies, similarly to *V.cholerae* mutants lacking VpsT (*S32, S33*).

CsgD and VpsT share an overall similar architecture harboring a C-terminal LuxR type HTH domain and an N-terminal REC domain with homology to FixJ/LuxR family response regulators of two-component signal transduction systems. Similar to VpsT, the REC domain of CsgD contains the conserved aspartate residue (D^{59} in CsgD) predicted to be phosphorylated, but lacks crucial residues necessary for phosphotransfer. The mechanism by which CsgD gets activated is yet unknown. As discussed in the main text, the c-di-GMP binding motif is not conserved in CsgD, but other small molecules may bind to its receiver domain to regulate CsgD activity (*S34*).

2.A2. Supplemental Figures

Figure 2.4: Conservation of VpsT homologs in related Vibrio species. (A) Sequence alignment of VpsT homologs from various *Vibrio* species generated with ClustalW2 (*S35*) and formatted with ESPript (*S36*). Key residues responsible for c-di-GMP binding, nucleotide-dependent, and constitutive dimerization are marked with closed arrows. Asterisks highlight residues directly involved in coordination of c-di-GMP. Open arrows mark the degenerate phosphorylation switch, i.e. residues involved in magnesium coordination, phosphorylation, and phosphotransfer-dependent conformational changes in canonical REC domains. The following sequences were used to generate the alignment: *Vibrio cholerae* O1 biovar *El Tor* str. N16961 (NP_233336), *Vibrio parahaemolyticus* 16 (ZP_05117817), *Vibrio parahaemolyticus* RIMD 2210633 (NP_800957), *Vibrio harveyi* ATCC BAA-1116 (ABU73058), *Vibrio fischeri* ES114 (YP_205791), *Vibrio shilonii* AK1 (ZP_01865459), *Vibrio alginolyticus* 12G01 (ZP_01258564), *Vibrio sp.* Ex25 (ZP_04922131), *Vibrio vulnificus* YJ016 (NP_936438).
(B) Sequence conservation mapped onto the solvent-accessible surface of VpsT. Conservation is presented as a color gradient from green to red (0%-100% conservation). Arrows and circles highlight structurally and functionally important motifs.
(C) Surface conservation mapping based on alignment including VpsT-like protein sequences. As previously discussed, the latter carry a proline substitution at position 3 of the c-di-GMP binding pocket (W[L/F/M]PR) (not shown in the alignment in A). The surface shows significantly less conservation, especially in the interaction interfaces suggesting subfamily-specific dimerization.


A
VCA_0952_VpsT
VCA_0952_VpsT
VPMS16_882
VPA_1447_CpsS
VIBHAR_05152
VF_2408
VSAK1_16597
V12G01_05626
VEx25_A1160
VVA_0382
B
c-di-GMP-dep. interface
DNA binding
nuc.-indep.interface
c-di-GMP
180°
c-di-GMP binding site
Conservation
100%
0%
C

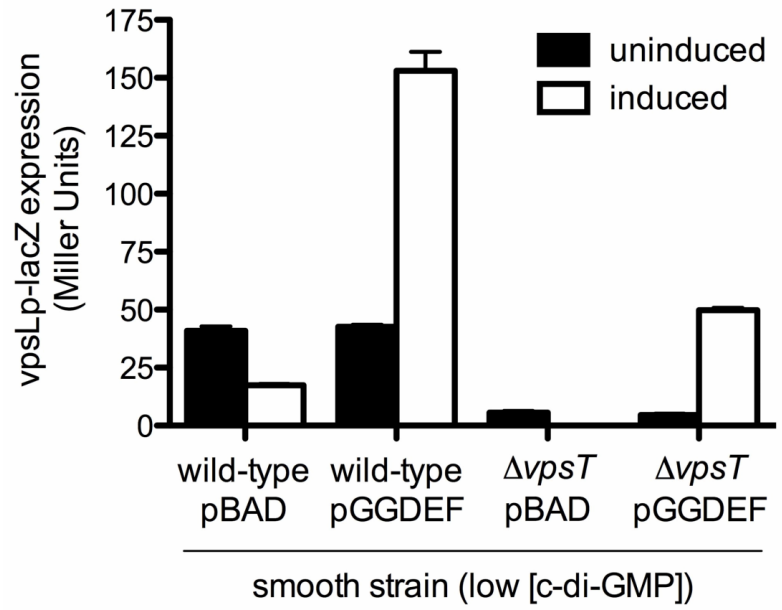
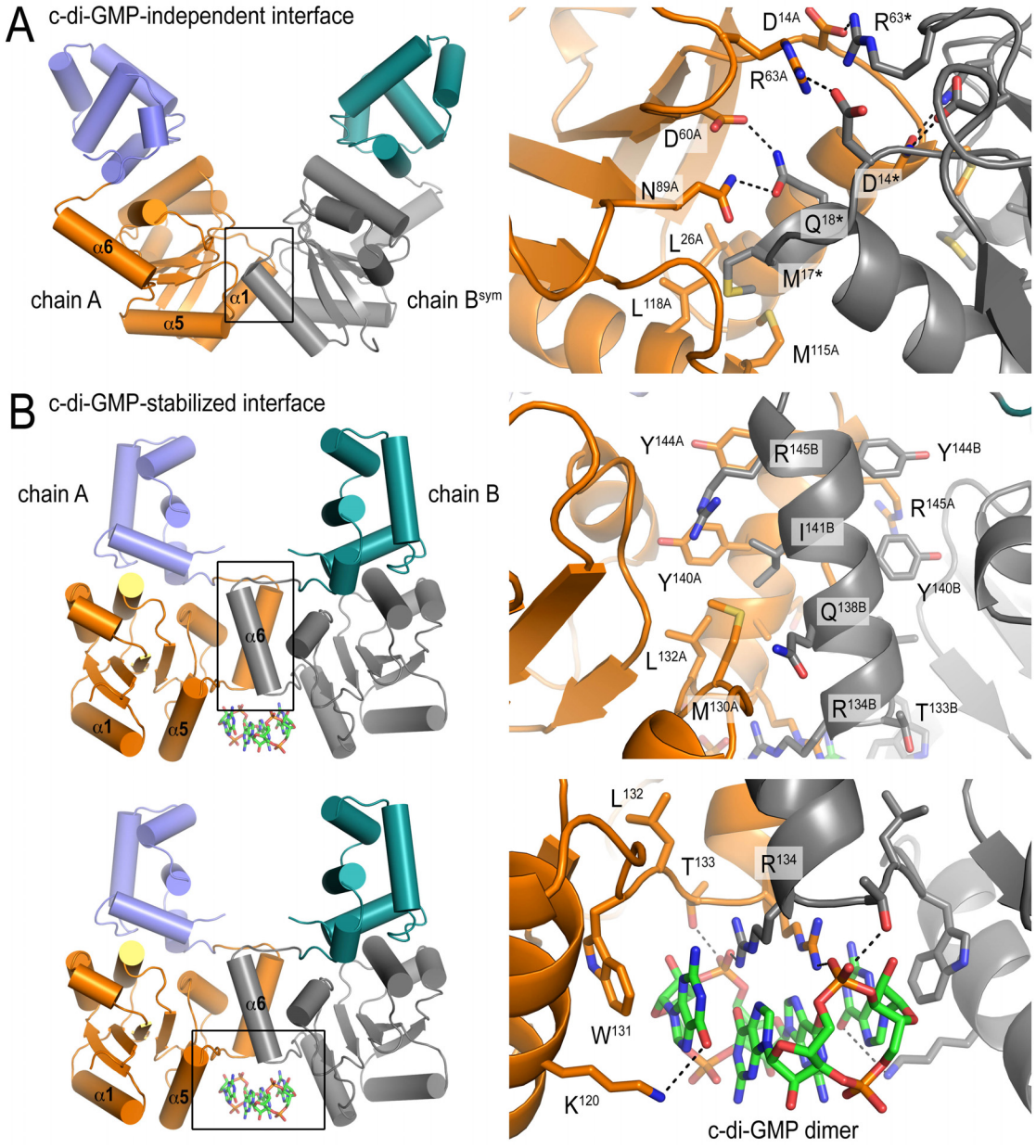


Figure 2.5: Cyclic di-GMP dependence of VpsT-mediated gene transcription. *vpsL* gene expression in different genetic backgrounds harboring a single-copy chromosomal *vpsLp-lacZ* fusion. Data are mean of 8 replicates -/+ SD. In this experiment, a smooth wild-type or *vpsT*-deletion strain was used that have low levels of c-di-GMP (*S2*). Induced expression of an active diguanlylate cyclase increases cellular c-di-GMP concentration (*S2, S37*). *vpsL* expression was driven to a large extent by VpsT in a c-di-GMP-dependent manner. A smaller fraction is independent of VpsT but requires an elevated c-di-GMP level. Although the analysis may be complicated by the usage of different genetic backgrounds, which may explain the differences in magnitude of *vpsL* expression comparing uninduced and induced samples, the data is supported by and consistent with other experiments described here that demonstrate that c-di-GMP and VpsT act in concert (Data courtesy of Jiunn C.N. Fong, Nicholas J. Shikuma, and Fitnat H. Yildiz).

Figure 2.6: Dimerization and nucleotide-binding interfaces.
(A) The c-di-GMP-independent dimerization interface. A crystallographic dimer is shown in cartoon presentation (chain A-chain B^{sym}) (left). The close-up view shows the dimerization interface (right). Interfacial residues are shown as sticks and labeled appropriately, where A and asterisk (*) identify residues belonging to chain A and chain B^{sym}, respectively. Pair-wise interactions spanning the interface are also denoted. **(B)** c-di-GMP-dependent interactions. A crystallographic dimer formed by two protomers in one asymmetric unit (chain A-chain B) is shown in cartoon presentation (left). The interaction is mediated primarily through helices $\alpha 6$ of each chain and is stabilized by the coordination of an intercalated c-di-GMP dimer at the base. A close-up view shows the protein interface (top-right). The second close-up view shows the c-di-GMP binding pocket (bottom-right). Protein interface residues, as well as residues participating in electrostatic and π-stacking interactions with the nucleotide are depicted as sticks and appropriately labeled. Conservation of key residues participating in the formation of each interface is shown in the table at the bottom.



	17	131	132	133	134	141
VpsT (e.g. VCA0952)	[M/L]	W	[L/F/M]	[T/S]	R	[I/V]
VpsT-like (e.g. VC0396)	n.c.	W	[L/F/M]	P	R	[I/L]
CsgD (E. coli)	L	Y	F	[T/S]	Q	I

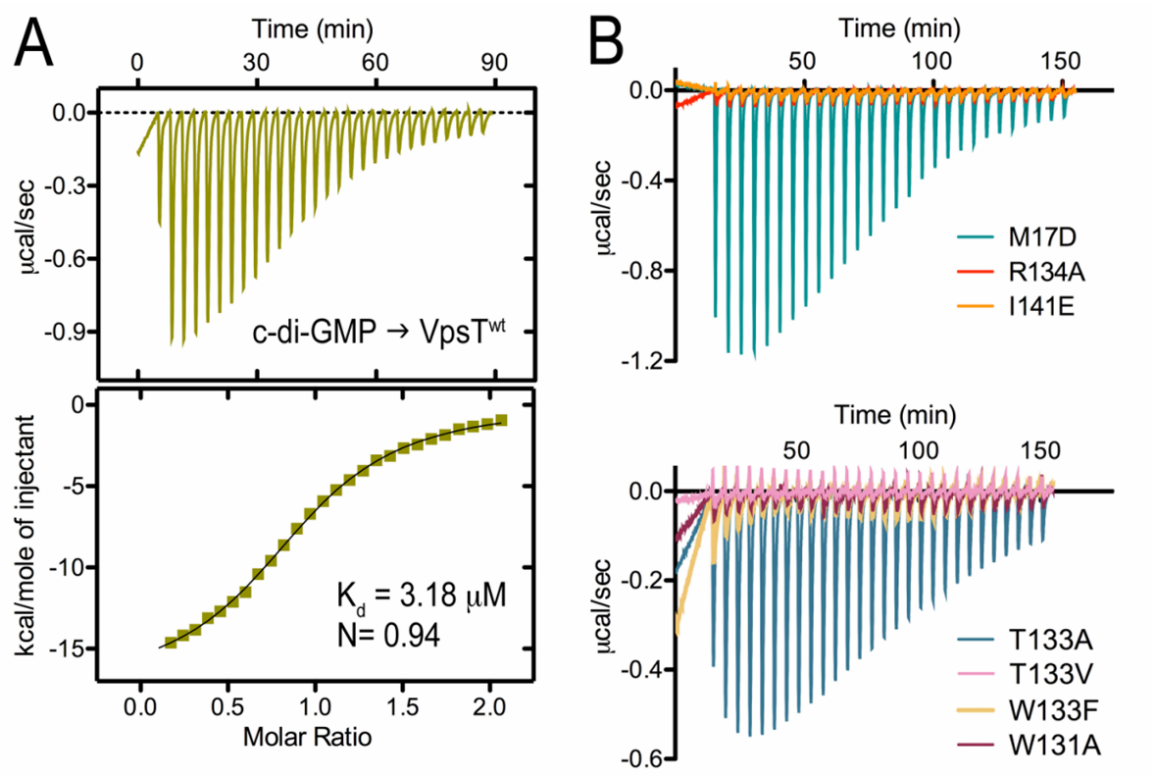
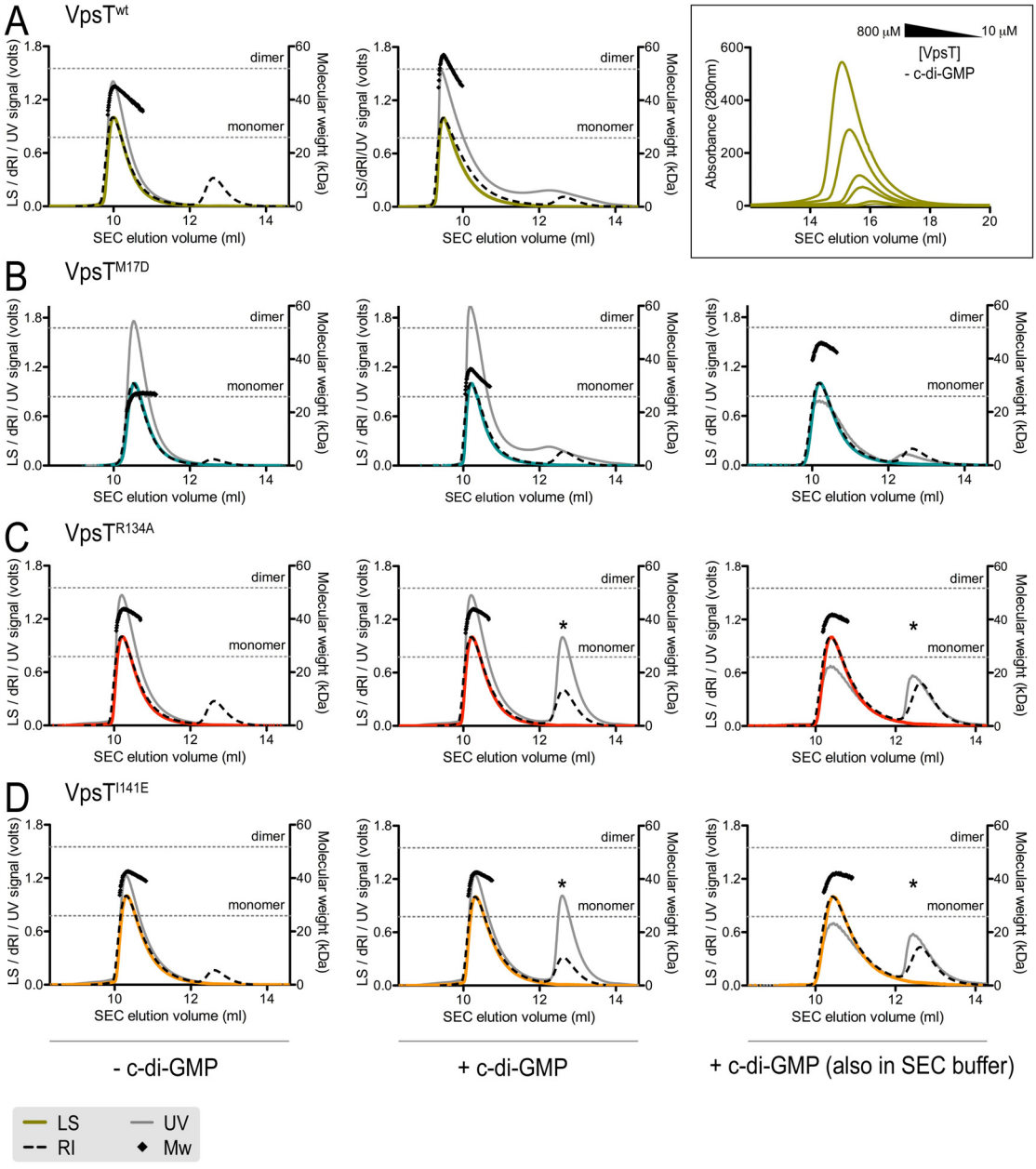


Figure 2.7: Isothermal titration calorimetry supports c-di-GMP binding *in vitro.* **(A)** Wild-type (wt) VpsT binds c-di-GMP with an apparent affinity in the low micromolar range (K_d ~ 3.2 µM). Binding occurs with a 1:1 stoichiometry in an entropically unfavorable (ΔS ~ -34 kcal/mole) exothermic reaction (ΔH ~ $-1.7x10^4$ kcal/mole). The top panel shows baseline-corrected data collected at 20°C, while the bottom shows the results of curve-fitting using a single/independent site binding model.
(B) Baseline corrected data collected at 20°C for the $VpsT^{M17D}$, $VpsT^{R134A}$, $VpsT^{I141E}$, $VpsT^{W131F}$, $VpsT^{W131A}$, $VpsT^{T133A}$ and $VpsT^{T133V}$ mutants. While $VpsT^{M17D}$ complexes c-di-GMP with similar to wild-type affinity (K_d ~ 2.8 µM) and $VpsT^{T133A}$ with slightly reduced affinity (K_d ~ 7.4 µM), nucleotide binding was not detectable for the other VpsT mutants.

Figure 2.8: VpsT oligomerization state in solution.
(A) Wild-type VpsT. The gel filtration profile of nucleotide-free, wild-type (wt) VpsT is characterized by a concentration-dependent shift in the elution peak (box). The protein concentration range is indicated as injected onto the column. SEC-coupled multi-angle light scattering analysis of wild-type VpsT in presence (middle) and absence (left) of c-di-GMP are shown. The signal from the 90°-scattering detector is shown in color, the signal from the refractive index detector is shown as dashed line and the UV absorbance is plotted in grey. Average molecular weights are plotted in black against the right Y-axis, as calculated every second across the protein elution peak. Theoretical molecular weights corresponding to those of a monomer and a dimer are indicated as horizontal dashed, grey lines. Rapid equilibria between monomers and dimeric assemblies are detected as species with intermediate molecular weights. Experiments evaluating the effects of c-di-GMP were conducted after pre-incubation with excess nucleotide, followed by separation in a mobile phase lacking c-di-GMP. Injected protein and nucleotide concentrations were 400 μM and 600 μM, respectively.
(B) $VpsT^{M17D}$. The mutant was analyzed as described above. In addition, experiments were carried out with 40 μM c-di-GMP in running buffer (right) to stabilize the VpsT•c-di-GMP complex during SEC.
(C) $VpsT^{R134A}$. The mutant was analyzed as described in A and B. Asterisks denote small molecule peaks in the UV absorbance, plausibly due to the elution of unbound nucleotide.
(D) $VpsT^{I141E}$. The mutant was analyzed as described above. Results were similar to those obtained with $VpsT^{R134A}$. Results from SEC-coupled multi-angle light scattering and ITC are summarized in Table S2. The monomeric molecular weight obtained with $VpsT^{M17D}$ compared to higher molecular weights for the rest of the mutants and the wild-type protein indicates constitutive VpsT dimerization through the nucleotide-independent interface in the absence of nucleotide. Increased molecular weights and shifts in the elution peaks for the wild-type and $M^{17}D$ proteins in the presence of c-di-GMP are indicative of a nucleotide-dependent change in oligomerization through the corresponding interface.


A
VpsT^wt
B
VpsT^M17D
C
VpsT^R134A
D
VpsT^I141E
dimer
monomer
LS / dRI / UV signal (volts)
Molecular weight (kDa)
SEC elution volume (ml)
800 µM
10 µM
[VpsT]
- c-di-GMP
Absorbance (280nm)
*
- c-di-GMP
+ c-di-GMP
+ c-di-GMP (also in SEC buffer)
LS
UV
RI
Mw

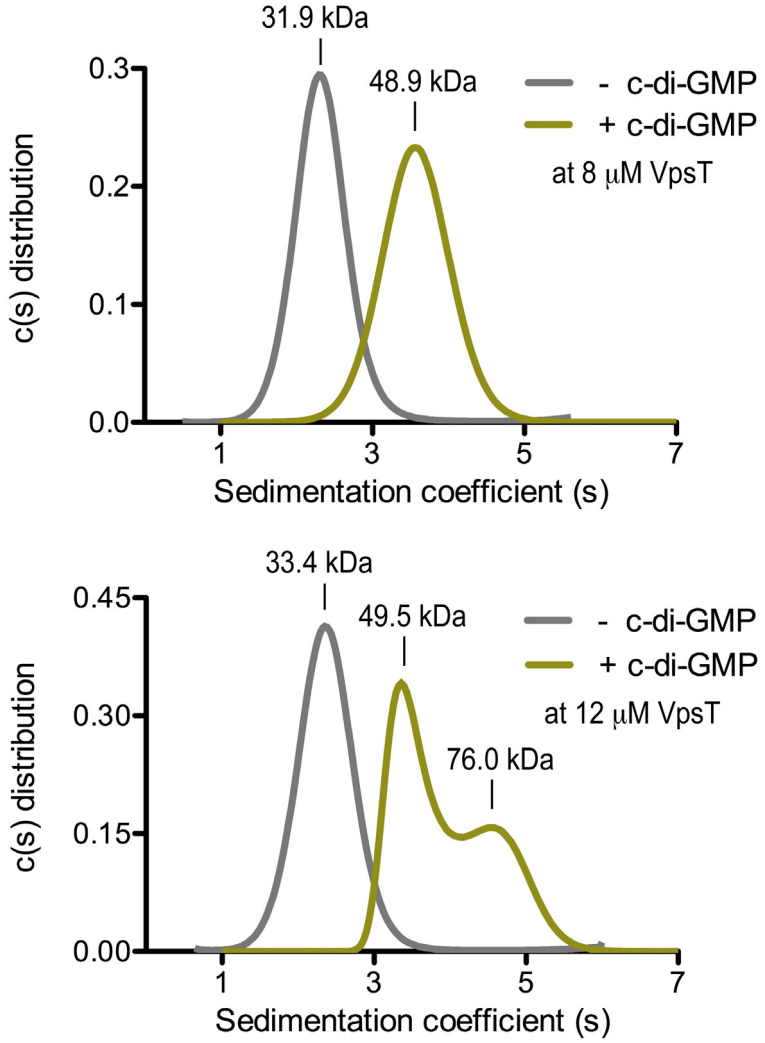
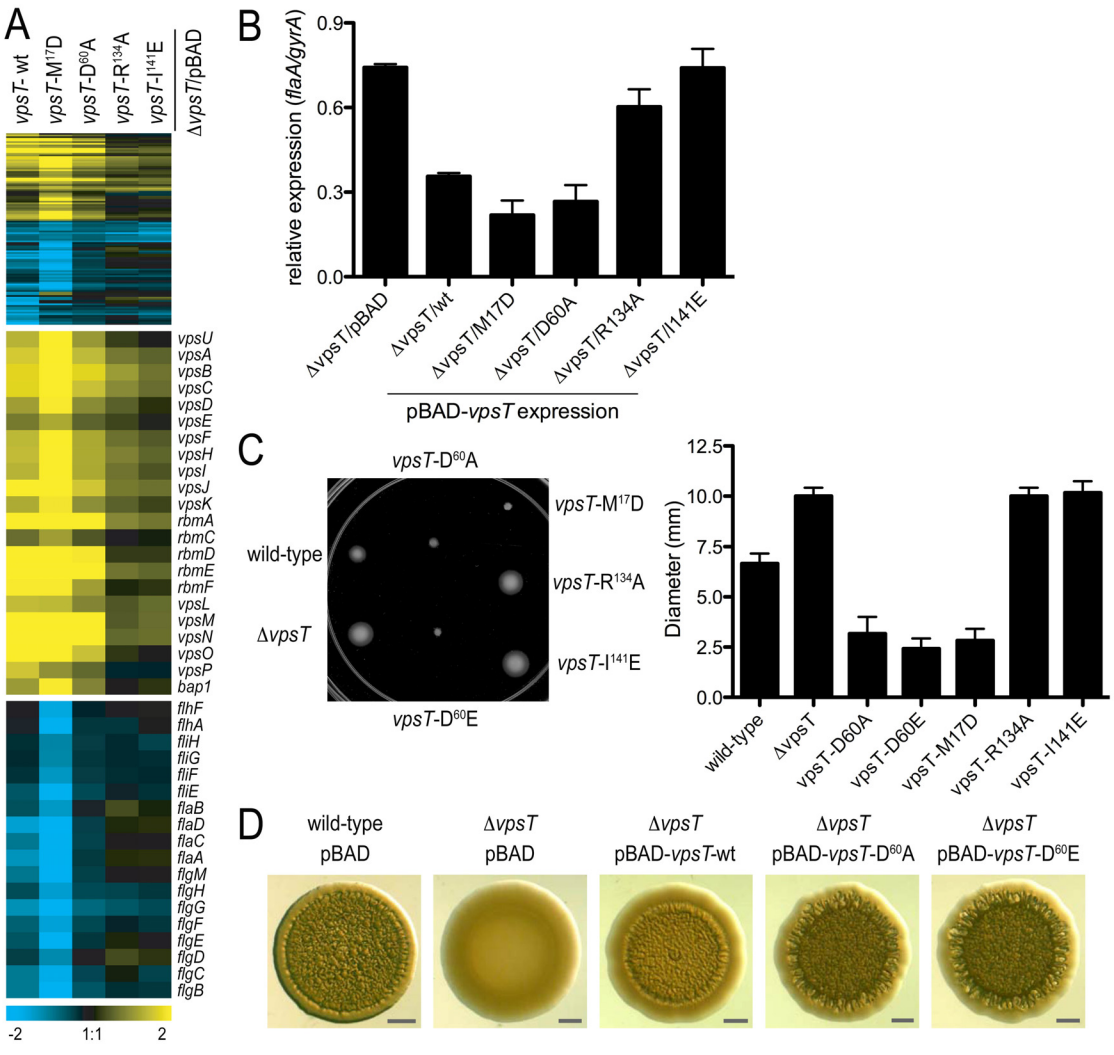


Figure 2.9: Analytical ultracentrifugation of wild-type VpsT. VpsT was analyzed by sedimentation velocity analytical ultracentrifugation at 8 and 12 μM protein concentration in the presence and absence of c-di-GMP (25 μM). Molecular weights were analyzed by using the program SedFit.

Figure 2.10: Functional analysis of VpsT with mutations at the putative phosphorylation site (Data courtesy of Jiunn C.N. Fong, Nicholas J. Shikuma, and Fitnat H. Yildiz).
(A) Whole genome expression profiling. Top, compact heatmap in a log2-based pseudocolor scale (yellow, induced; blue, repressed) comparing differentially expressed genes in a Δ*vpsT* strain expressing wild-type (wt) or mutated VpsT versions compared to the vector control (midpanel, expression profiles of *vps*-I and *vps*-II clusters and genes encoding matrix proteins; bottom, expression profiles of flagellar biosynthesis genes)
(B) Quantitative PCR for *flaA* expression. Expression of *flaA* in a *vpsT*-deletion strain (Δ*vpsT*) expressing wild-type (wt) and mutant VpsT proteins from a pBAD plasmid. Concentration of the inducer arabinose is 0.1% as used in the β-galactosidase and expression profiling assays. Expression of the housekeeping *gyrA* gene is used for normalization. Data are mean of 3 replicates -/+ SD.
(C) Motility phenotypes on semisolid LB agar plates. For strains expressing mutants of VpsT, single chromosomal insertion mutants are shown. The graph shows the mean migration zone diameter of each strain. Data are mean of 8 replicates -/+ SD.
(D) Spot morphology. A wild-type rugose strain carrying the vector (pBAD) and Δ*vpsT* strains carrying the vector or plasmids containing wild-type or mutant *vpsT* are shown (bars=1 mm). Mutations in the putative phosphorylation site designed to produce a constitutively inactive or active state, $VpsT^{D60A}$ and $VpsT^{D60E}$, respectively, show identical phenotypes indicating that phosphorylation is unlikely a regulatory mechanism for VpsT function.


A
vpsT- wt
vpsT-M17D
vpsT-D60A
vpsT-R134A
vpsT-I141E
ΔvpsT/pBAD
vpsU
vpsA
vpsB
vpsC
vpsD
vpsE
vpsF
vpsH
vpsI
vpsJ
vpsK
rbmA
rbmC
rbmD
rbmE
rbmF
vpsL
vpsM
vpsN
vpsO
vpsP
bap1
flhF
flhA
fliH
fliG
fliF
fliE
flaB
flaD
flaC
flaA
flgM
flgH
flgG
flgF
flgE
flgD
flgC
flgB
-2
1:1
2
B
relative expression (flaA/gyrA)
0.9
0.6
0.3
0.0
ΔvpsT/pBAD
ΔvpsT/wt
ΔvpsT/M17D
ΔvpsT/D60A
ΔvpsT/R134A
ΔvpsT/I141E
pBAD-vpsT expression
C
vpsT-D60A
vpsT-M17D
wild-type
vpsT-R134A
ΔvpsT
vpsT-I141E
vpsT-D60E
Diameter (mm)
12.5
10.0
7.5
5.0
2.5
0.0
wild-type
ΔvpsT
vpsT-D60A
vpsT-D60E
vpsT-M17D
vpsT-R134A
vpsT-I141E
D
wild-type pBAD
ΔvpsT pBAD
ΔvpsT pBAD-vpsT-wt
ΔvpsT pBAD-vpsT-D60A
ΔvpsT pBAD-vpsT-D60E

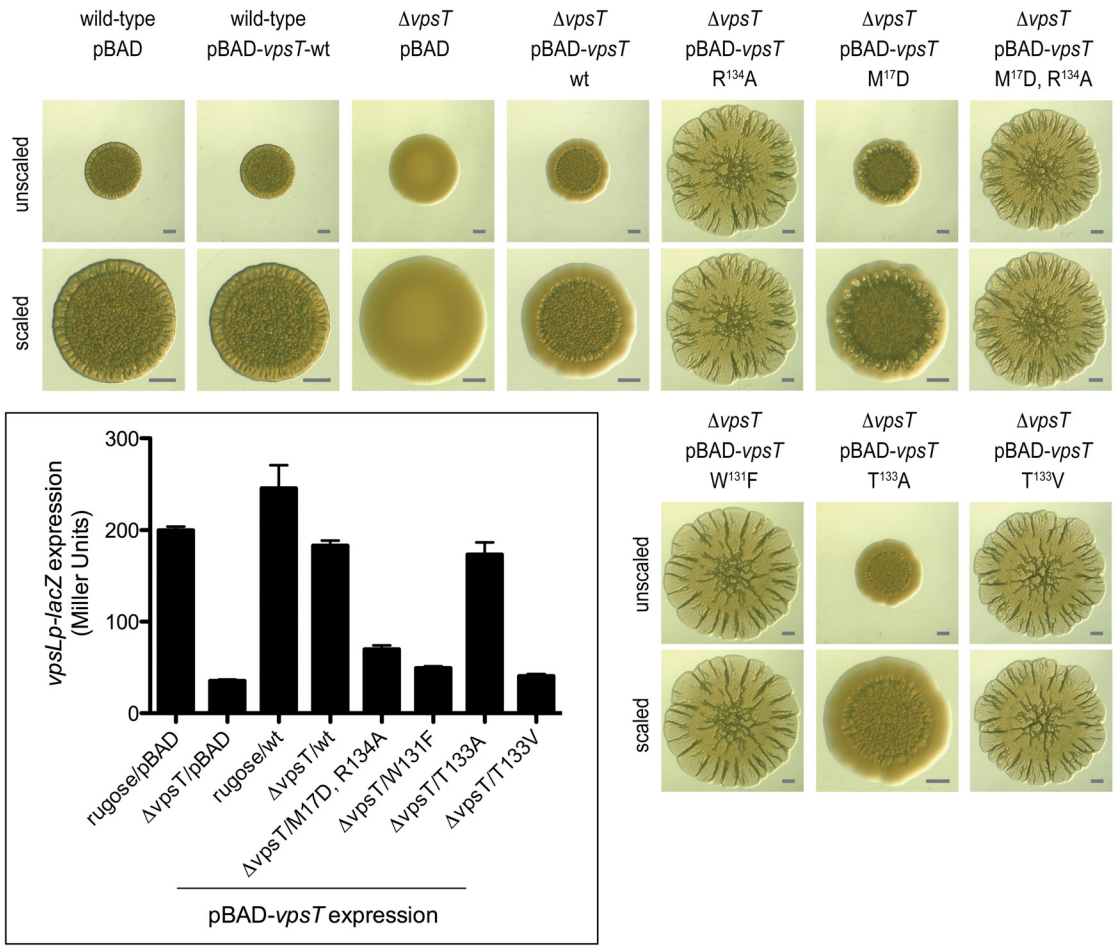
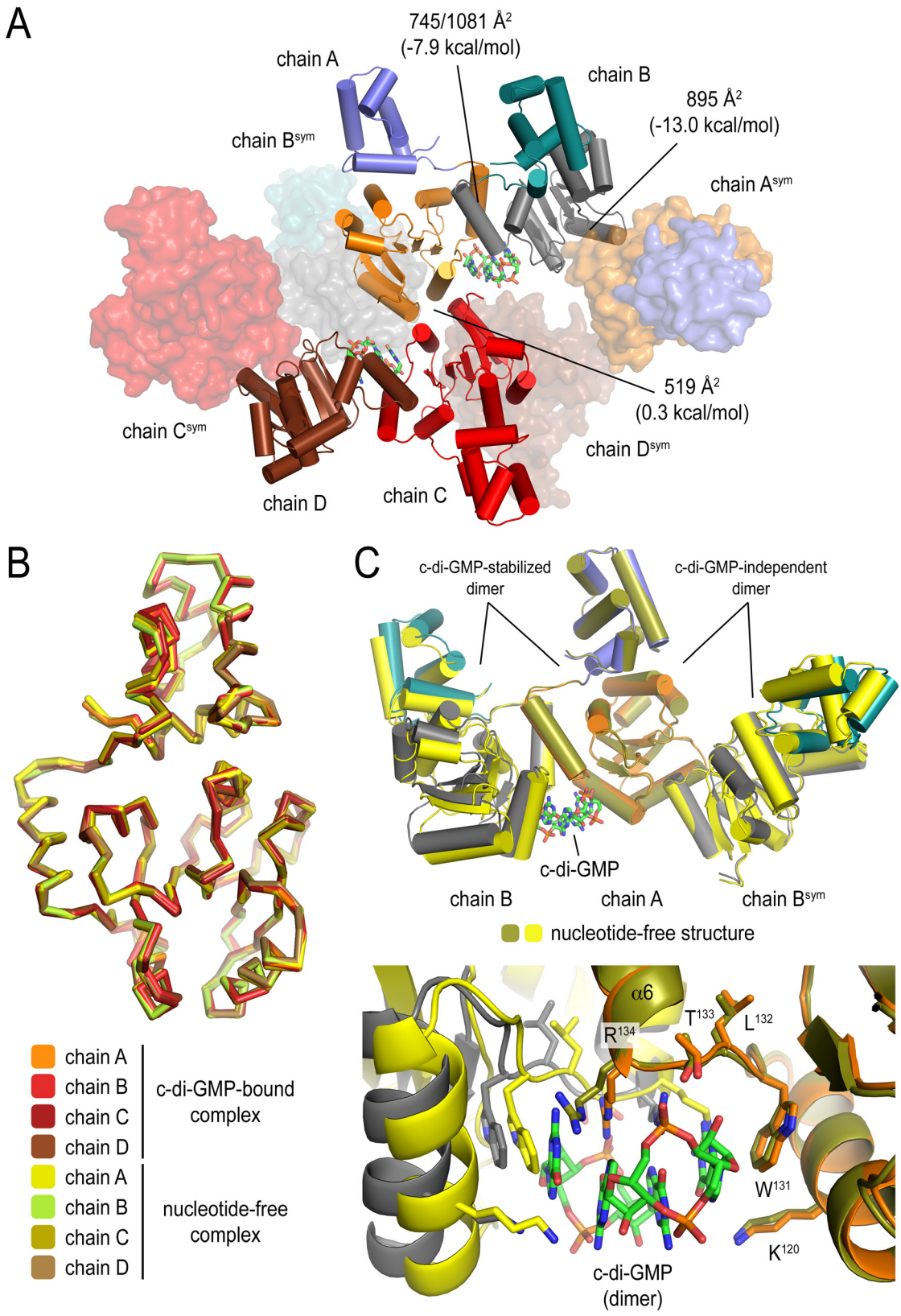
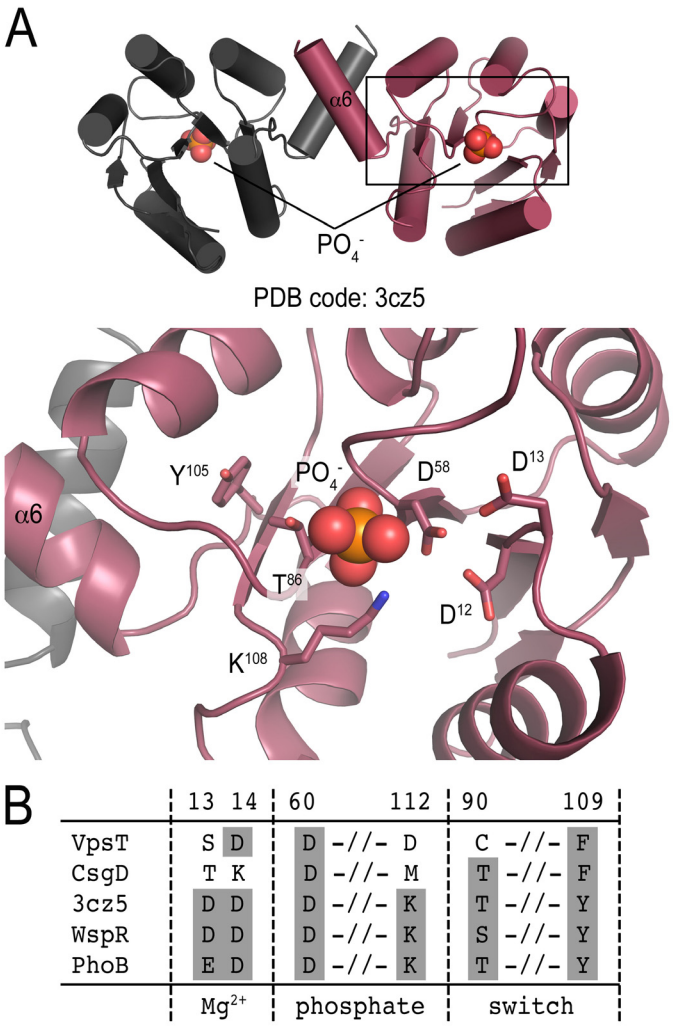


Figure 2.11: Spot morphology phenotypes for additional nucleotide binding mutants of VpsT. A wild-type rugose strain carrying the vector (pBAD) or a plasmid containing wild-type VpsT, and Δ*vpsT* strains carrying the vector or plasmids containing wild-type or mutant *vpsT* are shown (top, unscaled; bottom, scaled to similar diameter, bars=1 mm). The double VpsT$^{M17D/R134A}$ mutant shows a phenotype observed for the single VpsT^{R134A} and VpsT^{I141E} mutants, indicating that c-di-GMP-driven dimerization is dominant in VpsT function. The boxed inset shows *vpsL* gene expression in different genetic backgrounds harboring a single-copy chromosomal *vpsLp-lacZ* fusion. Data are mean of 8 replicates -/+ SD (Data courtesy of Jiunn C.N. Fong, Nicholas J. Shikuma, and Fitnat H. Yildiz).

.

Figure 2.12: Comparison of c-di-GMP-bound and nucleotide-free VpsT.
(A) Asymmetric unit and crystal packing. The asymmetric unit contains four VpsT protomers (cartoon presentation), and is shown for the c-di-GMP-bound VpsT crystal. Adjacent symmetry mates are shown in transparent surface presentation illustrating the polymerization of VpsT in the crystals mediated by the c-di-GMP-stabilized and nucleotide-independent interfaces. Interfacial surface areas and interaction free energy gain estimations were calculated using the PISA server (*S17*).
(B) Structural comparison of VpsT promoters. Superposition of the α-carbon backbone of all 8 symmetry-unrelated protomers in the unliganded and complex crystal structures of VpsT shows almost identical protein conformation (rmsd of 0.5-0.9Å over all atoms).
(C) Structural comparison of oligomeric assemblies of c-di-GMP-bound and nucleotide-free VpsT. Superposition of the corresponding crystallographic trimers using molecule A as the reference shows minor adjustments in the packing (top), where the VpsT protomers move as rigid bodies. A close-up view of the c-di-GMP binding site shows alternative rotamer conformation for arginine R^{134}, as well as narrowing of the nucleotide-binding pocket in the apo-state (bottom).


A
745/1081 Å²
(-7.9 kcal/mol)
chain A
chain B
895 Å²
(-13.0 kcal/mol)
chain B^sym
chain A^sym
519 Å²
(0.3 kcal/mol)
chain C^sym
chain D^sym
chain D
chain C
B
chain A
chain B
chain C
chain D
c-di-GMP-bound complex
chain A
chain B
chain C
chain D
nucleotide-free complex
C
c-di-GMP-stabilized dimer
c-di-GMP-independent dimer
c-di-GMP
chain B
chain A
chain B^sym
nucleotide-free structure
α6
T133
L132
R134
W131
K120
c-di-GMP (dimer)



B

	13	14	60		112	90		109
VpsT	S	D	D	-//-	D	C	-//-	F
CsgD	T	K	D	-//-	M	T	-//-	F
3cz5	D	D	D	-//-	K	T	-//-	Y
WspR	D	D	D	-//-	K	S	-//-	Y
PhoB	E	D	D	-//-	K	T	-//-	Y
	Mg^{2+}		phosphate			switch		

Figure 2.13: Structure of a REC domain structurally related to VpsT.
(A) Crystal structure of the receiver domain of a LuxR-like response regulator from *Aurantimonas sp. SI85-9A1* (PDB code 3cz5; Malashkevich et al.; unpublished). A DALI search identified the aforementioned protein as a structural homolog to VpsT (*S38*). Similarly to VpsT, this receiver domain contains an additional helix α6, mediating dimerization contacts in the crystal lattice (top). However, no ligand is stabilizing the interaction, and a bound phosphate ion at the putative phosphorylation site (bottom) suggests control by phosphorylation. Residues involved in magnesium binding, phosphorylation, and phosphotransfer-dependent conformational changes are depicted as sticks and appropriately labeled. The phosphate ion bound at the putative phosphorylation site is shown as spheres.
(B) Conservation of residues involved in phosphotransfer as compared to canonical (WspR and PhoB) and divergent (CsgD and VpsT) receiver domains. Residue numbers in the table correspond to the VpsT sequence.

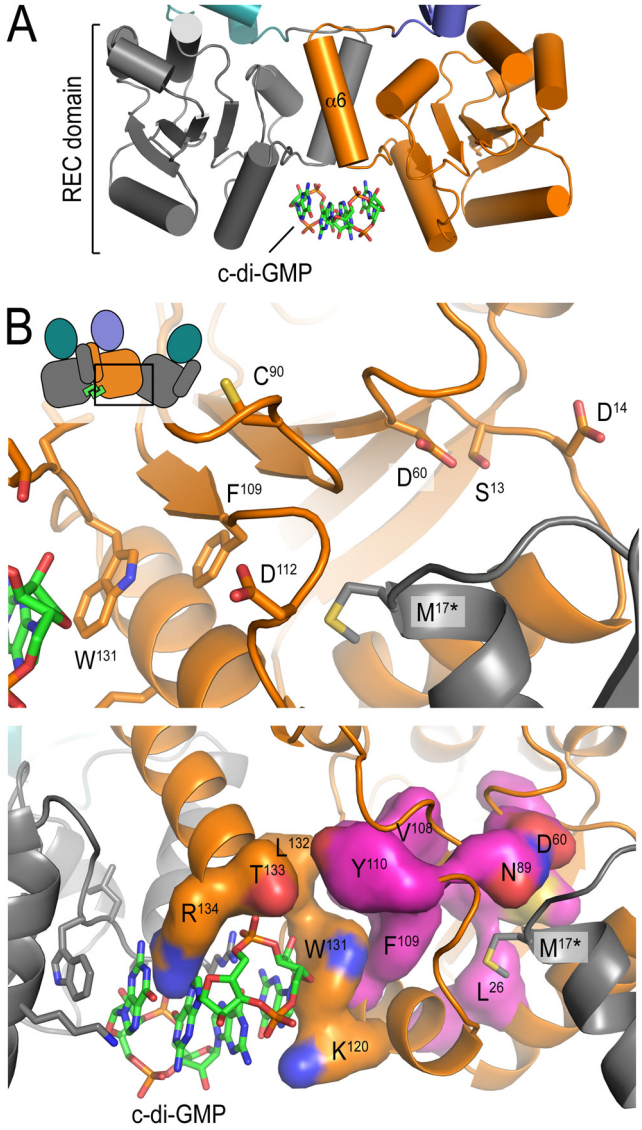
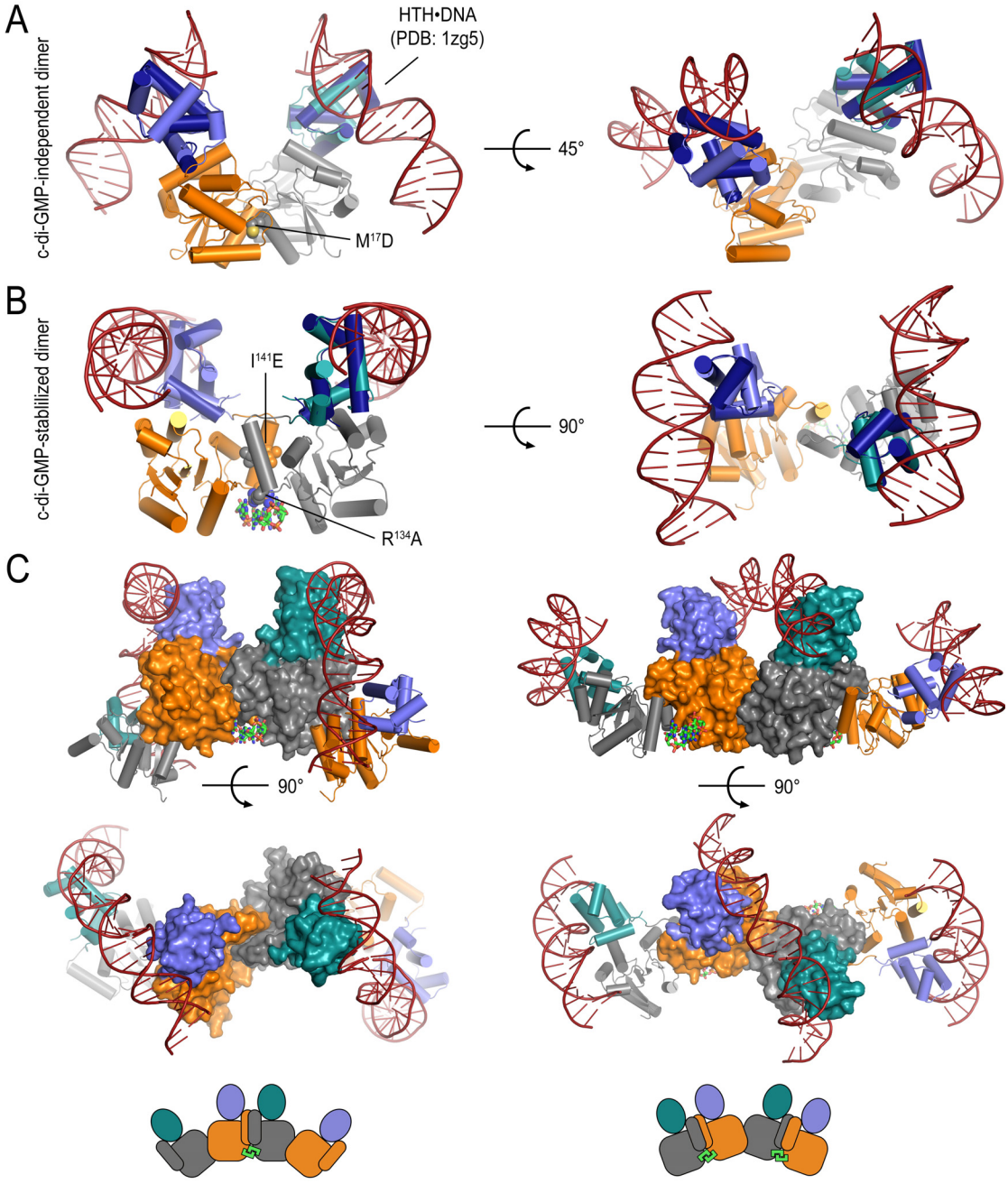


Figure 2.14: Conserved residues form a path through the REC domain connecting the c-di-GMP-independent interface and the c-di-GMP binding site. **(A)** Structure of the receiver domain dimer showing c-di-GMP bound at the base of helices α6.

(B) Close-up view of the putative phosphorylation site. Residues involved in phosphotransfer in canonical receiver domains are shown as sticks, as well as residue M^{17} from an adjacent molecule, stabilizing the constitutive dimer interface, and residue W^{131} from the c-di-GMP binding pocket (top). Conserved residues connecting the c-di-GMP-independent interface and the nucleotide binding site are shown in magenta. Interestingly, these include the degenerate phosphorylation site together with residues involved in phosphotransfer-related conformational switching, suggesting that they might form an allosteric path for phosphorylation-dependent α6 dimerization in other LuxR-like response regulators (see Figure 2.13).

Figure 2.15: Models for DNA binding to VpsT.
(A) Nucleotide-independent VpsT dimer. The dimer is shown in two views that are separated by a 45°-rotation along the x-axis. The structure of the HTH domain of NarL bound to cognate DNA (*S16*) was superimposed onto the HTH domain of VpsT to illustrate the DNA binding mode of VpsT dimers.
(B) Cyclic di-GMP-stabilized VpsT dimer. The dimer is shown in two orthogonal views. DNA binding was modeled as described above.
(C) Tetrameric models of VpsT•DNA complexes. Based on the crystal packing contacts, two plausible tetrameric assemblies can be constructed. In the left panel, c-di-GMP bridges two nucleotide-independent dimers. The DNA was used from a structure of the catabolite activator protein, a transcriptional regulator that bends DNA in a 90°-angle (*S21*). In the right panel, two c-di-GMP-driven dimers associate via the nucleotide-independent dimerization interface of VpsT. The central DNA is taken from a catabolite activator protein•DNA complex structure, the lateral DNA fragments are taken from a NarL•DNA complex structure. The distribution of VpsT binding sites on the chromosomes is likely to determine the feasibility of these models for transcriptional regulation.


A
c-di-GMP-independent dimer
HTH•DNA
(PDB: 1zg5)
M17D
45°
B
c-di-GMP-stabililized dimer
I141E
R134A
90°
C
90°
90°

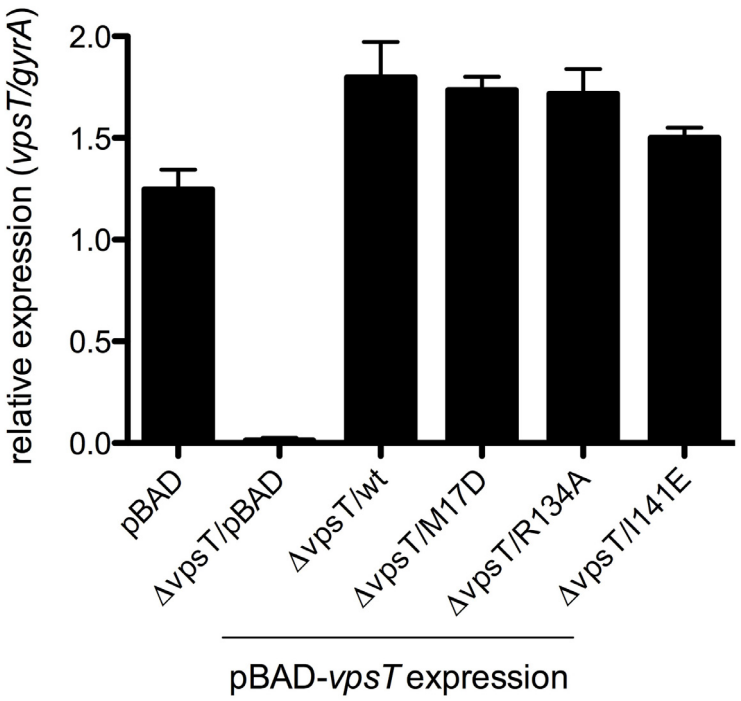


Figure 2.16: Quantitative PCR (qPCR) results show modest VpsT over-expression. Expression of wild-type (wt) and mutant VpsT proteins from a pBAD plasmid introduced in a *vpsT*-deletion strain (Δ*vpsT*) is compared to chromosome-driven VpsT expression in the wild-type rugose strain carrying an insert-less vector. Concentration of the inducer arabinose is 0.1% as used in the β-galactosidase and expression profiling assays. Expression of the housekeeping *gyrA* gene is used for normalization. Data are mean of 3 replicates -/+ SD (Data courtesy of Jiunn C.N. Fong, Nicholas J. Shikuma, and Fitnat H. Yildiz).

.

2.A3. Supplemental tables

Table S1: Data collection and refinement statistics

Data collection	**VpsT•c-di-GMP**	**VpsT (nucleotide free)**		
X-ray source	NSLS, X29	NSLS, X29	NSLS, X29	NSLS, X29
Wavelength (Å)	1.0809	0.9788 (peak)	0.9794 (inflection)	1.0809 (remote)
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$	$P4_12_12$
Unit cell parameters				
a, b, c (Å)	121.7, 121.7, 208.2	121.4, 121.4, 198.4	121.4, 121.4, 198.9	121.9, 121.9, 199.2
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution range (Å)	50.0-2.8 (2.97-2.8)[a]	50.0-3.1 (3.27-3.1)	50.0-3.2 (3.35-3.2)	50.0-3.4 (3.65-3.4)
No. of reflections				
Total	555982 (88936)	521519 (76688)	475332 (61878)	329892 (39399)
Unique	38433 (6048)	52138 (8187)	47985 (7137)	32884 (4451)
Completeness (%)	97.8 (97.3)	99.2 (96.7)	98.2 (90.8)	86.4 (72.7)
Redundancy	14.5 (14.7)	10.0 (9.4)	9.9 (8.7)	10.0 (8.8)
$I/\sigma(I)$	18.2 (4.0)	11.8 (3.2)	12.5 (3.7)	14.8 (5.3)
R_{meas} (%)	11.7 (78.8)	13.1 (84.3)	11.8 (66.7)	11.5 (48.1)
Refinement				
R_{work}/R_{free} (%)	24.6/29.0	24.8/30.8		
r.m.s. deviations				
Bond length (Å)	0.007	0.011		
Bond angles (°)	1.049	1.271		
No. of atoms				
Protein	6649	6640		
Water	49	---		
c-di-GMP/Tartrate	184/20	---		
Ave. B-factors (Å^2)				
Protein	54.6	74.6		
Water	41.2	---		
c-di-GMP/Tartrate	38.5/56.8	---		
Ramachandran (%)[b]				
Favored	90.4	89.7		
Allowed	9.0	9.9		
Generous	0.7	0.4		
Disallowed	0.0	0.0		

(a) Values for the highest resolution bin. (b) Calculated with PROCHECK.

Table S2: Oligomeric state and c-di-GMP binding affinity of VpsT

Protein	Property	- c-di-GMP	+ c-di-GMP	
			Preincubation	/ In SEC buffer
$VpsT^{wt}$				
	Molecular weight (MALS)	42.3 kDa	52.6 kDa	/ n.a.
	Polydispersity (MALS)	1.006 ± 0.03	1.004 ± 0.03	/ n.a.
	c-di-GMP binding affinity / stoichiometry (ITC)		3.18 μM / 0.94	
$M^{17}D$				
	Molecular weight (MALS)	25.9 kDa	35.1 kDa	/ 44.0 kDa
	Polydispersity (MALS)	1.007 ± 0.08	1.008 ± 0.07	/ 1.003 ± 0.04
	c-di-GMP binding affinity / stoichiometry (ITC)		2.81 μM / 1.06	
$R^{134}A$				
	Molecular weight (MALS)	41.5 kDa	41.8 kDa	/ 40.5 kDa
	Polydispersity (MALS)	1.008 ± 0.05	1.005 ± 0.05	/ 1.001 ± 0.04
	c-di-GMP binding affinity (ITC)		n.d.	
$I^{141}E$				
	Molecular weight (MALS)	40.6 kDa	40.8 kDa	/ 40.4 kDa
	Polydispersity (MALS)	1.004 ± 0.05	1.003 ± 0.05	/ 1.003 ± 0.04
	c-di-GMP binding affinity (ITC)		n.d.	
BSA				
	Molecular weight (MALS)	62.1 kDa	n.a.	/ 64.1 kDa
	Polydispersity (MALS)	1.001 ± 0.02	n.a.	/ 1.001 ± 0.03

Table S3: Gene expression profiles. Differentially expressed genes (≥2-fold) in rugose *ΔvpsT* harboring *vpsT*-WT, *vpsT*-$M^{17}D$, *vpsT*-$D^{60}A$, *vpsT*-$R^{134}A$, or *vpsT*-$I^{141}E$ in comparison to the same strain containing pBAD alone. Differentially expressed genes were determined using SAM software with criteria of a ≥2-fold change in gene expression and a false discovery rate of ≤3% (Data courtesy of Jiunn C.N. Fong, Nicholas J. Shikuma, and Fitnat H. Yildiz).

Open Reading Frame and Functional Category	Gene	Function	*vpsT*-WT	*vpsT*-$M^{17}D$	*vpsT*-$D^{60}A$	*vpsT*-$R^{134}A$	*vpsT*-$I^{141}E$
Biofilm Related Functions							
VC0916	*vpsU*	phosphotyrosine protein phosphatase	2.57	4.70	3.40		
VC0917	*vpsA*	UDP-N-acetylglucosamine 2-epimerase	3.02	6.21	3.80		
VC0918	*vpsB*	UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase	3.29	8.27	5.44	2.18	
VC0919	*vpsC*	serine acetyltransferase-related protein	3.28	6.50	4.70		
VC0920	*vpsD*	exopolysaccharide biosynthesis protein	2.19	3.83	2.36		
VC0921	*vpsE*	polysaccharide export protein		2.36	2.13		
VC0922	*vpsF*	hypothetical protein	2.66	4.92	3.52		
VC0924	*vpsH*	capK protein putative	2.71	4.91	3.53		
VC0925	*vpsI*	polysaccharide biosynthesis protein	2.55	5.32	3.75		
VC0926	*vpsJ*	hypothetical protein	4.03	5.89	4.88		
VC0927	*vpsK*	UDP-N-acetyl-D-mannosamine transferase	2.41	3.70	3.09		
VC0928	*rbmA*	hypothetical protein	7.36	8.80	5.24		
VC0930	*rbmB*	hemolysin-related protein		2.20			
VC0931	*rbmC*	conserved hypothetical protein	9.50	4.82	2.93		
VC0932	*rbmD*	hypothetical protein	12.30	10.45	5.91		
VC0933	*rbmE*	hypothetical protein	3.83	3.87	3.05		
VC0934	*vpsL*	capsular polysaccharide biosynthesis glycosyltransferase	2.46	2.50			
VC0935	*vpsM*	hypothetical protein	19.14	19.72	11.95		
VC0936	*vpsN*	polysaccharide export-related protein	18.79	16.67	9.32		
VC0937	*vpsO*	exopolysaccharide biosynthesis protein	5.83	5.22	3.20		
VC0938	*vpsP*	hypothetical protein	2.91	2.01			
VC1029	*cdgB*	cyclic-di-guanylate protein	2.15				
VC1888	*bap1*	hemolysin-related protein	2.14	3.84	2.69		
VC2455	*vpvB*	*Vibrio* phase variation protein		7.09	3.64		
VC2456	*vpvA*	*Vibrio* phase variation protein		4.12	2.52		
VCA0074	*cdgA*	GGDEF family protein		2.09			
VCA0075	VCA0075	hypothetical protein		2.17			
Chemotaxis and Motility							
VC1298	VC1298	methyl-accepting chemotaxis protein	0.40				
VC1413	VC1413	methyl-accepting chemotaxis protein	0.32	0.17			
VC1763	VC1763	chemotaxis protein MotB-related protein		0.50			
VC1898	VC1898	methyl-accepting chemotaxis protein		0.35			
VC2065	*cheY-3*	chemotaxis protein		0.48			
VC2068	*flhF*	flagellar biosynthetic protein		0.34			
VC2069	*flhA*	flagellar biosynthetic protein		0.31			
VC2131	*fliH*	flagellar assembly protein		0.48			
VC2132	*fliG*	flagellar motor switch protein		0.47			
VC2133	*fliF*	flagellar M-ring protein		0.41			
VC2134	*fliE*	flagellar hook-basal body complex protein		0.26	0.41		

VC2142	*flaB*	flagellin		0.40	0.45		
VC2143	*flaD*	flagellin	0.37	0.20	0.25		
VC2187	*flaC*	flagellin		0.26			
VC2188	*flaA*	flagellin core protein	0.42	0.27	0.34		
VC2189	VC2189	hypothetical protein	0.18	0.15	0.31		
VC2191	*flgM*	flagellar hook-associated protein		0.28			
VC2194	*flgH*	flagellar L-ring protein		0.41			
VC2195	*flgG*	flagellar basal-body rod protein	0.47	0.25	0.43		
VC2196	*flgF*	flagellar basal-body rod protein		0.37			
VC2197	*flgE*	flagellar hook protein		0.28			
VC2198	*flgD*	basal-body rod modification protein		0.45			
VC2199	*flgC*	flagellar basal-body rod protein		0.36			
VC2200	*flgB*	flagellar basal-body rod protein		0.21	0.32		
VCA0864	VCA0864	methyl-accepting chemotaxis protein		2.29			
Hypothetical							
VC0102	VC0102	hypothetical protein	0.48				
VC0132	VC0132	hypothetical protein			2.89		
VC0138	VC0138	hypothetical protein		2.46			
VC0184	VC0184	hypothetical protein	2.08				
VC0283	VC0283	hypothetical protein		2.11			
VC0823	VC0823	hypothetical protein					0.50
VC1420	VC1420	hypothetical protein		0.47			
VC1538	VC1538	hypothetical protein	0.34				
VC1933	VC1933	hypothetical protein	0.48				
VC1997	VC1997	hypothetical protein			2.05		
VC2046	VC2046	hypothetical protein		0.44			
VC2058	VC2058	hypothetical protein		0.49			
VC2207	VC2207	hypothetical protein	0.46	0.21	0.45		
VC2304	VC2304	hypothetical protein					0.43
VC2331	VC2331	hypothetical protein	2.06	2.66			
VC2372	VC2372	hypothetical protein			6.07		
VC2667	VC2667	hypothetical protein		2.40			
VCA0215	VCA0215	hypothetical protein			2.72		
VCA0284	VCA0284	hypothetical protein		0.46			0.47
VCA0556	VCA0556	hypothetical protein		0.46			
VCA0672	VCA0672	hypothetical protein		0.48			
VCA0874	VCA0874	hypothetical protein		0.43			
VCA0934	VCA0934	hypothetical protein		0.43			
VCA0935	VCA0935	hypothetical protein	0.33	0.41			
VCA1075	VCA1075	hypothetical protein	0.27				
VCA1076	VCA1076	hypothetical protein	0.21				
VC1151	VC1151	conserved hypothetical protein	0.50				
VC1710	VC1710	conserved hypothetical protein	0.43				
VC2206	VC2206	conserved hypothetical protein		0.35			
VC2386	VC2386	conserved hypothetical protein		0.49			
VCA0055	VCA0055	conserved hypothetical protein	2.73	5.90	4.12		
VCA0167	VCA0167	conserved hypothetical protein		2.54	2.42		
VCA0536	VCA0536	conserved hypothetical protein	0.31				

Regulatory Functions							
VC0665	*vpsR*	sigma-54 dependent transcriptional regulator		2.18			
VC1118	VC1118	transcriptional regulator putative		0.40			
VC1349	VC1349	sensory box sensor histidine kinase/response regulator		2.03			
VCA0166	*cspA*	cold shock transcriptional regulator					0.46
VCA0933	VCA0933	cold shock domain family protein	0.41	0.39		0.39	0.37
Others							
VC0194	*ggt*	gamma-glutamyltranspeptidase		2.27			
VC0284	VC0284	TonB system receptor	2.57	4.51	3.11		
VC0436	*rpmA*	ribosomal protein L27					0.45
VC0475	*irgA*	enterobactin receptor			0.47		
VC0586	VC0586	carbonic anhydrase putative	0.50				
VC0679	*rpsT*	ribosomal protein S20	0.41				
VC1301	*sdaC*-1	serine transporter		2.23	2.04		
VC1584	*ankB*	ankB protein	2.25				
VC1585	*katB*	catalase	2.39	2.18			
VC1620	VC1620	hypothetical protein	0.33	0.36	0.42		
VC1621	VC1621	agglutination protein	0.41				
VC1622	VC1622	outer membrane protein putative	0.46				
VC1672	*tag*	DNA-3-methyladenine glycosidase I		4.55			
VC1862	VC1862	amino acid ABC transporter permease protein		0.48			
VC1935	VC1935	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase-related protein		2.06			
VC1936	VC1936	phosphatidate cytidylyltransferase putative		2.30			
VC2305	*ompK*	outer membrane protein	0.49				
VC2336	VC2336	methionyl-tRNA synthetase-related protein	2.42	2.43			
VC2416	VC2416	2, 3-cyclic-nucleotide 2-phosphodiesterase		2.31			
VC2582	*rpsH*	ribosomal protein S8	2.00				
VCA0057	*phrB*-2	deoxyribodipyrimidine photolyase	2.09	3.94	2.54		
VCA0064	*hutR*	heme receptor					2.02
VCA0088	*gltP*-2	proton/glutamate symporter	2.04				
VCA0518	*fruB*	PTS system fructose-specific			0.50		
VCA0860	*malS*	alpha-amylase			0.46		
VCA1028	*ompS*	maltoporin			0.34		

Table S4: Bacterial strains and plasmids

Strain or plasmid	Relevant properties	Source
***E. coli* strain**		
DH10B	F⁻ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80*lacZ*ΔM15 Δ*lacX74 recA1 endA1 ara*Δ*139* Δ(*ara leu*)*7697 galU galK* λ⁻ *rpsL* (Str^r) *nupG*	Invitrogen
S17-1 (λ *pir*)	Tp^r Sm^r *recA, thi, pro,* r_K^- m_K^+ *RP4:2-Tc*:MuKm Tn7 λ *pir*	(*39*)
DH5α	*fhuA2*Δ*(argF-lacZ)U169 phoA glnV44* ϕ80 *lacZ*ΔM15 *gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB
T7 Express	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^S*)2 [dcm] R(zgb-210::Tn10--*Tet^S*) endA1* Δ*(mcrC-mrr)114::IS10*	NEB
T7 Express Crystal	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^S*)2 [dcm] R(zgb-210::Tn10--*Tet^S*) endA1 metB1* Δ*(mcrC-mrr)114::IS10*	NEB
***V. cholerae* strains**		
FY_Vc_1	*Vibrio cholerae* O1 El Tor A1552, smooth variant, Rif^r	(*40*)
FY_Vc_2	*Vibrio cholerae* O1 El Tor A1552, rugose variant, Rif^r	(*40*)
FY_Vc_616	FY_Vc_1 carrying chromosomal *vpsLp-lacZ*, Rif^r	(*3*)
FY_Vc_618	FY_Vc_2 carrying chromosomal *vpsLp-lacZ*, Rif^r	This study
FY_Vc_5	FY_Vc_2 Δ*vpsT*	(*4*)
FY_Vc_4435	FY_Vc_618 Δ*vpsT*	This study
FY_Vc_3463	FY_Vc_616 Δ*vpsT*	(*3*)
FY_Vc_5045	FY_Vc_618 carrying chromosomal *vpsT*(M17D)	This study
FY_Vc_5046	FY_Vc_618 carrying chromosomal *vpsT*(D60A)	This study
FY_Vc_5047	FY_Vc_618 carrying chromosomal *vpsT*(D60E)	This study
FY_Vc_5048	FY_Vc_618 carrying chromosomal *vpsT*(R134A)	This study
FY_Vc_5049	FY_Vc_618 carrying chromosomal *vpsT*(I141E)	This study
Plasmids		
pET28a/His-*sumo*	modified pET28a (Novagen) yielding N-terminally hexahistidine-tagged SUMO fusion proteins, Kan^r	This study
pETsumo_wt	pET28a/His-*sumo*::*vpsT*, Kan^r	This study
pETsumo_$M^{17}D$	pET28a/His-*sumo*::*vpsT*(M17D), Kan^r	This study
pETsumo_$R^{134}A$	pET28a/His-*sumo*::*vpsT*(R134A), Kan^r	This study
pETsumo_$I^{141}E$	pET28a/His-*sumo*::*vpsT*(I141E), Kan^r	This study
pETsumo_$W^{131}F$	pET28a/His-*sumo*::*vpsT*(W131F), Kan^r	This study
pETsumo_$W^{131}A$	pET28a/His-*sumo*::*vpsT*(W131A), Kan^r	This study
pETsumo_$T^{133}A$	pET28a/His-*sumo*::*vpsT*(T133A), Kan^r	This study
pETsumo_$T^{133}V$	pET28a/His-*sumo*::*vpsT*(I133V), Kan^r	This study
pBAD/*myc*-His-B	Arabinose-inducible expression vector with C-terminal myc epitope and six-His tags, Ap^r	Invitrogen
pFY-876	pBAD/*myc*-His-B::*vpsT*, Ap^r	This study
pFY-877	pBAD/*myc*-His-B::*vpsT*(M17D), Ap^r	This study
pFY-842	pBAD/*myc*-His-B::*vpsT*(W131F), Ap^r	This study
pFY-881	pBAD/*myc*-His-B::*vpsT*(T133A), Ap^r	This study
pFY-843	pBAD/*myc*-His-B::*vpsT*(T133V), Ap^r	This study
pFY-844	pBAD/*myc*-His-B::*vpsT*(R134A), Ap^r	This study
pFY-846	pBAD/*myc*-His-B::*vpsT*(I141E), Ap^r	This study
pFY-884	pBAD/*myc*-His-B::*vpsT*(M17D/R134A), Ap^r	This study
pFY-903	pBAD/*myc*-His-B::*vpsT*(D60A), Ap^r	This study
pFY-904	pBAD/*myc*-His-B::*vpsT*(D60E), Ap^r	This study
pFY-217	pGP704-*sacB*28::*vpsLp-lacZ* transcriptional fusion; Ap^r	(*10*)
pBAD33	pACYC184 ori, araC P_{araBAD}, Cm^r	(*41*)
pAT1662	pBAD33::VCA0956-His_6, Cm^r	(*37*)
pFY-959	pGP704-*sacB*28::*vpsT*(M17D), Ap^r	This study
pFY-960	pGP704-*sacB*28::*vpsT*(D60A), Ap^r	This study
pFY-961	pGP704-*sacB*28::*vpsT*(D60E), Ap^r	This study
pFY-962	pGP704-*sacB*28::*vpsT*(R134A), Ap^r	This study
pFY-963	pGP704-*sacB*28::*vpsT*(I141E) Ap^r	This study

2.A4. Material and Methods

Bacterial strains, plasmids and culture conditions

The bacterial strains and plasmids used in this study are listed in Table 2.S4. All *V. cholerae* and *E. coli* strains were grown aerobically, at 30°C and 37°C, respectively. Unless otherwise noted, growth medium consisted of Luria Bertani (LB) medium (1% Tryptone, 0.5% Yeast Extract, 0.5% NaCl, pH 7.5). LB-agar and LB-soft agar contained 1.5% (wt/vol) and 0.3% (wt/vol) granulated agar (Difco), respectively. Concentrations of antibiotics used, where appropriate, were as follows: ampicillin (100 μg/ml), rifampicin (100 μg/ml), kanamycin (50 μg/ml), and chloramphenicol (5 μg/ml).

Recombinant DNA techniques

DNA manipulations were carried out using standard molecular techniques. VpsT point mutants were generated using QuikChange site-directed mutagenesis kit (Stratagene), following the manufacturer's instructions. In-frame deletions and *vpsLp-lacZ* single-copy chromosomal reporter strains were generated as described previously (*S1-3*). All point mutations and chromosomal deletions/insertions were sequence-verified.

Protein expression and purification

The coding region corresponding to full-length VpsT from *V. cholerae* O1 El Tor (VCA0952) (*S4*) was amplified by standard PCR and cloned into a modified pET28a expression plasmid (Novagen) yielding N-terminally hexahistidine-tagged SUMO fusion proteins. The hexahistidine-tagged SUMO-moiety was cleavable using the protease Ulp-1 from *S. cerevisiae*.

Native and selenomethionine-derivatized proteins were overexpressed in *E.*

coli T7 Express or T7 Crystal Express cells, respectively (NEB). For the expression of native proteins, cells were grown in Terrific Broth (TB) media supplemented with 50 μg/ml kanamycin at 37°C. At a cell optical density corresponding to an absorbance of 0.8-1.0 at 600 nm (OD_{600}), the temperature was reduced to 18°C, and protein production was induced with 1 mM IPTG. Selenomethionine-derivatized proteins were produced in cells grown in M9 minimal media supplemented with 50 μg/ml kanamycin, 1 μg/ml thiamine, 1 μg/ml biotin, 0.4% glucose and 40 μg/ml of each of the 20 amino acids with selenomethionine substituting for methionine. Protein expression was induced at OD_{600} of 0.4-0.5. After 16 hours, cells were harvested by centrifugation, resuspended in NiNTA buffer A (25 mM Tris-Cl, pH 7.5, 550 mM NaCl and 20 mM imidazole), and flash-frozen in liquid nitrogen.

After cell lysis by sonication and removal of cell debris by centrifugation, clear lysates were loaded onto NiNTA columns (HisTrap; GE Healthcare) equilibrated in NiNTA buffer A. The resin was washed with 20 column volumes of NiNTA buffer A, and proteins were eluted in a single step of NiNTA buffer A supplemented with 500 mM imidazole. Proteins were buffer exchanged into desalting buffer (25 mM Tris-HCl, pH 7.5, 550 mM NaCl, 10mM imidazole, 5mM β-mercaptoethanol), and affinity tags were removed by incubation with the yeast protease Ulp-1 at 4°C overnight. Cleaved proteins were collected in the flow-through during NiNTA affinity chromatography, and were subjected to size exclusion chromatography on a Superdex 200 column (GE Healthcare) equilibrated in gel filtration buffer (25 mM Tris-HCl, pH 7.5, 550 mM NaCl, 2mM DTT). Proteins were concentrated on a Centricon ultrafiltration device (10 kDa cutoff; Millipore) to a final concentration of approximately 1-4 mM. Protein aliquots were flash frozen in liquid nitrogen and stored at -80°C.

Crystallization, data collection and structure determination

Crystals were obtained by hanging drop vapor diffusion by mixing equal volumes of protein (10-40 mg/ml) and reservoir solution followed by incubation at 20°C. For crystallization of the c-di-GMP bound state, protein incubated with c-di-GMP was subjected to size exclusion chromatography for removal of unbound nucleotide prior to crystallization. The reservoir solution contained 0.1 M Tris-HCl pH 7.0, 0.8 M Potassium sodium tartrate tetrahydrate, 3-5% Polyethylene glycol monomethyl ether 5,000, and 8-12% xylitol. Crystals appeared within 3-10 days with typical dimensions of 0.40 mm x 0.08 mm x 0.08 mm. For cryo-protection, crystals were soaked in reservoir solution supplemented with 25% xylitol. Cryo-preserved crystals were flash-frozen and stored in liquid nitrogen. Data was collected on frozen crystals at 100 K using synchrotron radiation at the National Synchrotron Light Source (NSLS, Brookhaven National Laboratory, beamline X29).

Data reduction was carried out with the software package XDS (*S5*). Experimental phases were obtained from multi-wavelength anomalous diffraction (MAD) experiments on crystals grown from selenomethionine-derivatized proteins. By using the software package HKL2MAP/Shelx (*S6*), 28 out of 32 heavy atom positions could be determined. Solvent flattening was carried out by using the program ShelxE (*S6*). The structure of c-di-GMP-bound VpsT was determined by molecular replacement using the software package PHENIX (*S7*) with the apo-structure as the search model. Refinement in PHENIX (*S7*) and COOT (*S8*) yielded the final models. Data collection and refinement statistics are summarized in Table 2.S1. Illustrations were made in Pymol (DeLano Scientific).

Large-Scale Enzymatic production of c-di-GMP

Large amounts of c-di-GMP were synthesized enzymatically using a highly

active mutant of the diguanylate cyclase WspR and GTP as a substrate (*S9*). High purity of the compound was achieved by preparative reverse-phase HPLC followed by lyophilization. In a final step of the product analysis, it was enzymatically tested as a substrate for phosphodiesterases. Cyclic di-GMP concentration was determined based on absorbance at 254 nm in comparison with commercially obtained standard of known concentration (Biolog Life Science Institute).

Size-exclusion chromatography (SEC)-coupled static multi-angle light scattering

For SEC-coupled multi-angle light scattering, purified protein (~10 μg/μl or 400 μM, injected concentration) was subjected to SEC using a Shodex KW-803 column (JM Science, Inc.) equilibrated overnight in gel filtration buffer (25 mM Tris-HCl pH 7.5, 400-600 mM NaCl, and 2 mM DTT). Where specified, wild-type or mutant VpsT was incubated with excess c-di-GMP for 30 minutes at room temperature prior to injection. In additional sets of experiments, c-di-GMP was also added to the gel filtration buffer at a concentration of 40 μM. The chromatography system was coupled to a 3-angle light scattering detector (miniDAWN TREOS) and a refractive index detector (Optilab rEX) (Wyatt Technology). Data were collected every second at a flow rate of 0.5 ml/min. Data analysis was carried out using the program ASTRA, yielding the molar mass and mass distribution (polydispersity) of the sample. For normalization of the light scattering detectors and data quality control, monomeric bovine serum albumin (BSA; Sigma) was used.

Reverse-phase HPLC

SEC eluted protein peaks from above were collected, concentrated to a final concentration of 10 μg/μl, heat denatured at 95°, and centrifuged at 14,000 rpm for 10 minutes. Resulting supernatants were filtered through Microcon Centrifugal Filter

Units (Millipore, 10 kDa cut-off) and separated on a C18 reverse-phase column using a methanol-phosphate gradient (buffer A: 100 mM potassium phosphate monobasic, pH 6.0; buffer B: 30% methanol/70% buffer A). Protein bound c-di-GMP was identified by comparison to a nucleotide standard.

Analytical ultracentrifugation

Sedimentation velocity experiments were carried out using an XL-I analytical ultracentrifuge (Beckman Coulter) equipped with an AN-60 Ti rotor. Proteins (8 and 12 μM) were diluted in buffer (25 mM Tris-HCl pH 7.5, 550 mM NaCl, 2 mM DTT) in the absence or presence of c-di-GMP (25 μM), and were analyzed at a centrifugation speed of 130,000 x g. Data collection was carried out at 280 nm, followed by data analysis using the program SedFit (version 11.0).

Isothermal titration calorimetry

Apparent dissociation constants (K_d) and stoichiometry of interactions were measured by isothermal titration calorimetry (ITC) using a VP calorimeter (Microcal, Amherst, MA). Calorimetric titrations of c-di-GMP (250 μM in the syringe; 10 μl injections) and wild-type or mutant VpsT (25 μM in the cuvette) were carried out at 20°C in assay buffer (25 mM Tris-HCl pH 7.5, 550 mM NaCl, 2mM DTT) with a spacing of 180 or 300 sec between injections. ITC data were analyzed by integrating heat effects normalized to the amount of injected protein and curve-fitting based on a single-site binding model using the Origin software package (Microcal). The dissociation constant was derived from the data by using standard procedures.

RNA isolation

V. cholerae cells were grown aerobically overnight in LB in the absence of

arabinose. Cultures were diluted 1:500 in fresh media with the inducer arabinose (0.1%) and grown aerobically at 30°C with shaking at 200 rpm to an OD_{600} of 0.3 to 0.4. Two ml aliquots of cultures were collected by centrifugation for 2 min at room temperature. Cell pellets were immediately resuspended in 1 ml Trizol reagent (Invitrogen) and stored at -80°C. Total RNA was isolated from the cell pellets according to the manufacturer's instructions (Invitrogen). Contaminating DNA was removed by incubating RNA with RNase-free DNase I (Ambion), and an RNeasy Mini kit (Qiagen) was used to clean up RNA after DNase digestion.

Quantitative PCR (qPCR)

qPCR was performed as described previously (*S3*). Briefly, cDNA was synthesized from 1 μg of RNA from each sample using an iScript cDNA Synthesis Kit (Bio-Rad). The product was used as a template in a subsequent PCR reaction using Expand High Fidelity PCR System (Roche). PCR reaction conditions were as follows: 94°C for 2 min, then 25 cycles of 94°C for 30 sec, 55°C for 30 sec and 72°C for 30 sec, and a final 72°C for 2 min. Amplified products were analyzed on a 1.5% agarose gel and quantified using ImageQuant software (Molecular Dynamics). Intensities of each DNA band were normalized to the corresponding gyrA band. Three biological replicates were conducted for each treatment tested and reactions lacking reverse transcriptase were used as negative controls.

Electromobility shift assays

Electromobility shift assays (EMSAs) were performed using purified wild-type or mutant VpsT proteins and DNA fragments tiling the *vpsL* promoter region. Briefly, biotinylated primers were used to amplify the corresponding chromosomal regions by standard PCR using genomic DNA as a template. Commercially available

60 bp duplex biotin end-labeled Epstein-Barr Nuclear Antigen (EBNA) DNA (Pierce) was used as a negative control. Binding reactions contained final concentrations of 5 nM labeled DNA, 1 μM protein, and 0.05 μg/μl Poly-dI-dC in the binding buffer (10 mM Tris-HCl pH 7.5, 50 mM KCl, 75 mM NaCl, 10 mM $MgCl_2$, 5% xylitol, 1 mM DTT, 0.1 mM EDTA). Where specified, c-di-GMP was added to a final concentration of 50 μM. After 40 min incubation at room temperature, DNA was resolved in 5% TBE-polyacrylamide gels (BioRad) with 0.5X TBE as running buffer (45 mM Tris Base, 45 mM Boric Acid, 1 mM EDTA). DNA mobility was visualized using LightShift Chemiluminescent EMSA kit (Pierce) following the manufacturer's instructions.

β-Galactosidase assays

V. cholerae cells were grown aerobically overnight in LB in the absence of arabinose. Cultures were diluted 1:500 in fresh media with the inducer arabinose (0.1%) and grown aerobically at 30°C with shaking at 200 rpm to an OD_{600} of 0.3 to 0.4. β-galactosidase assays were carried out in MultiScreen 96-well microtiter plates fitted onto a MultiScreen filtration system (Millipore) using a previously published procedure (*S10*). The assays were repeated with two biological replicates and at least six technical replicates.

Gene expression profiling

Microarrays used in this study were performed as described previously (*S3*), except reference RNA was obtained from a Δ*vpsT V. cholerae* strain (Fy_Vc_4435) harboring pBAD/*myc*-His-B grown to mid-exponential phase (OD_{600} of 0.3 to 0.4) in LB with the inducer arabinose (0.1%), inoculated (1:500 dilution) with overnight grown culture. Differentially regulated genes were determined using three biological

replicates and two technical replicates for each treatment (6 data points for each spot) with the SAM software (*S11*) using 2-fold differences in gene expression and 3% false discovery rate (FDR) as cut-off values. Microarray data has been deposited in the NCBI Gene Expression Omnibus (GEO) database.

Spot morphology and motility assays

Analysis of spot morphologies of strains carrying plasmids with either wild-type VpsT or VpsT point mutants ($VpsT^{M17D}$, $VpsT^{R134A}$, or $VpsT^{I141E}$) were carried out by spotting 2 μl of 1:200-diluted overnight cultures onto LB agar plates supplemented with 100 μg/ml ampicillin and 0.02% arabinose. Spot cultures were incubated for 1 day at 30°C and photographed. Spot morphologies shown are representation of two independent biological replicates. Motility assays were carried out with LB-soft agar (0.3% agar) inoculated from a single colony grown overnight on LB agar at 30°C. After incubation for 18 to 20 h at 30°C, the migration zone diameter was measured and compared between strains.

SUPPLEMENTAL REFERENCES

S1. Fong, J. C., K. Karplus, et al. (2006). "Identification and characterization of RbmA, a novel protein required for the development of rugose colony morphology and biofilm structure in Vibrio cholerae." *J Bacteriol* **188**(3): 1049-59.

S2. Lim, B., S. Beyhan, et al. (2006). "Cyclic-diGMP signal transduction systems in Vibrio cholerae: modulation of rugosity and biofilm formation." *Mol Microbiol* **60**(2): 331-48.

S3. Shikuma, N. J., J. C. Fong, et al. (2009). "Overexpression of VpsS, a hybrid sensor kinase, enhances biofilm formation in Vibrio cholerae." *J Bacteriol* **191**(16): 5147-58.

S4. Casper-Lindley, C. and F. H. Yildiz (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." *J Bacteriol* **186**(5): 1574-8.

S5. Kabsch, W. (1993). "Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants." *J. Appl. Cryst.* **26**: 795-800.

S6. Sheldrick, G. M. (2008). "A short history of SHELX." *Acta Crystallogr A* **64**(Pt 1): 112-22.

S7. Adams, P. D., R. W. Grosse-Kunstleve, et al. (2002). "PHENIX: building new software for automated crystallographic structure determination." *Acta Crystallogr D Biol Crystallogr* **58**(Pt 11): 1948-54.

S8. Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-32.

S9. De, N., M. Pirruccello, P.V. Krasteva et al. (2008). "Phosphorylation-independent regulation of the diguanylate cyclase WspR." *PLoS Biol* **6**(3): e67.

S10. Shikuma, N. J., J. C. Fong, et al. (2009). "Overexpression of VpsS, a hybrid sensor kinase, enhances biofilm formation in Vibrio cholerae." *J Bacteriol* **191**(16): 5147-58.

S11. Tusher, V. G., R. Tibshirani, et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." *Proc Natl Acad Sci U S A* **98**(9): 5116-21.

S12. Chan, C., R. Paul, et al. (2004). "Structural basis of activity and allosteric control of diguanylate cyclase." *Proc Natl Acad Sci U S A* **101**(49): 17084-9.

S13. Egli, M., R. V. Gessner, et al. (1990). "Atomic-resolution structure of the cellulose synthase regulator cyclic diguanylic acid." *Proc Natl Acad Sci U S A* **87**(8): 3235-9.

S14. Liaw, Y. C., Y. G. Gao, et al. (1990). "Cyclic diguanylic acid behaves as a host molecule for planar intercalators." *FEBS Lett* **264**(2): 223-7.

S15. Gao, R. and A. M. Stock (2009). "Biological insights from structures of two-component proteins." *Annu Rev Microbiol* **63**: 133-54.

S16. Maris, A. E., M. Kaczor-Grzeskowiak, et al. (2005). "Primary and secondary modes of DNA recognition by the NarL two-component response regulator." *Biochemistry* **44**(44): 14538-52.

S17. Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." *J Mol Biol* **372**(3): 774-97.

S18. Brombacher, E., C. Dorel, et al. (2003). "The curli biosynthesis regulator CsgD co-ordinates the expression of both positive and negative determinants for biofilm formation in Escherichia coli." *Microbiology* **149**(Pt 10): 2847-57.

S19. Tyson, K. L., A. I. Bell, et al. (1993). "Definition of nitrite and nitrate response elements at the anaerobically inducible Escherichia coli nirB promoter: interactions between FNR and NarL." *Mol Microbiol* **7**(1): 151-7.

S20. Baikalov, I., I. Schroder, et al. (1996). "Structure of the Escherichia coli response regulator NarL." *Biochemistry* **35**(34): 11053-61.

S21. Schultz, S. C., G. C. Shields, et al. (1991). "Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees." *Science* **253**(5023): 1001-7.

S22. Matthews, K. S. (1992). "DNA looping." *Microbiol Rev* **56**(1): 123-36.

S23. Schleif, R. (1992). "DNA looping." *Annu Rev Biochem* **61**: 199-223.

S24. Lobell, R. B. and R. F. Schleif (1991). "AraC-DNA looping: orientation and distance-dependent loop breaking by the cyclic AMP receptor protein." *J Mol Biol* **218**(1): 45-54.

S25. Zhang, Z., S. Kim, et al. (2006). "Polymorphism of the signaling molecule c-di-GMP." *J Am Chem Soc* **128**(21): 7015-24.

S26. Zoltowski, B. D. and B. R. Crane (2008). "Light activation of the LOV protein vivid generates a rapidly exchanging dimer." *Biochemistry* **47**(27): 7012-9.

S27. Guvener, Z. T. and L. L. McCarter (2003). "Multiple regulators control capsular polysaccharide production in Vibrio parahaemolyticus." *J Bacteriol* **185**(18): 5431-41.

S28. Ferreira, R. B., L. C. Antunes, et al. (2008). "Vibrio parahaemolyticus ScrC modulates cyclic dimeric GMP regulation of gene expression relevant to growth on surfaces." *J Bacteriol* **190**(3): 851-60.

S29. Kim, Y. K. and L. L. McCarter (2007). "ScrG, a GGDEF-EAL protein, participates in regulating swarming and sticking in Vibrio parahaemolyticus." *J Bacteriol* **189**(11): 4094-107.

S30. Hickman, J. W. and C. S. Harwood (2008). "Identification of FleQ from

Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor." *Mol Microbiol* **69**(2): 376-89.

S31. Romling, U. (2005). "Characterization of the rdar morphotype, a multicellular behaviour in Enterobacteriaceae." *Cell Mol Life Sci* **62**(11): 1234-46.

S32. Romling, U., M. Rohde, et al. (2000). "AgfD, the checkpoint of multicellular and aggregative behaviour in Salmonella typhimurium regulates at least two independent pathways." *Mol Microbiol* **36**(1): 10-23.

S33. Romling, U., W. D. Sierralta, et al. (1998). "Multicellular and aggregative behaviour of Salmonella typhimurium strains is controlled by mutations in the agfD promoter." *Mol Microbiol* **28**(2): 249-64.

S34. Chirwa, N. T. and M. B. Herrington (2003). "CsgD, a regulator of curli and cellulose synthesis, also regulates serine hydroxymethyltransferase synthesis in Escherichia coli K-12." *Microbiology* **149**(Pt 2): 525-35.

S35. Larkin, M. A., G. Blackshields, et al. (2007). "Clustal W and Clustal X version 2.0." *Bioinformatics* **23**(21): 2947-8.

S36. Gouet, P., E. Courcelle, et al. (1999). "ESPript: analysis of multiple sequence alignments in PostScript." *Bioinformatics* **15**(4): 305-8.

S37. Tischler, A. D. and A. Camilli (2004). "Cyclic diguanylate (c-di-GMP) regulates Vibrio cholerae biofilm formation." *Mol Microbiol* **53**(3): 857-69.

S38. Holm, L., S. Kaariainen, et al. (2008). "Searching protein structure databases with DaliLite v.3." *Bioinformatics* **24**(23): 2780-1.

S39. de Lorenzo, V. and K. N. Timmis (1994). "Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons." *Methods Enzymol* **235**: 386-405.

S40. Yildiz, F. H. and G. K. Schoolnik (1999). "Vibrio cholerae O1 El Tor: identification of a gene cluster required for the rugose colony type,

exopolysaccharide production, chlorine resistance, and biofilm formation." *Proc Natl Acad Sci U S A* **96**(7): 4028-33.

S41. Guzman, L. M., D. Belin, et al. (1995). "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter." *J Bacteriol* **177**(14): 4121-30.

ACKNOWLEDGMENTS

Work on the structural and functional characterization of VpsT would not have been possible without the outstanding contributions from members of the Yildiz Lab at UC Santa Cruz and the Sondermann Lab at Cornell University. I am deeply grateful to Fitnat Yildiz and my mentor, Dr. Holger Sondermann, for help with the experimental design, data analysis, and manuscript writing; to Jiunn C. N. Fong and Nicholas J. Shikuma for conducting the cell-based experiments; to Sinem Beyhan for providing certain strains and reagents; and to Marcos V. A. S. Navarro for providing access to synchrotron data collection and for general advice in crystallographic data analysis.

APPENDIX B

Additional targets and future directions

2.B1. Identification of VpsT consensus binding sites in promoter/enhancer DNA

The global effects of c-di-GMP mediated signaling, where distinct responses affecting motility, secretion, and biofilm formation are concertedly tuned to rapidly adapt bacteria's physiology and virulence potential (*B1*), make regulators of gene expression likely candidates for c-di-GMP responsive effectors. Such addaptational strategy was recently confirmed in the literature by the identification of protein transcription factors (FleQ of *P. aeruginosa,* Clp of *X. campestris*), as well as a riboswitch class in some messenger RNAs, to directly sense c-di-GMP and regulate gene expression at the transcription initiation and post-transcriptional levels, respectively (*B2-B5*). Here, we presented the crystal structure of *Vibrio cholerae* VpsT and identified it as a yet another transcription regulator which inversely controls biofilm formation and motility in response to fluctuations in intracellular c-di-GMP.

In rugose cells, VpsT positively regulates and shares an overall similar regulon with VpsR, a transcription factor homologous to c-di-GMP sensing FleQ of *Pseudomonas aeruginosa* (*B6-B9*). Bioinformatic analysis of the upstream non-coding regions of genes that are differentially expressed between the rugose (R) variant and an RΔ*vpsR* strain, has led to the identification of a highly significant palindrome motif, later confirmed experimentally to directly bind VpsR (*B10, B11*). Since VpsR and VpsT regulate by and large the same set of genes (*B8*), similar bioinformatic approaches have failed to distinguish between binding sites for the two proteins, assuming that they are non-overlapping. The potential for hierarchical gene

regulation, where VpsT acts indirectly by stimulating or repressing the expression of other transcriptional regulators, including VpsR, further complicates such analyses.

As discussed in the previous section, the DNA binding domain of VpsT shares high percent sequence similarity with the helix-turn-helix motifs of NarL and CsgD (Fig. 2.17) (*B9*). However, most of the conserved residues participate in protein-DNA interactions through the DNA phosphate backbone, contributing little or no sequence specificity (*B12*). Thus, while consensus DNA motifs for CsgD and NarL binding have been previously proposed, VpsT recognition sites remain to be experimentally determined (*B13-B15*).

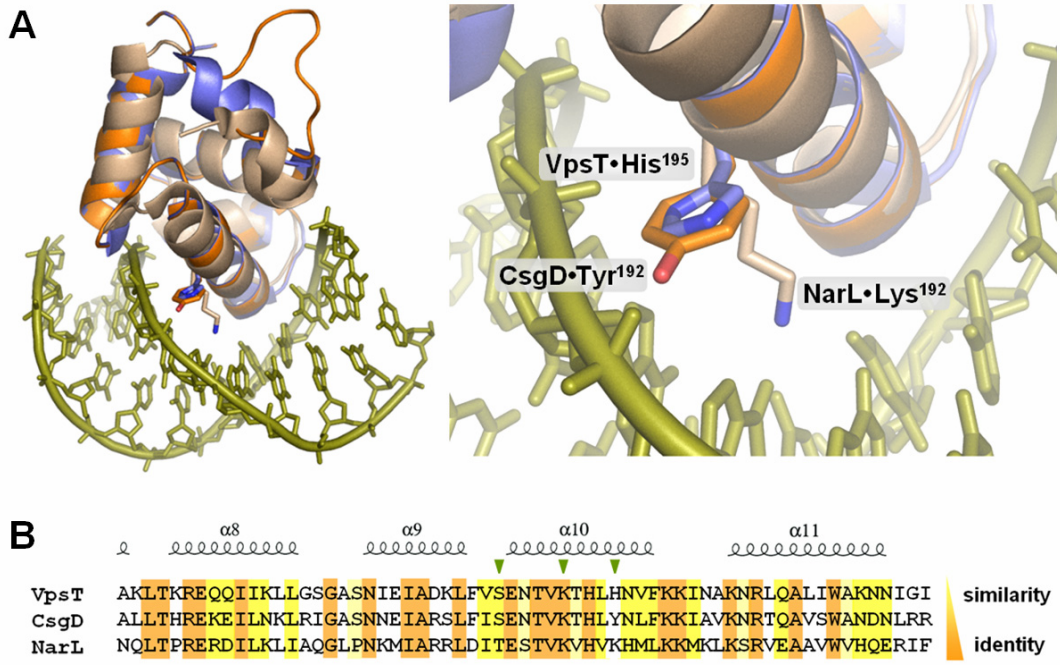


Figure 2.17: Structural basis for sequence specificity.
(A) Structural alignment of the DNA-binding domain of VpsT (slate, pdb entry: 3klo), the modeled DNA-binding domain of CsgD (orange, Swiss Model), and the DNA-binding domain of NarL bound to its cognate DNA (wheat and olive, pdb entry: 1zg5). Lys192 is one of the few NarL residues participating in direct hydrogen bonding with DNA nucleobases (*B12*, *B15*). While it fits in the large groove of DNA to stabilize the protein-DNA complex, corresponding residues in VpsT and CsgD are divergent and sterically incompatible.
(B) Alignment of the DNA-binding modules of the three proteins (*B9*). Residues participating in direct contacts with nucleobases are marked with green symbols.

In order to establish the specific DNA sequences responsible for VpsT-mediated gene regulation a number of sophisticated approaches can be employed. For the selection of high-affinity DNA motifs that would be later suitable for crystallographic studies, Systematic Evolution of Ligands by EXponential enrichment (SELEX or *in vitro* selection) is a primary method of choice (*B16, B17*). Briefly, a double stranded DNA library of oligonucleotides, harboring a central random region (30-40 base pairs) flanked by tails of fixed sequence to be used for PCR amplification and sequencing, would be subjected to VpsT affinity chromatography. To this end, VpsT expressed either as an N-terminal hexahistidine SUMO fusion protein (His_6-SUMO-VpsT) or as a C-terminally tagged VpsT-MYC-His_6 construct is shown to remain functional and can bind c-di-GMP *in vivo* (data not shown). *In vitro* bound oligonucleotides will be eluted, amplified, sequenced, and subjected to subsequent rounds of selection with increasing stringency of the binding conditions. In order to determine possible conformation-specific binding sequences, *in vitro* selection experiments with wild-type VpsT and VpsT point mutants of choice must be performed in parallel and in the presence or absence of nucleotide.

Although a powerful approach, oftentimes stringent *in vitro* selection yields sequences with much higher binding affinity and significant sequence divergence as compared to naturally occurring consensus motifs. In fact, transcriptional regulation in the cell often involves cooperative binding of multiple factors, that have otherwise weak affinities for their cognate DNA recognition sites (*B18-B21*). To overcome this, alternative or complementary to SELEX are Chromatin Immuno-Precipitation (ChIP) approaches (*B22-B24*). Briefly, *V.cholerae* strains expressing functionally tagged wild-type VpsT or VpsT point mutants would be grown under conditions favoring expression of biofilm determinants. Harvested cells would be subjected to mild formaldehyde crosslinking conditions to stabilize protein-DNA complexes. Following

cell lysis and fragmentation of DNA, VpsT-DNA complexes would be immunoprecipitated by using standard affinity purification protocols. After reversal of the crosslinks, protein-bound DNA would be purified and subjected to low-cost next-generation sequencing for identification of consensus binding motifs.

Finally, a more targeted approach for identification of VpsT-specific DNA motifs would involve DNase I footprinting approaches (*B25, B26*). In our promoter-binding studies we showed that VpsT binds in a c-di-GMP dependent manner to multiple sites in the VpsL promoter region, without the requirement of additional factors for complex formation (*B6*). In addition, we have established protocols for the expression and purification of several VpsR constructs, the highly conserved core RNA Polymerase of *E.coli*, and σ-factors of *V.cholerae* likely involved in promoter recognition and transcription initiation (Fig. 2.18). By using DNase I footprinting as an analytical tool, the characterization of protein protected DNA motifs in different binding reactions would help to identify not only the VpsT-specific DNA recognition sequence, but also potential cooperativity in factor binding, σ-factor specificity, and repositioning of complex components *en route* to transcription initiation.

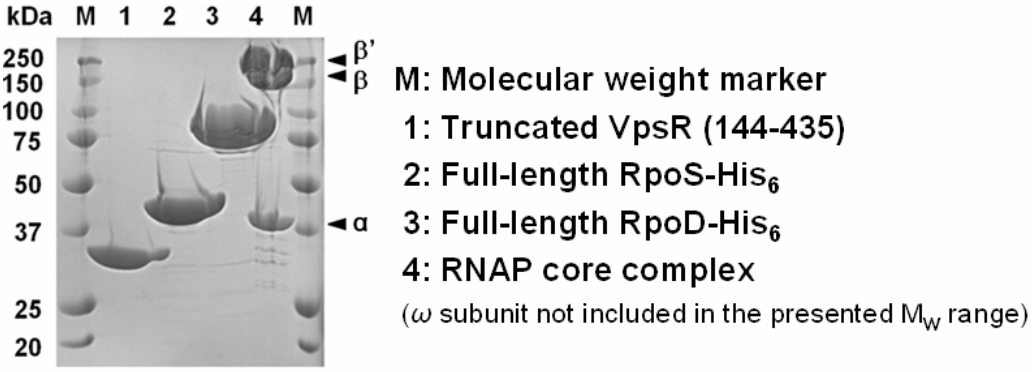


Figure 2.18: Purified VpsR, σ-factors, and core RNA Polymerase. Proteins were expressed in *E.coli* BL21 cells and purified through NiNTA affinity chromatography. RpoS and RpoD σ-factors, as well as the VpsR construct were cloned out of genomic *Vibrio cholerae* DNA. Core RNA Polymerase was derived from *Escherichia coli* and shares almost complete sequence identity with its *Vibrio cholerae* homolog (*B9*).

2.B2. Determining the structure of c-di-GMP•VpsT•DNA complexes

We have presented here the crystal structures of VpsT in its nucleotide-free and c-di-GMP bound states (*B6*). As discussed in the previous section, the two structures are almost identical apart from a small rigid body displacement and minor side chain rearrangements in the nucleotide-binding pocket. The near-perfect superposition of the symmetry-unrelated protomers and the large area of buried surface at the interdomain interface, indicate that the observed relative orientation of the two domains is likely preserved in solution, especially considering the high percent (67%) solvent content in the VpsT crystals (*B27, B28*).

Identification of minimal DNA sequences capable of high affinity binding to apo-VpsT or c-di-GMP•VpsT is crucial for pursuing the structural determination of protein-DNA complexes in crystallographic studies. Oligonucleotides selected in high-stringency *in vitro* selection would be used in crystallization trials with commercially available sparse-matrix screens and following established protocols. In addition, Small Angle X-ray Scattering (SAXS) can be used as an in-solution approach for the determination of low-resolution structural models. SAXS-derived distance distribution functions, molecular weights, and *ab initio* envelope models provide a fast and powerful tool to explore the assembly states and domain-domain orientation of proteins and nucleoprotein complexes, sensitive to detect even subtle conformational changes or multiple conformations in solution (*B29-B31*).

Based on the relative orientation of the DNA-binding domains of VpsT in the crystal structures, we hypothesize that its regulatory function on DNA leads to major changes in DNA architecture, such as DNA bending and/or looping upon promoter binding and nucleotide recognition (*B6*). The fact that nucleotide-bound VpsT recognizes multiple and relatively remote regions in its cognate *vpsL* promoter is also consistent with such a hypothesis. Both modes of DNA deformations have been

previously described as regulatory modes in modulating bacterial gene expression (*B32-B35*). Looping has been established as the primary mode of action for the λ repressor and AraC, whereas proteins such as the catabolite activator protein are shown to introduce a 90° bend in the DNA helix upon protein-DNA complex formation. To test our hypothesis and reveal structural changes in DNA upon VpsT engagement, negative-stain electron microscopy would be a primary method of choice to directly image the structure of naked and c-di-GMP•VpsT-bound DNA (*B36, B37*). Briefly, samples would be spotted onto poly-lysine-coated carbon-formvar grids and negatively stained with uranyl acetate. Wild-type VpsT and VpsT point mutants characterized by different oligomerization potential would be used in parallel preparations. If necessary, DNA-bound VpsT can be further visualized by gold immunolabeling or similar enhancing approaches (*B38*).

As an alternative approach to electron microscopy, atomic force microscopy (AFM) can be used to directly visualize protein-induced changes in DNA architecture (*B39, B40*). Advantages of the technique include the ease of sample preparation, where mica sample carriers require only a positively charged coating (polylysine or poly-L-ornithine), and no subsequent sample staining or enhancement. As recently shown for similar protein-promoter DNA systems (*B41*), the inherent high resolution and three-dimensional information that the method provides could also serve to directly derive the oligomeric state of DNA-bound VpsT protein variants.

2.B3. Structure-function analyses of *V. cholerae* VpsR and *P. aeruginosa* FleQ

As mentioned above, VpsT functions in concert with VpsR, an AAA+ domain-containing protein similar in sequence and domain organization to the c-di-GMP dependent transcription factor FleQ of *Pseudomonas aeruginos*a. The two proteins contain a relatively divergent N-terminal response receiver domain, followed

by more conserved AAA+ σ^{54}-interaction domain and a C-terminal helic-turn-helix DNA-binding motif (*B42, B43*).

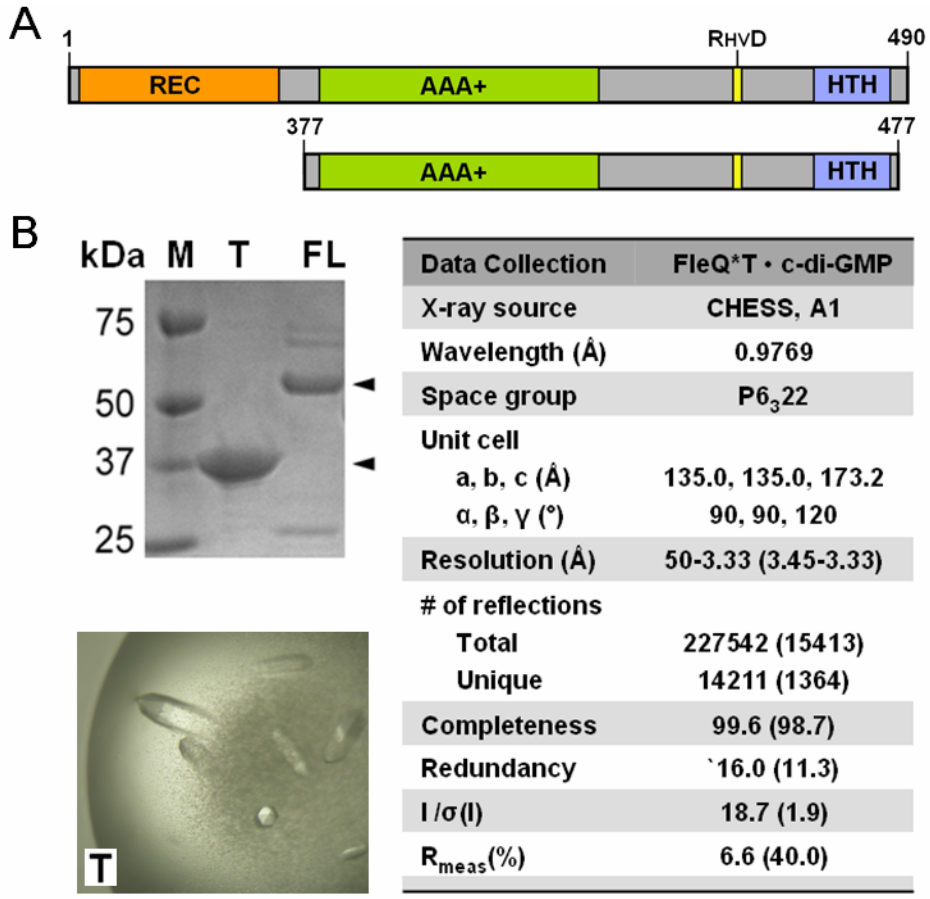
When bound to DNA, FleQ acts as a master activator of flagellar gene expression, as well as a repressor of the *pel* exopolysaccharide synthesis operon (*B44*, *B2*). C-di-GMP binding to the protein is an "off" switch for DNA binding and thus leads to motility inhibition and biofilm formation. Nucleotide recognition is shown to occur independently of the N-terminal receiver domain, but other than that nothing is known about the mode of interaction (*B2*). Unlike VpsR, FleQ of *Pseudomonas aeruginosa* appears to function independently of a VspT protein homolog.

Similarly to the function of VpsT *in vivo*, gene expression under the control of VpsR appears dependent on a c-di-GMP signaling input (*B45*). Whether VpsR, like FleQ, is on itself a c-di-GMP sensing effector, or VpsT function suffices to relay the second messenger's signal remains to be experimentally determined. While the former scenario would identify VpsR as a novel c-di-GMP receptor, the latter would suggest direct interaction between the two proteins, which would likely occur in a c-di-GMP dependent manner.

To this end, we have expressed full-length FleQ as well as truncated FleQ and VpsR constructs lacking the divergent N-terminal receiver domains (Fig. 2.18 and Fig. 2.19). As a positive control, we confirmed that FleQ binds c-di-GMP, and that binding is independent of the N-terminal receiver domain. An established SEC/HPLC coupled chromatography assay and/or ITC would be useful to determine binding of c-di-GMP to VpsR. Crystallographic trials using commercially available sparse-matrix screens would be used in attempts toward crystallization and structure solving.

To date, we have identified preliminary conditions for the crystalization of truncated FleQ (137-477) (Fig. 2.18) The native protein crystallizes in the presence of c-di-GMP in a $P6_322$ space group (Fig. 2.18B). The crystals are characterized by

extremely high solvent content (79%), resulting in suboptimal anisotropic diffraction. Experimental phases were obtained by Single-wavelength Anomalous Diffraction (SAD) experiments on crystals grown from selenomethionine-derivatized protein (data not shown). Further model building, refinement, and functional assays would be necessary to determine biologically significant interfaces, residues participating in c-di-GMP coordination, and the ligand's effects on FleQ structure and function.



Data Collection	FleQ*T • c-di-GMP
X-ray source	CHESS, A1
Wavelength (Å)	0.9769
Space group	$P6_322$
Unit cell	
a, b, c (Å)	135.0, 135.0, 173.2
α, β, γ (°)	90, 90, 120
Resolution (Å)	50-3.33 (3.45-3.33)
# of reflections	
Total	227542 (15413)
Unique	14211 (1364)
Completeness	99.6 (98.7)
Redundancy	`16.0 (11.3)
I /σ(I)	18.7 (1.9)
R_{meas}(%)	6.6 (40.0)

Figure 2.19: Expression, purification and preliminary structural studies on FleQ of *Pseudomonas aeruginosa.*
(A) Domain diagram of full-length (FL) and truncated (T) FleQ with preserved functions in c-di-GMP sensing and DNA recognition. The linker region between the AAA+ σ-54 interacting domain and the helix-turn-helix motif contains a conserved RxxD motif. As shown by us and others, similar motifs serve as allosteric c-di-GMP binding sites in active diguanylate cyclases or divergent homologs.
(B) Preliminary structural studies: Expression, crystalization and data collection statistics.

2.B4. Detection of putative VpsT-VpsR interactions.

The way VpsT and VpsR work together and with the core transcriptional machinery in the cell remains largely unknown. One possibility, supported by the fact that the two proteins regulate mostly the same sets of genes *in vivo*, is that VpsT and VpsR directly interact to translate the c-di-GMP message in controlled expression of biofilm determinants.

Surprisingly, analysis of the R*ΔvpsT*, R*ΔvpsR*, and R*ΔvpsTΔvpsR* mutants (gene knockouts in the rugose morphotype) has suggested that the two proteins have distinct roles in determining biofilm architecture (*B8*). All three deletions cause rugose to smooth colony conversion, and the R*ΔvpsR* and R*ΔvpsTΔvpsR* mutants develop identically, with only small microcolonies or single cells attached to the substratum. In contrast, under certain growth conditions R*ΔvpsT* can develop well-differentiated biofilms and the magnitude of the deletion effects on gene expression is by and large lower than that for the R*ΔvpsR* or R*ΔvpsTΔvpsR* mutants (*B8*). This presents several possible scenarios for the regulation of the VpsT/VpsR system. It is plausible that VpsR is essential and sufficient for biofilm formation, has c-di-GMP binding capacity similar to that of FleQ in *Pseudomonas aeruginosa* (*B2*), but utilizes VpsT as an activity enhancing partner through direct or indirect protein interactions. Alternatively, VpsR could be sufficient for the basal transcription of biofilm determinants but would require VpsT to inhibit motility and boost biofilm formation as a response to elevated c-di-GMP levels. As discussed in the previous chapter, the *Vibrio cholerae* genome encodes for an additional VpsT/CsgD-like protein, VC0396, which might compensate for some of the VpsT functions in the R*ΔvpsT* background. VC0396 contains a conserved WLPR motif at the putative c-di-GMP binding site (*B42*), which is a sequence capable of c-di-GMP recognition when introduced in the nucleotide-binding pocket of *Vibrio cholerae* VpsT (data not shown).

To test for putative VpsT-VpsR interactions, a variety of *in vitro* or cell-based assays could be employed. Pull-down experiments, using differentially tagged purified proteins, or gel filtration and gradient co-sedimentation experiments with tag-free protein constructs, are among the various available *in vitro* approaches. Similar isolated systems present the opportunity to screen for a variety of binding conditions, such as the absence or presence of c-di-GMP and the differential dimerization of VpsT by the use of appropriate interface mutants. It is possible however that the native formation of a VpsT/VpsR complex is further controlled by post-translational modifications, small molecule binding, or cooperative interactions in higher-order protein or protein-DNA complexes. In addition, purification of heterologously expressed full-length VpsR has proved challenging (data not shown), indicating that the protein might need additional cellular factors for stability. To account for this, differentially tagged VpsT and VpsR can be expressed from vectors allowing inducible heterologous co-expression for *in situ* complex formation. Alternatively, functional *vpsT*- and *vpsR-tag* fusions can be introduced in the genome of the R*ΔvpsTΔvpsR Vibrio* mutant to allow for protein expression under the control of native promoters at the genes' original locations. *In situ* complex formation can then be detected by metal affinity pull-downs or co-immunoprecipitation depending on the used protein-tag fusions. The latter approach would also provide the opportunity to screen for VpsT/VpsR complex formation at different growth conditions or developmental stages and, if coupled to mass-spectroscopic analysis, to identify additional binding partners for the two master regulators of biofilm formation.

REFERENCES

B1. Jenal, U. and J. Malone (2006). "Mechanisms of cyclic-di-GMP signaling in bacteria." *Annu Rev Genet* **40**: 385-407.

B2. Hickman, J. W. and C. S. Harwood (2008). "Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor." *Mol Microbiol* **69**(2): 376-89.

B3. Leduc, J. L. and G. P. Roberts (2009). "Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of Xanthomonas axonopodis pv. citri." *J Bacteriol* **191**(22): 7121-2.

B4. Chin, K. H., Y. C. Lee, et al. "The cAMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in Xanthomonas campestris." *J Mol Biol* **396**(3): 646-62.

B5. Sudarsan, N., E. R. Lee, et al. (2008). "Riboswitches in eubacteria sense the second messenger cyclic di-GMP." *Science* **321**(5887): 411-3.

B6. Krasteva, P. V., J. C. Fong, et al. "Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP." *Science* **327**(5967): 866-8.

B7. Yildiz, F. H., N. A. Dolganov, et al. (2001). "VpsR, a Member of the Response Regulators of the Two-Component Regulatory Systems, Is Required for Expression of vps Biosynthesis Genes and EPS(ETr)-Associated Phenotypes in Vibrio cholerae O1 El Tor." *J Bacteriol* **183**(5): 1716-26.

B8. Beyhan, S., K. Bilecen, et al. (2007). "Regulation of rugosity and biofilm formation in Vibrio cholerae: comparison of VpsT and VpsR regulons and epistasis analysis of vpsT, vpsR, and hapR." *J Bacteriol* **189**(2): 388-402.

B9. Larkin, M. A., G. Blackshields, et al. (2007). "Clustal W and Clustal X version

2.0." *Bioinformatics* **23**(21): 2947-8.

B10. Yildiz, F. H., X. S. Liu, et al. (2004). "Molecular analysis of rugosity in a Vibrio cholerae O1 El Tor phase variant." *Mol Microbiol* **53**(2): 497-515.

B11. Lin, W., G. Kovacikova, et al. (2007). "The quorum sensing regulator HapR downregulates the expression of the virulence gene transcription factor AphA in Vibrio cholerae by antagonizing Lrp- and VpsR-mediated activation." *Mol Microbiol* **64**(4): 953-67.

B12. Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." *J Mol Biol* **372**(3): 774-97.

B13. Maris, A. E., M. R. Sawaya, et al. (2002). "Dimerization allows DNA target site recognition by the NarL response regulator." *Nat Struct Biol* **9**(10): 771-8.

B14. Brombacher, E., C. Dorel, et al. (2003). "The curli biosynthesis regulator CsgD co-ordinates the expression of both positive and negative determinants for biofilm formation in Escherichia coli." *Microbiology* **149**(Pt 10): 2847-57.

B15. Maris, A. E., M. Kaczor-Grzeskowiak, et al. (2005). "Primary and secondary modes of DNA recognition by the NarL two-component response regulator." *Biochemistry* **44**(44): 14538-52.

B16. Blackwell, T. K. and H. Weintraub (1990). "Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection." *Science* **250**(4984): 1104-10.

B17. Djordjevic, M. (2007). "SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways." *Biomol Eng* **24**(2): 179-89.

B18. Hudson, J. M. and M. G. Fried (1990). "Co-operative interactions between the catabolite gene activator protein and the lac repressor at the lactose promoter." *J Mol Biol* **214**(2): 381-96.

B19. Bokal, A. J. t., W. Ross, et al. (1995). "The transcriptional activator protein

FIS: DNA interactions and cooperative interactions with RNA polymerase at the Escherichia coli rrnB P1 promoter." *J Mol Biol* **245**(3): 197-207.

B20. Schumacher, M. A., M. C. Miller, et al. (2002). "Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR." *Embo J* **21**(5): 1210-8.

B21. Balaeff, A., L. Mahadevan, et al. (2004). "Structural basis for cooperative DNA binding by CAP and lac repressor." *Structure* **12**(1): 123-32.

B22. Wade, J. T., K. Struhl, et al. (2007). "Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization." *Mol Microbiol* **65**(1): 21-6.

B23. Robertson, G., M. Hirst, et al. (2007). "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing." *Nat Methods* **4**(8): 651-7.

B24. Park, P. J. (2009). "ChIP-seq: advantages and challenges of a maturing technology." *Nat Rev Genet* **10**(10): 669-80.

B25. Hampshire, A. J., D. A. Rusling, et al. (2007). "Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands." *Methods* **42**(2): 128-40.

B26. Kinney, J. B., A. Murugan, et al. (2010). "Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence." *Proc Natl Acad Sci U S A* **107**(20): 9158-63.

B27. Matthews, B. W. (1968). "Solvent content of protein crystals." *J Mol Biol* **33**(2): 491-7.

B28. Kantardjieff, K. A. and B. Rupp (2003). "Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals." *Protein Sci* **12**(9): 1865-71.

B29. Koch, M. H., P. Vachette, et al. (2003). "Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution." *Q Rev Biophys* **36**(2): 147-227.

B30. Putnam, C. D., M. Hammel, et al. (2007). "X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution." *Q Rev Biophys* **40**(3): 191-285.

B31. Kulshina, N., N. J. Baird, et al. (2009). "Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch." *Nat Struct Mol Biol* **16**(12): 1212-7.

B32. Schultz, S. C., G. C. Shields, et al. (1991). "Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees." *Science* **253**(5023): 1001-7.

B33. Schleif, R. (1992). "DNA looping." *Annu Rev Biochem* **61**: 199-223.

B34. Matthews, K. S. (1992). "DNA looping." *Microbiol Rev* **56**(1): 123-36.

B35. Adhya, S. (1989). "Multipartite genetic control elements: communication by DNA loop." *Annu Rev Genet* **23**: 227-50.

B36. Griffith, J., A. Hochschild, et al. (1986). "DNA loops induced by cooperative binding of lambda repressor." *Nature* **322**(6081): 750-2.

B37. Su, W., S. Porter, et al. (1990). "DNA-looping and enhancer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial glnA promoter." *Proc Natl Acad Sci U S A* **87**(14): 5504-8.

B38. Stirling, J. W. (1990). "Immuno- and affinity probes for electron microscopy: a review of labeling and preparation techniques." *J Histochem Cytochem* **38**(2): 145-57.

B39. Lyubchenko, Y. L., L. S. Shlyakhtenko, et al. (1997). "Atomic force microscopic demonstration of DNA looping by GalR and HU." *Nucleic Acids*

Res **25**(4): 873-6.

B40. Rippe, K., M. Guthold, et al. (1997). "Transcriptional activation via DNA-looping: visualization of intermediates in the activation pathway of E. coli RNA polymerase x sigma 54 holoenzyme by scanning force microscopy."UI **270**(2): 125-38.

B41. Wang, H., L. Finzi, et al. (2009). "AFM studies of lambda repressor oligomers securing DNA loops." *Curr Pharm Biotechnol* **10**(5): 494-501.

B42. Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." *J Mol Biol* **215**(3): 403-10.

B43. Marchler-Bauer, A., J. B. Anderson, et al. (2009). "CDD: specific functional annotation with the Conserved Domain Database." *Nucleic Acids Res* **37**(Database issue): D205-10.

B44. Arora, S. K., B. W. Ritchings, et al. (1997). "A transcriptional activator, FleQ, regulates mucin adhesion and flagellar gene expression in Pseudomonas aeruginosa in a cascade manner." *J Bacteriol* **179**(17): 5574-81.

B45. Lim, B., S. Beyhan, et al. (2006). "Cyclic-diGMP signal transduction systems in Vibrio cholerae: modulation of rugosity and biofilm formation." *Mol Microbiol* ***60***(2): 331-48.

CHAPTER 3

Preface

Pseudomonas fluorescens, an important biocontrol microorganism, provides an attractive system to study bacterial biofilm formation. While its ability to form homobiofilms indicates that all components necessary for this signaling process are encoded in its own genome, key proteins involved in the different stages of bacterial attachment, macrocolony formation and biofilm dispersal have been recently identified.

In the following work, a collaborative effort between the Sondermann Lab at Cornell University and the O'Toole Lab at Dartmouth Medical School, we provide a complete mechanistic description for c-di-GMP mediated inside-out signaling, a novel paradigm in bacterial signal transduction. Cytoplasmic c-di-GMP levels, sensitive to environmental substrate availability, are sensed by a transmembrane receptor relaying the signal to a periplasmic effector partner, whose proteolytic activity determines biofilm initiation and dispersal. Such level of resolution, elucidating an entire c-di-GMP-mediated control circuit, from environmental signal to molecular output, is unprecedented for any other c-di-GMP pathway in the current literature.

Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis *

3.1. Abstract

The bacterial second messenger c-di-GMP has emerged as a central regulator for biofilm formation. Increased cellular c-di-GMP levels lead to stable cell attachment, which in *P. fluorescens* requires the widely conserved, transmembrane receptor LapD. Cyclic di-GMP binding to its degenerate phosphodiesterase domain is communicated via a HAMP relay to the periplasmic domain, triggering a sequestration of the protease LapG, thus preventing cleavage of the surface adhesin LapA. Here, we elucidate the mechanism of autoinhibition and activation of LapD. In the absence of c-di-GMP, the intracellular module is held in an inactive state that is disrupted by c-di-GMP binding. The active state involves dimerization of c-di-GMP-bound phosphodiesterase domains via conserved interactions found in c-di-GMP-degrading enzymes. Efficient mechanical coupling of the conformational changes across the membrane is realized through an extensively domain-swapped periplasmic fold. Our analyses identified a conserved system for the regulation of periplasmic proteases, common in many free-living and pathogenic species.

---

* Reproduced from [Marcos V.A.S. Navarro ¶, Peter D. Newell ¶, Petya V. Krasteva ¶, Debashree Chatterjee ¶, George A. O'Toole, and Holger Sondermann (*submitted*)]

¶ These authors contributed equally to this work. Author contributions are as follows: M.V.A.S.N., P.V.K., D.C., P.D.N., G.A.O. and H.S. designed research; M.V.A.S.N., P.V.K., D.C., and P.D.N. performed research; M.V.A.S.N., P.V.K., D.C., P.D.N., G.A.O. and H.S. analyzed data; and P.V.K. and H.S. wrote the manuscript.

3.2 Introduction

Bacterial biofilms arise from planktonic microbial cells that attach to surfaces and form sessile multicellular communities, relevant for survival in hostile habitats and for bacterial pathogenesis (*1*). Recent work has identified biofilm formation as a multiphase process with strict temporal and spatial regulation, often accompanied by adaptational strategies such as phenotypic variation, development of antibiotic resistance, and virulence gene expression (*2, 3*). On the cellular level, functional differentiation including changes in motility, cell adhesion, and secretion are among the many processes driving bacterial biofilm formation. Such a plethora of physiological responses inevitably begs the question of how regulation is achieved, and a nucleotide unique to bacteria, cyclic dimeric GMP (c-di-GMP), has emerged as a key signaling molecule in this process (*4, 5*)

Cyclic di-GMP is a monocyclic RNA dinucleotide that functions as an intracellular second messenger exerting control at the transcriptional, translational, and posttranslational levels (*6*). It is generated from two GTP molecules by GGDEF domain-containing diguanylate cyclases, and degraded by phosphodiesterases (PDEs) containing either EAL or HD-GYP protein domains (*7-10*). Once generated, c-di-GMP is believed to act in a protein-bound form, eliciting localized, rather than more general, diffusive signals (*5*). To date, few but strikingly diverse c-di-GMP receptors have been identified. Protein domains involved in c-di-GMP signal recognition include PilZ domains (*11, 12*), a non-canonical receiver domain in VpsT of *Vibrio cholerae* (*13*), the AAA σ54 interaction domain-containing transcription factor FleQ of *Pseudomonas aeruginosa* (*14*), and the cyclic nucleotide monophosphate-binding domain in Clp of *Xanthomonas campestris* (*15*). In other cases, c-di-GMP turnover domains can also serve as sensors for the nucleotide. For example, in GGDEF domain-containing proteins, an RxxD motif can serve as a c-di-GMP-binding

inhibitory site to regulate either the activity of active enzymes (e.g. PleD of *Caulobacter crescentus*, WspR of *Pseudomonas aeruginosa*) (*16, 17*), or to mediate protein-protein interactions in degenerate homologs (e.g. PelD of *Pseudomonas aeruginosa*, CdgG of *Vibrio cholerae*) (*18,19*).

Proteins participating in c-di-GMP-mediated signal transduction are often composed of multiple domains, allowing for a variety of regulatory inputs, signal ramifications and/or physiological responses (*20*). For example, a large number of bacterial proteins contain both, GGDEF and EAL domains in the same polypeptide chain. These proteins fall into three main categories based on their catalytic activity: Tandem domain-containing proteins with both, diguanylate cyclase and phosphodiesterase activity; proteins with only one active domain, in which the degenerate, inactive domain serves a regulatory function; and proteins in which both domains are degenerate, likely to work as c-di-GMP receptors (*21, 22*). Despite the vast abundance of this signaling module, structural and mechanistic insight regarding their function and regulation is sparse.

The transmembrane protein LapD belongs to the last group, containing degenerate GGDEF and EAL domains that lack catalytic activity, but capable of c-di-GMP binding via its divergent phosphodiesterase domain (*23*). LapD is required for stable cell attachment and biofilm formation in *Pseudomonas fluorescens* and *Pseudomonas putida* (*24-26*), where it controls the localization of the cell surface adhesin LapA in a c-di-GMP-dependent manner (*23, 26*). Binding of c-di-GMP to the LapD EAL domain is relayed to the periplasmic output domain through an inside-out signaling mechanism that utilizes a juxtamembrane HAMP domain, a relay module often found in bacterial transmembrane receptors. As a result, the amount and stability of LapA at the cell surface is increased, leading to stable cell attachment and biofilm formation (*23*).

Recent work by Newell *et al.* reveals the complete c-di-GMP signaling circuit by which LapD controls cell attachment in response to phosphate availability (*27*). For wild-type LapD, c-di-GMP binding appears to induce a conformational change, which activates the receptor. As a consequence, the affinity of the periplasmic domain towards the cysteine protease LapG increases, limiting its access to LapA. Perturbations in the HAMP domain by deletion of some key elements yield a constitutively active receptor, independent of nucleotide binding. However, what prevents LapD from adopting an active conformation and how nucleotide binding translates into an output signal has remained unknown.

Here, we elucidate the autoinhibition and activation mechanism of the c-di-GMP receptor LapD from *Pseudomonas fluorescens*. The crystal structure of the cytoplasmic module containing the GGDEF-EAL tandem domains reveals the presence of an autoinhibitory latch formed by a helical extension of the HAMP domain. In this inactive state, the GGDEF domain functions as a lid restricting nucleotide access to the EAL domain module. The crystal structure of dimeric nucleotide-bound EAL domains provides insight in the conformational changes resulting from c-di-GMP binding. Based on the crystal structure of the periplasmic output domain of LapD, we identify functionally important residues and propose a model for the regulation of its activity in inside-out signal transduction. Finally, our structural studies highlight many conserved features that allowed us to identify similar signaling systems in a variety of bacterial strains including common pathogens such as *Vibrio cholerae* and *Legionella pneumophila*.

2.3. Results and Discussion

Autoinhibition of the intracellular module of LapD in the off-state

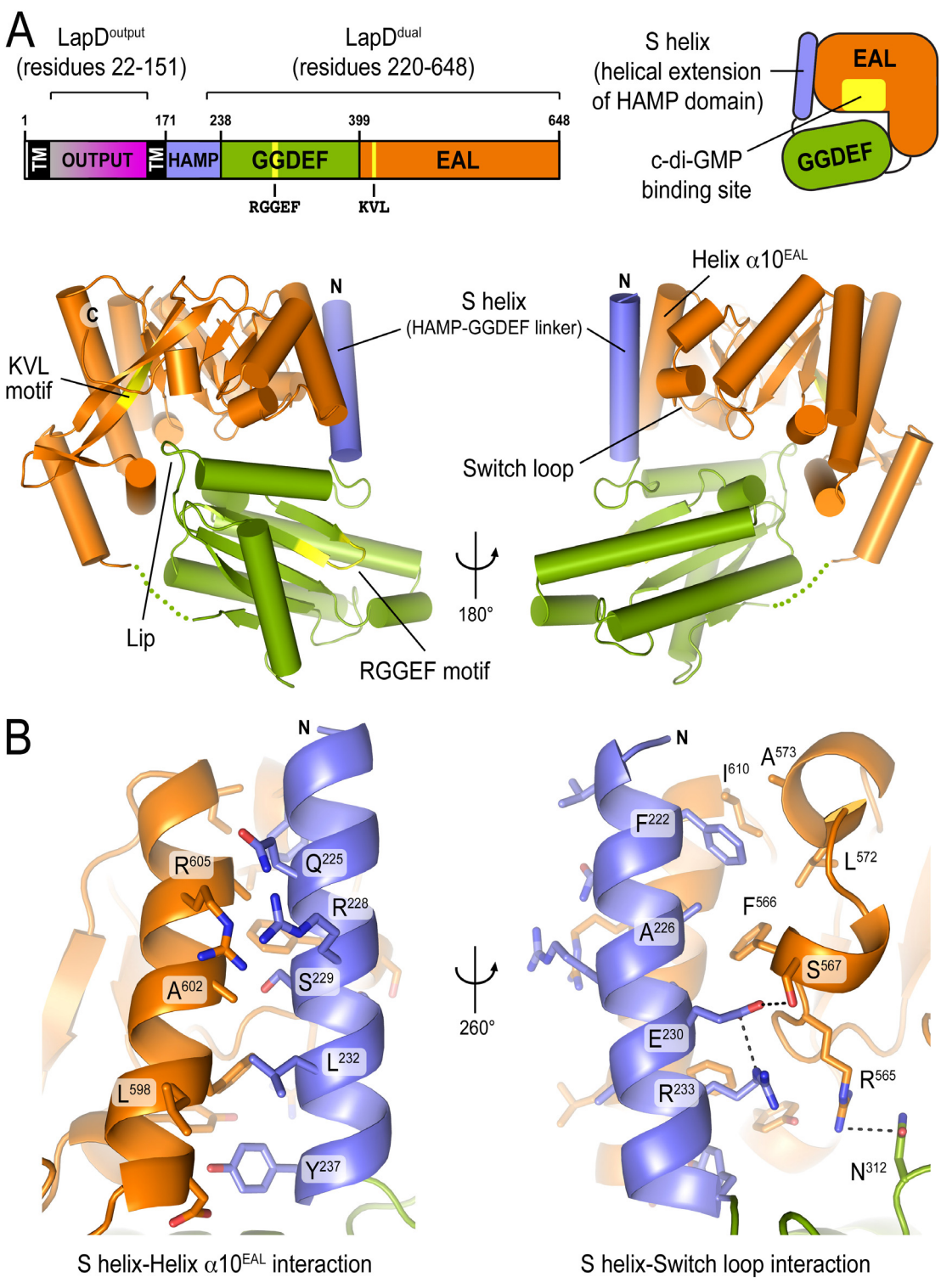
In order to elucidate the molecular mechanism that regulates LapD function, we determined the crystal structure of the intracellular module of *P. fluorescens* LapD, comprising a HAMP-GGDEF domain linker segment and the degenerate GGDEF-EAL domain module ($LapD^{dual}$; residues 220-648) (Fig. 3.1). Based on secondary structure predictions, the linker forms a continuation of the second HAMP domain helix (Fig. 3.8). We will refer to this motif as the signaling or S helix in analogy to helical extensions found in association with other HAMP domains, where they are involved in transducing signals through the HAMP domain to the adjacent signaling modules (*28-30*).

$LapD^{dual}$ was purified to homogeneity by standard chromatography. Crystals were obtained by hanging drop vapor diffusion (space group $P3_2$; one molecule in the asymmetric unit). Selenomethione-substituted protein crystals diffracted X-rays to 2.5 Å (Table S1). The structure was solved by single-wavelength anomalous dispersion (SAD) phasing, and initial phases were extended to higher resolution obtained from native protein crystals. The model was built into the experimentally phased map and further refined at 2.0 Å resolution. We also obtained a second crystal form involving different crystal packing contacts (space group I23, one molecule in the asymmetric unit), yet the overall conformation of LapD in the two crystals is identical (rmsd of 0.9 Å over all atoms; Fig. 3.9A). In both cases, the biologically significant unit was predicted to be a monomer based on energetic and geometric estimations (*31*). In addition, we solved the structure of the isolated EAL domain of LapD bound to c-di-GMP in two different crystal forms (residues 399-648; $LapD^{EAL}$•c-di-GMP; Fig. 3.9B), which will be discussed below.

LapD^{dual} adopts a compact, bilobal conformation (Fig. 3.1A). The GGDEF domain, comprising the N-terminal lobe, caps the nucleotide-binding pocket of the EAL domain, which presents the C-terminal lobe of the tandem-domain structure. The EAL domain buttresses the N-terminal S helix via predominantly hydrophobic interactions, burying 1170 Å^2 (Fig. 3.1). The binding groove on the EAL domain, which accommodates the S helix, consists of the helix α10 and an adjacent loop. The latter has been identified as a conserved motif in catalytically active EAL domain-containing phosphodiesterases, in which it is involved in dimerization, substrate specificity and metal ion coordination in the active site (*32, 33*). This loop was referred to as loop 6 in SadR/RocR and β5-α5 loop in the light-regulated phosphodiesterase BlrP1. In an attempt to introduce a more general nomenclature, we will refer to this motif as the switch loop.

In addition to the S helix-EAL domain interaction, the GGDEF domain contacts the nucleotide-binding surface of the EAL domain at multiple points, forming a loosely packed interface that buries 1620 Å^2 of surface area (Fig. 3.1A; Fig. 3.10A and 3.10B). One such contact, the salt bridge between an arginine residue (R^{450}), located just downstream of the signature EAL motif (KVL in LapD) at the center of the c-di-GMP-binding site, and a glutamate residue (E^{262}), presented by a loop of the GGDEF domain ("lip"), forms a particularly close interaction incompatible with c-di-GMP binding (Fig. 3.10A and 3.10B). Although apo-LapD is in a c-di-GMP-binding incompetent state, the binding site is not completely occluded and may allow for competing out the inhibitory interactions by c-di-GMP (Fig. 3.10B).

Figure 3.1: Autoinhibited structure of the cytoplasmic domain of LapD (Data courtesy of Marcos V.A.S. Navarro).
(A) Crystal structure of apo-LapD^{dual}. The domain organization of LapD from *P. fluorescens* Pf0-1 is shown. The degenerate sequence of the GGDEF and EAL signature motifs are indicated. The crystal structure of the LapD^{dual} (residues 220-648) is shown as ribbon presentation and colored according to the diagrams. The signaling or S helix forms an extension of second HAMP domain helix. The switch loop is sensitive to the nucleotide binding state of the EAL domain and is involved for dimerization and catalysis in active phosphodiesterases. Two views, separated by a 180° rotation, are shown.
(B) The autoinhibitory latch. A close-up view of the S helix-EAL domain interface is shown with residues involved in direct, pair-wise interactions shown as sticks. Two views, separated by a 260° rotation, are shown. Helix α10 and the switch loop form a surface buttressing the S helix.


A
LapD^output
(residues 22-151)
LapD^dual
(residues 220-648)
1
171
238
399
648
TM
OUTPUT
TM
HAMP
GGDEF
EAL
RGGEF
KVL
S helix
(helical extension
of HAMP domain)
EAL
c-di-GMP
binding site
GGDEF
C
N
KVL
motif
S helix
(HAMP-GGDEF linker)
Helix α10^EAL
N
Switch loop
Lip
180°
RGGEF motif
B
N
R605
Q225
R228
S229
A602
L232
L598
Y237
260°
N
I610
A573
F222
L572
F566
A226
S567
E230
R565
R233
N312
S helix-Helix α10^EAL interaction
S helix-Switch loop interaction

The loop that connects the S helix to the GGDEF domain adopts a conformation that is identical to the linkage between active diguanylate cyclase domains and their regulatory domains (Fig. 3.10C). The conformation is stabilized by a salt bridge between two strictly conserved residues that are located at the beginning of the connecting loop and just upstream of the signature GGDEF motif (^{318}RGGEF^{322} in LapD), respectively: Asp^{239} in the loop and Arg^{316} in the GGDEF domain of LapD; Asp^{174} and Arg^{249} in WspR; Asp^{292} and Arg^{366} in PleD (*16, 17, 34, 35*). This interaction likely constraints the loop conformation, restricting the overall rotational freedom of the GGDEF domains relative to their associated regulatory modules, the S helix in the case of LapD and the response receiver domains in the case of PleD and WspR.

In summary, the structural analysis of the cytoplasmic domain of LapD revealed that in the absence of c-di-GMP, the protein resides in an inhibited conformation, with the GGDEF domain forming a lid restricting access of c-di-GMP to the EAL domain. Nucleotide binding would disrupt this conformation likely accompanied by a major conformational change.

Structural model of the on-state: Crystal structure of LapD^{EAL}•c-di-GMP

The crystal structure of LapD^{EAL} bound to c-di-GMP was solved by molecular replacement (Table S1). We obtained crystals in two independent conditions yielding two different crystal forms (space group $C222_1$, one molecule per asymmetric unit; and space group $P6_522$, two molecules per asymmetric unit; Fig. 3.9B). While the majority of the crystal packing contacts were different, both crystal forms maintained a common dimer of EAL domains, and the resulting structures superimposed almost perfectly (rmsd of 0.6 Å over all atoms).

Cyclic di-GMP binding preserved the overall conformation of the EAL domain observed in the apo-LapD^{dual} structure (rmsd of 0.6 angstrom over all atoms) (Fig. 3.2), comparable to the lack of major conformational changes upon nucleotide binding to the EAL domains of YkuI and FimX (*36, 37*). Minor changes in the nucleotide-binding pocket are confined to four c-di-GMP-coordinating residues that adopt an alternate side chain rotamer conformation (Fig. 3.2A).

The most notable conformational change upon c-di-GMP binding occurs in the switch loop (Fig. 3.2B). Nucleotide binding and the absence of the S helix allow the loop to restructure, resulting in the switching of the conserved phenylalanine residue F^{566} (Fig. 3.2B; Fig. 3.8). In apo-LapD^{dual}, the side chain of F^{566} faces inward and is located at the center of the S helix-binding interface (Fig. 3.1B). In contrast, the switch loop adopts a conformation in the c-di-GMP-bound structure positioning F^{566} so that it can participate in homodimerization (Fig. 3.2B and Fig. 3.3). The symmetric LapD^{EAL} domain dimer is reminiscent of the oligomeric state in active EAL domain-containing phosphodiesterases, such as in *Pseudomonas aeruginosa* SadR/RocR, *Bacillus subtilis* YkuI and the BLUF domain-regulated photoreceptor BlrP1 from *Klebsiella pneumoniae*, where dimerization is involved in positioning of the general base for efficient catalysis (Fig.11) (*32, 33, 36*).

Most importantly, dimerization of the c-di-GMP-bound EAL domains is incompatible with the autoinhibited conformation observed in the crystals of LapD^{dual} (Fig. 3.3C). The surface occupied by the S helix overlaps significantly with the homodimerization interface, which indicates that nucleotide-induced conformational changes will include the displacement the GGDEF domain and the S helix. More generally, the preservation of EAL domain dimerization in LapD and the conformational change of the switch loop upon c-di-GMP binding suggest their importance for signaling and regulation in GGDEF-EAL domain containing proteins.

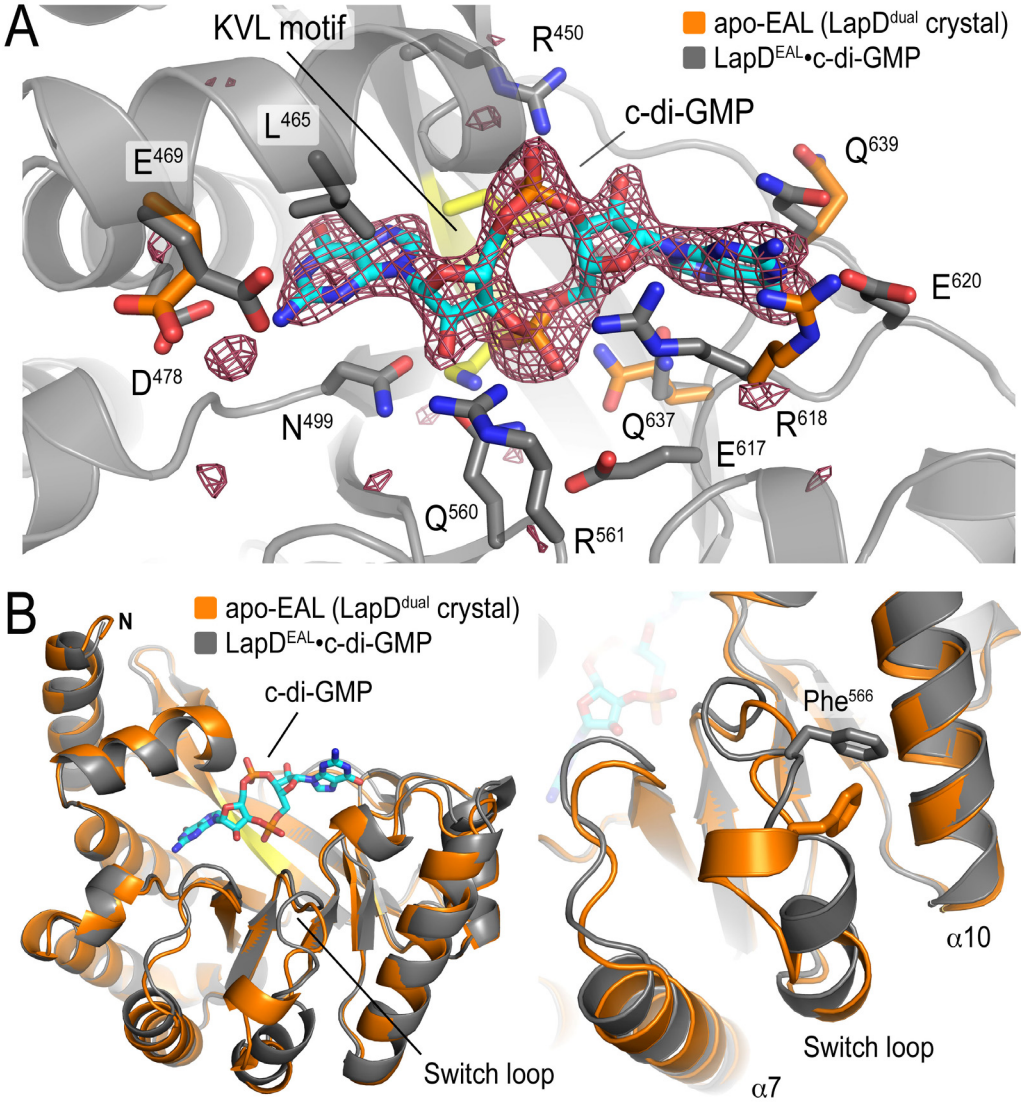
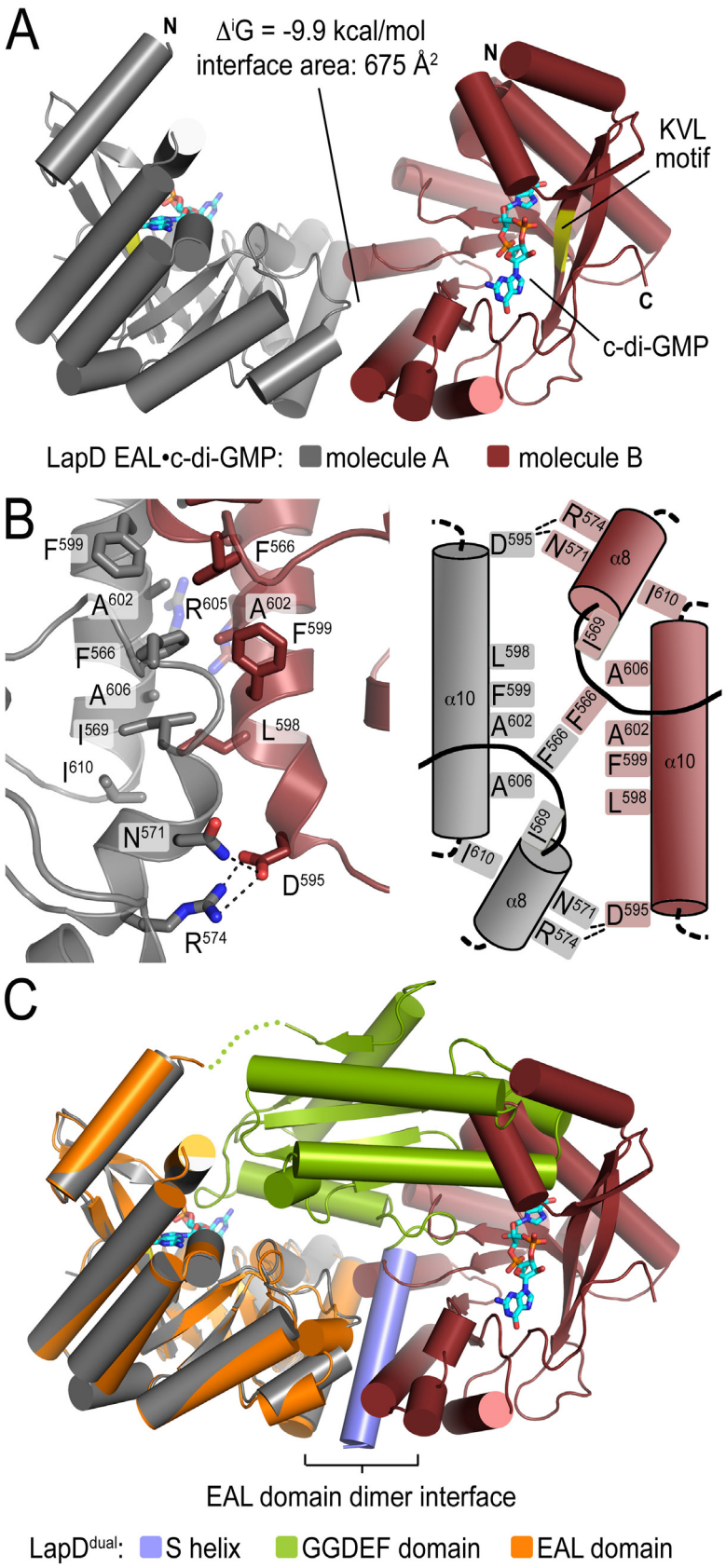


Figure 3.2: Comparison between the nucleotide-free and c-di-GMP-bound EAL domain of LapD.

(A) Crystal structure of LapD^{EAL}•c-di-GMP. The c-di-GMP-bound structure of LapD^{EAL} (gray) was superimposed onto the nucleotide-free structure of LapD^{dual} (orange residues). The S helix and GGDEF domain were omitted for clarity. A close-up view of the nucleotide binding pocket is shown with residues involved in c-di-GMP binding presented as sticks. The ($|F_o|$-$|F_c|$) electron density map is shown as calculated from a model prior to inclusion of nucleotide and is contoured at 3.5σ.

(B) Conformational change of the switch loop. Cyclic di-GMP binding and absence of the S helix allow the switch loop to adopt an alternative conformation (orange: apo-LapD^{dual}; gray: LapD^{EAL}•c-di-GMP). As a consequence, the side chain of Phe^{566}, a residue involved in both, S-helix interaction in LapD^{dual} and dimerization of LapD^{EAL}, changes position.

Figure 3.3: Dimerization of c-di-GMP-bound LapD^{EAL}.
(A) EAL domain dimerization. In both crystal forms obtained for LapD^{EAL}•c-di-GMP we observe symmetric dimerization between protomers involving helix α10 and the switch loop. Dimerization buries 1350 Å^2 of surface area (interface area times 2), and was predicted to be energetically favorable (*31*).
(B) Dimer interface. A close-up view (left panel) and cartoon diagram (right panel) of the dimer interface is shown.
(C) Comparison of apo-LapD^{dual} and LapD^{EAL}•c-di-GMP. The EAL domain from the crystal structure of nucleotide-free LapD^{dual} was superimposed on one c-di-GMP-bound EAL domain from dimeric LapD^{EAL}. LapD^{dual} is colored as shown in Fig. 3.1.


A
N
ΔiG = -9.9 kcal/mol
interface area: 675 Å2
N
KVL
motif
C
c-di-GMP
LapD EAL•c-di-GMP: molecule A molecule B
B
F599
F566
A602
R605
A602
F599
F566
A606
I569
L598
I610
N571
D595
R574
D595
N571
R574
α8
I610
I569
L598
F599
A602
α10
A606
F566
F566
A606
A602
F599
α10
L598
I569
I610
α8
N571
D595
R574
C
EAL domain dimer interface
LapDdual: S helix GGDEF domain EAL domain

Analysis of the autoinhibition and activation mechanism of LapD in solution

Structure-guided mutations were introduced into LapD to assess the functional relevance of the autoinhibitory conformation and EAL domain dimerization (Fig. 3.4A). Site-directed mutations were introduced into the S helix, predicted to weaken its interaction with the EAL domain without affecting EAL domain dimerization propensity ($F^{222}A$, $F^{222}E$, $S^{229}D$, $E^{230}A$ or $L^{232}E$; Fig. 3.1B). Another set of mutations targeted the GGDEF-EAL domain interface, focusing on changes in the GGDEF domain that would not interfere with EAL domain function ($M^{252}E$, $E^{262}A$ or $E^{333}A$; Fig. 3.10A). Finally, A^{602} was identified as a residue at the center of the EAL domain dimerization interface (Fig. 3.3B). The structure of apo-$LapD^{dual}$ showed A^{602} at the periphery of the S helix-EAL interaction, suggesting that perturbations at this site may maintain the autoinhibited state (Fig. 3.1B). To assess how buried the native residues are in the autoinhibited state, surface accessibility of their side chains was calculated based on the crystal structure of apo-$LapD^{dual}$ (Fig. 3.4B).

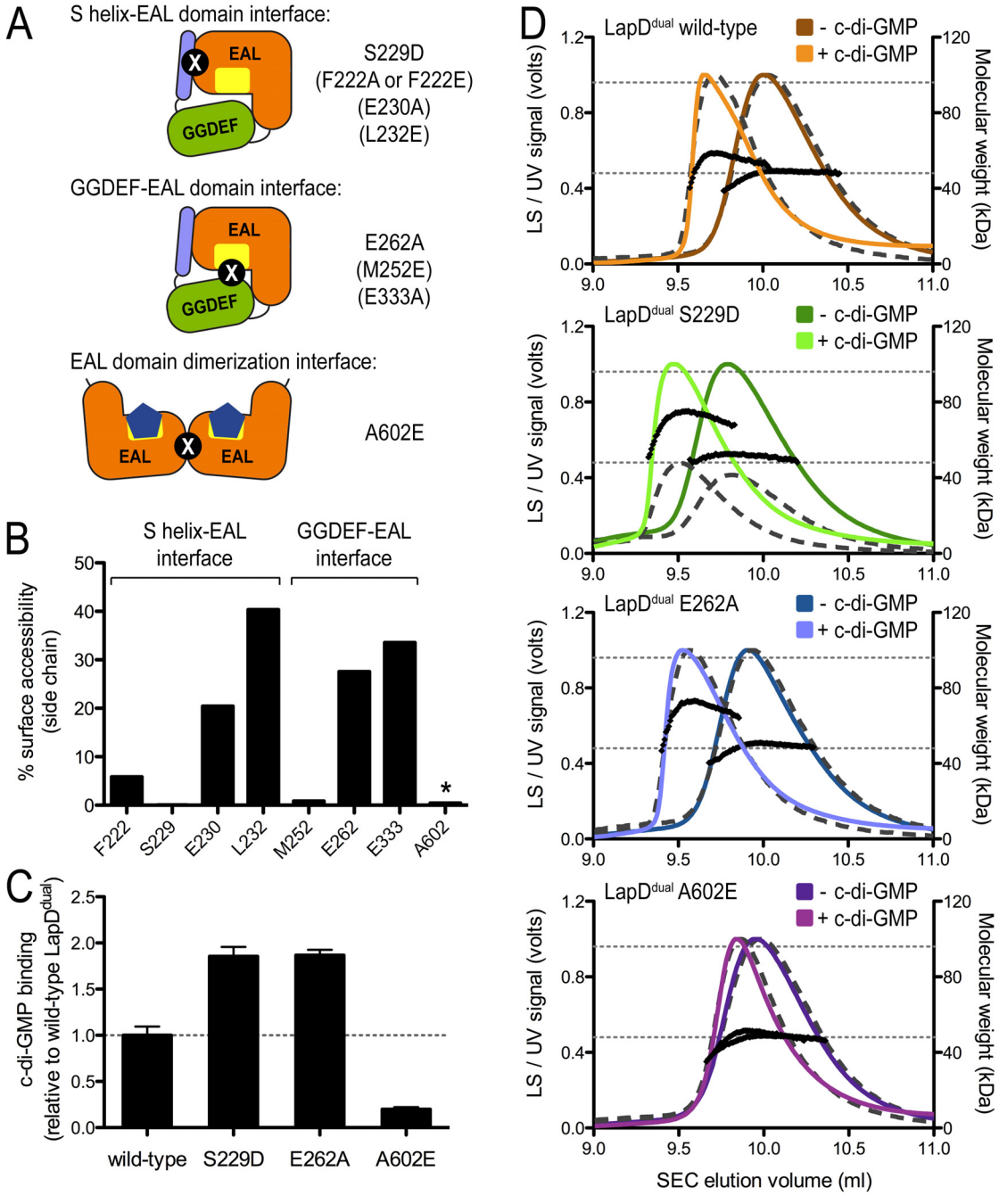
First, we determined the effect of mutating representative residues on nucleotide binding to $LapD^{dual}$ protein variants. Following incubation with c-di-GMP, proteins were subjected to gel filtration for removal of unbound nucleotide. Eluted protein peaks were collected, concentrated, normalized for protein content, and heat denatured to release affinity-bound c-di-GMP. Filtered supernatants were analyzed by reverse-phase high-performance liquid chromatography (*17*), and c-di-GMP binding was compared to that of the wild-type protein (Fig. 3.4C). Both, a single-point mutation in the latch or one in the lip of the GGDEF domain that directly occludes the nucleotide pocket ($S^{229}D$ and $E^{262}A$, respectively), increased c-di-GMP binding to $LapD^{dual}$ by twofold. In contrast, replacing A^{602} with a glutamate residue significantly reduced the amount of c-di-GMP bound to $LapD^{dual}$, arguing toward cooperativity between nucleotide binding and EAL domain dimerization. Similar results were

obtained for the isolated EAL domain mutant (Fig. 3.12B).

We next analyzed the oligomerization state of $LapD^{dual}$ and $LapD^{EAL}$ protein variants in solution, by using static multi-angle laser scattering (Fig. 3.4D and 3.12B). This method provides the absolute molecular weight and hence quaternary structure of proteins eluting from a gel filtration column, and yields comparable results to analytical ultracentrifugation experiments in a significantly shorter experimental time frame (*38*). The wild-type $LapD^{dual}$ protein elutes in a single peak from the size exclusion column with a molecular weight of 47.6 kDa, indicating a monomeric state in solution. Addition of c-di-GMP shifted the peak elution volume and increased the molecular weight slightly to 55.8 kDa, suggesting that nucleotide binding is accompanied by some degree of protein dimerization, although an additional change in molecular shape cannot be ruled out. Considering the c-di-GMP binding data, it is likely that only a fraction of LapD is bound to nucleotide under these conditions (Fig. 3.4C).

Both, the S helix-EAL and the GGDEF-EAL interface mutants ($S^{229}D$ and $E^{262}A$, respectively) showed more distinct shifts in molecular weight upon c-di-GMP binding, being intermediate between pure monomers and dimers (Fig. 3.4D). In general, intermediate molecular weights and the non-gaussian peak shape, as observed for these samples, are indicative of a fast exchange between monomeric and dimeric species, relative to the data acquisition time. In the absence of c-di-GMP, the $E^{262}A$ and more so the $S^{229}D$ mutant proteins elute earlier compared to wild-type but were still predominantly monomeric, which is indicative of a more elongated conformation of these mutants in solution. As predicted on the basis of the structural analysis, $LapD^{dual}$ or $LapD^{EAL}$ containing a glutamate substitution in place of A^{602} are monomeric in solution, independent of the presence of nucleotide (Fig. 3.4D and 3.12B).

Figure 3.4: Cyclic di-GMP binding and quaternary structure of LapD^{dual} in solution.
(A) Mutant categories. Structure-guided, site-directed mutants in LapD are illustrated. Mutations in brackets were used in experiments shown in Fig. 3.5 and Fig. 3.13.
(B) Surface accessibility in LapD^{dual} of residues targeted for mutagenesis. Based on the LapD^{dual} crystal structure, surface accessibility of the side chains was calculated using naccess (*39*). The asterisk marks a residue at the center of the EAL domain dimerization interface.
(C) Nucleotide binding to wild-type and mutant LapD^{dual}. Purified LapD^{dual} (wild-type, $S^{229}D$, $E^{262}A$ or $A^{602}E$) was incubated in the presence of c-di-GMP. Excess nucleotide was removed by gel filtration, and protein-bound c-di-GMP levels were assessed by reverse-phase HPLC after heat-denaturation. Data are expressed relative to the amount bound to wild-type LapD^{dual}.
(D) Oligomerization of LapD^{dual} in solution. Size exclusion chromatography (SEC)-coupled multi-angle light scattering (MALS) analysis of wild-type and mutant LapD^{dual} in presence and absence of c-di-GMP are shown. The signal from the 90°-scattering detector is shown in color and the signal from the refractive index detector is shown as dashed line. Average molecular weights are plotted in black against the right Y-axis, as calculated every second across the protein elution peak. Theoretical molecular weights corresponding to those of a monomer and a dimer are indicated as horizontal dashed, grey lines. Injected protein and nucleotide concentrations were 250 μM and 500 μM, respectively.


A
S helix-EAL domain interface:
EAL
GGDEF
S229D
(F222A or F222E)
(E230A)
(L232E)
GGDEF-EAL domain interface:
EAL
GGDEF
E262A
(M252E)
(E333A)
EAL domain dimerization interface:
EAL
EAL
A602E
B
S helix-EAL interface
GGDEF-EAL interface
% surface accessibility (side chain)
F222
S229
E230
L232
M252
E262
E333
A602
C
c-di-GMP binding (relative to wild-type LapD^dual)
wild-type
S229D
E262A
A602E
D
LapD^dual wild-type
- c-di-GMP
+ c-di-GMP
LapD^dual S229D
LapD^dual E262A
LapD^dual A602E
LS / UV signal (volts)
Molecular weight (kDa)
SEC elution volume (ml)

In summary, LapD appears to be inhibited for efficient nucleotide binding by structural features involving the latch motif and occupancy of the c-di-GMP binding site by the GGDEF domain. Based on the observation that the $A^{602}E$ mutation, located in the EAL domain homodimer interface and outside of the c-di-GMP binding site, renders the protein monomeric and reduces nucleotide binding, we propose that dimerization and c-di-GMP binding are cooperative events in $LapD^{dual}$ and $LapD^{EAL}$. It is important to note that LapD is a dimeric receptor via its HAMP and output domains, and therefore EAL domain dimerization (and nucleotide binding) is likely more pronounced in the context of the full-length receptor.

Effect of structure-based mutations in LapD on biofilm formation

Stable biofilms of *Pseudomonas fluorescens* require LapD expression and the presence of c-di-GMP (*23*). To examine the contribution of inter-domain interactions to LapD's function *in vivo*, full-length LapD variants were assessed for their ability to promote biofilm formation in a Δ*lapD* mutant strain (Fig. 3.5A). We observed a range of phenotypes from a slight reduction in biofilm formation relative to the wild-type, to strong hyper-adherent phenotypes comparable to that observed when LapD is constitutively activated (Fig. 3.5A). The mutation that we predict to disrupt the S helix-EAL interface in the autoinhibited conformation, $S^{229}D$, caused an 'activated' phenotype, consistent with its increased nucleotide binding and dimerization propensity *in vitro* (Fig. 3.4 and 3.5A). Similar results were obtained with the mutant $F^{222}E$, whereas a less invasive alanine substitution was tolerated at this position.

In the apo-$LapD^{dual}$ structure, the E^{262} residue is positioned such that it would occlude binding of c-di-GMP to the EAL domain (Fig. 3.10B). Consistent with this and its increased binding of c-di-GMP, the $E^{262}A$ mutation results in an increase in biofilm formation relative to the wild-type allele (Fig. 3.5A). Yet, the $E^{262}A$ mutant

phenotype is not as extreme as that exhibited in the case of the $S^{229}D$ mutation, despite comparable increases in c-di-GMP binding and dimerization by these proteins *in vitro*. This suggests that the $E^{262}A$ mutant is still able to assume the auto-inhibited conformation *in vivo*, albeit at a lower threshold of c-di-GMP than the wild-type protein. Other mutations showed intermediate ($L^{232}E$, $M^{252}E$) or no significant changes ($F^{222}A$, $E^{230}A$, $E^{333}A$) in phenotype, roughly corresponding to their surface exposure in the autoinhibited state structure (Fig. 3.4B and 3.5A).

The $A^{602}E$ mutation, which disrupts the dimerization interface of the EAL domain and reduces steady state c-di-GMP binding *in vitro* (Fig. 3.4 and Fig. 3.12), led to a small but significant decrease in biofilm formation relative to the wild-type allele (Fig. 3.5A). The observation that the $A^{602}E$ mutant showed a minor loss-of-function *in vivo*, distinct from the more pronounced loss-of-function observed with mutants in the nucleotide binding pocket (*23*), argues that dimerization increases the stability of the nucleotide-bound state rather than being required for c-di-GMP binding per se. While this modest reduction in function *in vivo* seemed incongruous with the severe defect in dimerization and binding exhibited by the dual-domain and EAL domain construct *in vitro*, we further tested its significance by introducing the $A^{602}E$ mutation into activated alleles of LapD, $S^{229}D$ and $F^{222}E$, respectively. The reduction in biofilm formation in the double-mutants was substantial, corroborating that EAL domain dimerization plays a role in LapD function *in vivo* (Fig. 3.13A).

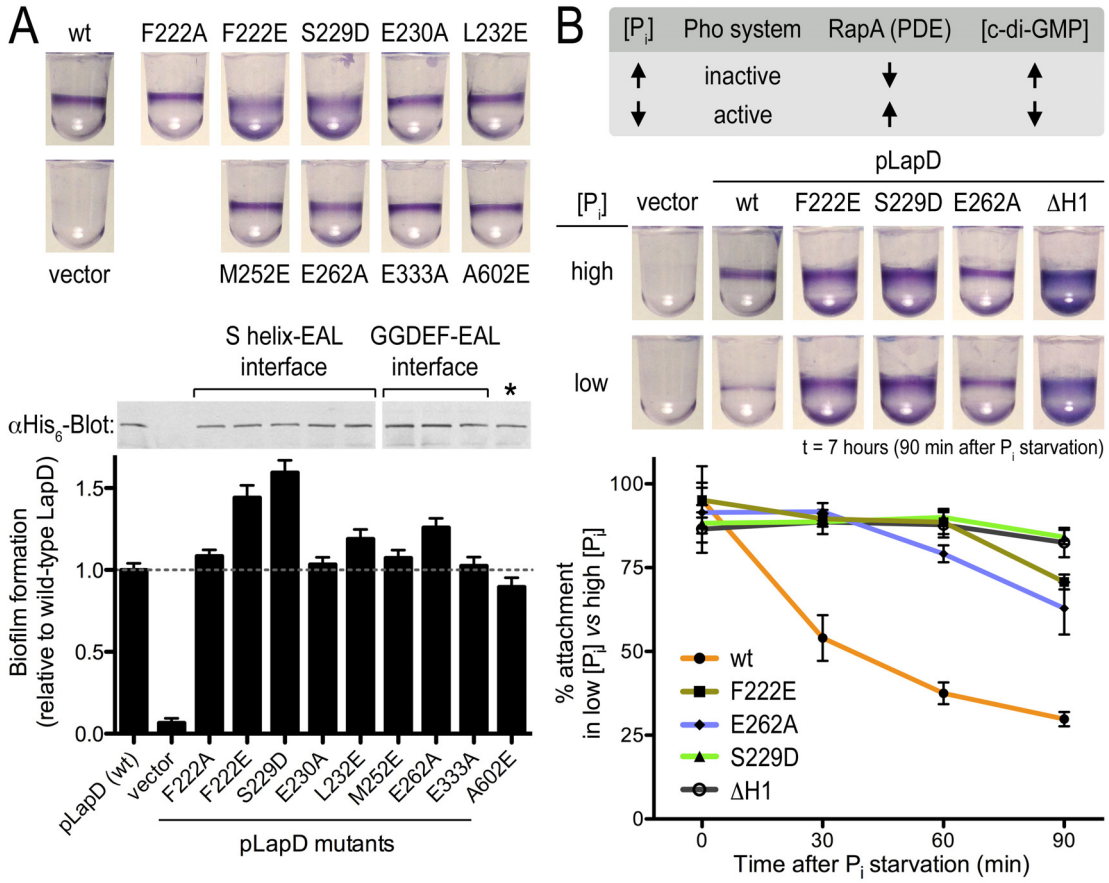


Figure 3.5: Phenotypic analyses of *lapD* mutants (Data courtesy of Peter D. Newell and George A. O'Toole).

(A) Biofilm phenotypes. Biofilm formation of Δ*lapD* cell expressing full-length, wild-type LapD, LapD point mutants or the insert-less expression vector was assessed. Crystal violet-stained biofilms (top) and their quantification (bottom) are shown. Data are means -/+ SD of 8 replicates. Protein levels were determined by Western blotting using a primary antibody that recognizes His_6-epitope at the C-terminus of LapD. The asterisk marks a residue at the center of the EAL domain dimerization interface.

(B) Phosphate-regulated c-di-GMP signaling. Phosphate starvation leads to the expression the active phosphodiesterase RapA and a reduction in cellular c-di-GMP concentration (*23*). LapD mutants were tested for their response to limiting phosphate concentration. Biofilm formation was monitored over 90 min after physiological activation of the Pho system in low-phosphate media, and compared to biofilm formation in phosphate-rich media. The mutant ΔH1 contains an activating deletion in the HAMP domain and has been described previously (*23*).

The single-mutants were also tested for their response to phosphate starvation, a physiological input for LapD-mediated signaling (*23, 40*). Limiting phosphate availability activates the Pho system, which as part of its response upregulates the expression of a c-di-GMP-specific phosphodiesterase, RapA. Reducing cellular c-di-GMP concentration downregulates wild-type LapD activity, which results in the release of the adhesin LapA from the cell surface and thus a reduction in biofilm formation (Fig. 3.5B, top) (*23*). Mutations in the S helix-EAL domain interface ($F^{222}E$, $S^{229}D$) or in the GGDEF lid that relieve occlusion of the c-di-GMP binding site failed to respond to phosphate starvation efficiently showing little to no reduction in biofilm formation (Fig. 3.5B). The effect was comparable to a deletion mutant described previously, in which a helical segment of the HAMP domain was removed, yielding a constitutively active, deregulated receptor (Fig. 3.5B) (*23*). Similar to the trends observed in the static biofilm assay (Fig. 3.5A), other mutants in LapD showed more subtle effects in the phosphate starvation experiments (Fig. 3.13B).

Collectively, these results suggest that the S helix-EAL is the dominant autoinhibitory feature responsible for positioning the GGDEF domain to occlude the c-di-GMP-binding pocket. Together these two interfaces establish autoregulation and appropriate control of LapD activation *in vivo*. In addition, EAL domain dimerization via a conserved mode of interaction is likely to contribute to the efficiency of the signaling system by stabilizing the activated conformation.

Crystal structure of LapD's output domain: A conserved, domain-swapped periplasmic domain

In order to shed light on how changes in the cytosolic domain are sensed in the periplasm, we determined the structure of the entire output domain (residues 22-151; Fig. 3.1A). The protein was purified to homogeneity by using standard

chromatography procedures. Crystals grown with selenomethionine-derivatized protein were obtained by hanging drop vapor diffusion, and diffracted X-rays to a maximum resolution of 1.8 Å (Table S1). The structure was solved by single-wavelength anomalous dispersion (SAD) phasing, using data collected at the selenium anomalous scattering peak wavelength. The final model consists of two molecules per asymmetric unit spanning residues 23-150 (Fig. 3.6A and 3.14A).

The periplasmic output domain of LapD forms an extensively interweaved, domain-swapped dimer sharing 3429 $Å^2$ interfacial surface area (1/3 of LapD's output domain molecular surface) (Fig. 3.6A and 3.14B). The dimer adopts an overall V-shaped conformation. Each arm of the fold consists of two α-helices and two β-strands contributed by one of the two protomers, complemented by two β-strands flanked by helical segments from the other. The N- and C-terminal helices of LapD's output domain presumably connect directly to the transmembrane helices and the HAMP domains. The two half-sites are linked via a long connecting segment that crosses over at the center of the dimer. The two protomers superimpose well except for a subtle rigid body rotation around the linker (Fig. 3.14A).

A DALI search comparing LapD's output domain to proteins in the PDB data base revealed structural similarity of its domain-swapped arms to the periplasmic domain of the sensor histidine kinase CitA (Z-score=5.4, rmsd of 2.5 Å) (*41, 42*). The periplasmic modules of CitA and related proteins show some homology to PAS domains and have been classified as PDC (PhoQ-DcuS-CitA) protein domains (*43, 44*). Such domains occur in many other bacterial transmembrane proteins, but unlike LapD's output domain, they are found to form a variety of regular, non-swapped dimers (*42, 43, 45*).

Figure 3.6: Structure-function analysis of the periplasmic output domain of LapD.

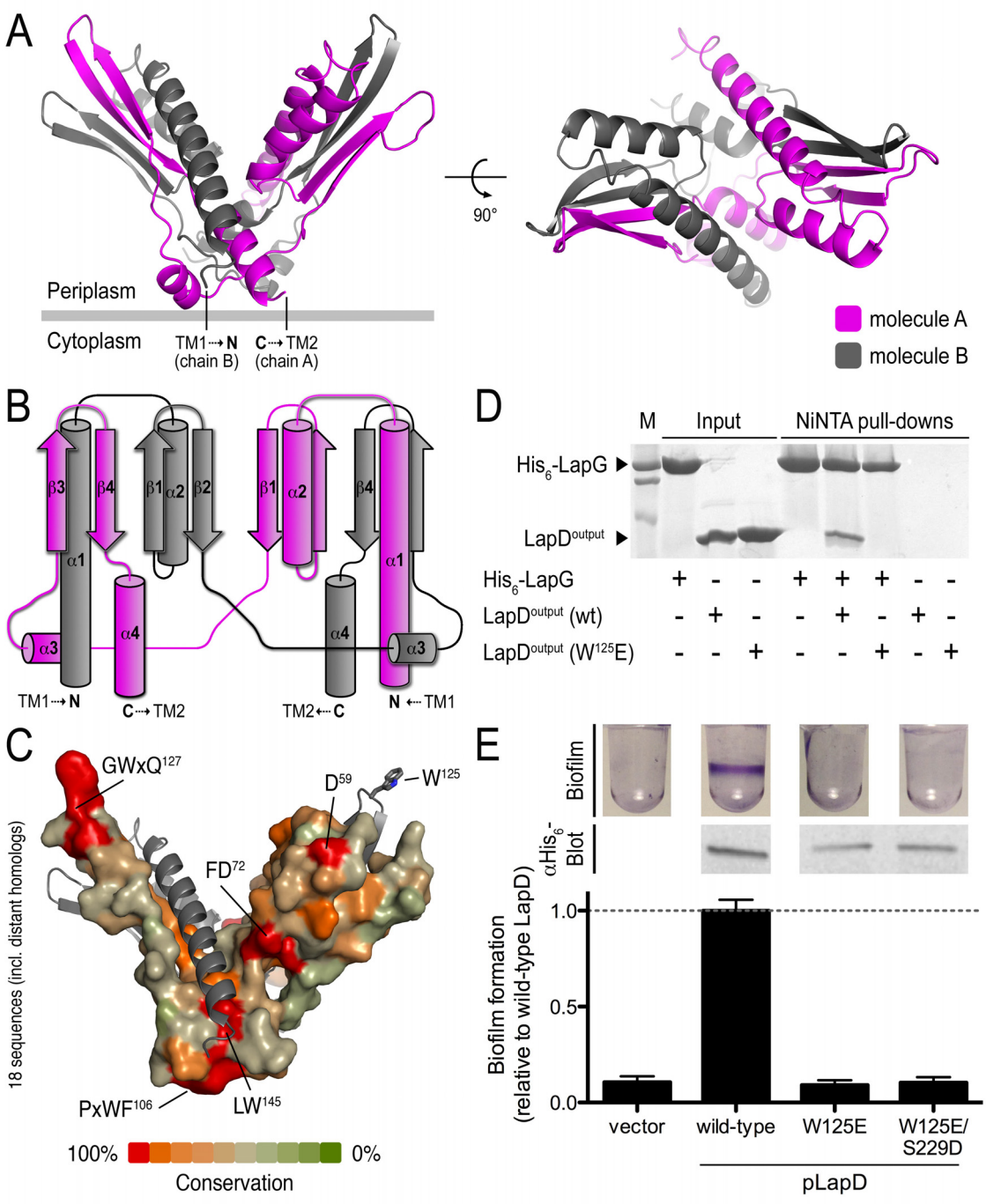
(A) Crystal structure of $LapD^{output}$. The crystal structure of the periplasmic output domain of LapD (residues 22-151) is shown as ribbon presentation with the two protomer chains colored in pink and gray, respectively. The relative position of the inner cell membrane (gray bar) and connection to the flanking transmembrane (TM) helices are indicated. Two orthogonal views are shown. (Protein expression and crystallization by Debashree Chatterjee; data collection and phase solution by Petya V. Krasteva; model building and refinement by Marcos V.A.S. Navarro and Holger Sondermann)

(B) Topology diagram. The diagram illustrates the domain-swapped structure of the dimeric output domain.

(C) Surface conservation. Based on an alignment of 18 sequences of LapD homologs, the sequence conservation was mapped onto the accessible surface of the output domain. One protomer is shown in surface presentation, the other is shown in ribbon presentation. Conserved motifs and individual residues are highlighted.

(D) $LapD^{output}$-LapG complex formation. Purified, hexahistidine-tagged LapG (His_6-LapG) was bound to NiNTA, and incubated in the absence or presence of untagged, wild-type $LapD^{output}$, or a $LapD^{output}$ mutant in which W125 has been replaced with a glutamate. The Coomassie-stained gel shows eluates of NiNTA-bound proteins (data courtesy of Debashree Chatterjee).

(E) Biofilm phenotypes and LapD stability. Biofilm formation of Δ*lapD* cell expressing full-length, wild-type LapD, LapD point mutants or the insert-less expression vector was assessed. Protein levels are shown by Western blotting for the His_6-epitope at the C-terminus of LapD (Data courtesy of Peter D. Newell and George A. O'Toole).


A
Periplasm
Cytoplasm
TM1⇢N (chain B)
C⇢TM2 (chain A)
90°
molecule A
molecule B
B
β3
β4
β1
α2
β2
α1
α3
α4
TM1⇢N
C⇢TM2
TM2⇠C
N⇠TM1
C
GWxQ127
D59
W125
FD72
PxWF106
LW145
18 sequences (incl. distant homologs)
100%
0%
Conservation
D
M
Input
NiNTA pull-downs
His6-LapG
LapDoutput
His6-LapG
LapDoutput (wt)
LapDoutput (W125E)
E
Biofilm
αHis6-Blot
Biofilm formation (relative to wild-type LapD)
1.0
0.5
0.0
vector
wild-type
W125E
W125E/ S229D
pLapD

A sequence alignment of 18 sequences was constructed, including LapD homologs from other *Pseudomonas* strains but also extending to more distantly related sequences from other bacterial genera (Fig. 3.8; Table S2). Mapping sequence conservation onto the accessible molecular surface revealed a few potentially important patches (Fig. 3.6C and 3.15A). The PxWF and LW segments (residues 103-106 and 144-145 of LapD, respectively) form a continuous surface at the bottom of the dimer. While the LW segment is part of the surface that accommodates the long N-terminal helix of the adjacent protomer, the PxWF is likely to interact with the inner membrane. The other striking feature is a strictly conserved loop connecting the strands β3 and β4 formed by the conserved GWxQ motif (residues 124-127 of LapD). W^{125} forms the most distal point of the periplasmic domain located at the center of the loop, and its side chain is in an outward facing rotamer conformation (Fig. 3.6C).

Given its strict conservation and peculiar conformation, we targeted W^{125} in a site-directed mutagenesis, replacing its side chain non-conservatively with a glutamate residue. The mutant output domain expressed and purified indistinguishably from the wild-type protein but had distinct functional properties. In a purified system using hexahistidine-tagged LapG, a periplasmic cysteine protease that binds to LapD's output domain in a c-di-GMP-dependent manner (*27*), we could efficiently pull-down the untagged wild-type output domain (Fig. 3.6D). This result indicates that in the absence of the transmembrane and cytoplasmic domains, the output domain adopts a LapG binding-competent state. In contrast, the output domain mutant $W^{125}E$ failed to interact with LapG in this assay. Consistent with these results, a full-length allele harboring the $W^{125}E$ mutation failed to restore LapD-dependent biofilm formation in a Δ*lapD* genetic background (Fig. 3.6E). The periplasmic loss-of-function mutation is also dominant over the highly activating $S^{229}D$ mutation when introduced in the same allele, underlining the functional importance of W^{125} in

transmitting cytosolic signaling events to the periplasm.

Structure-based model for the regulation of periplasmic proteases in bacteria

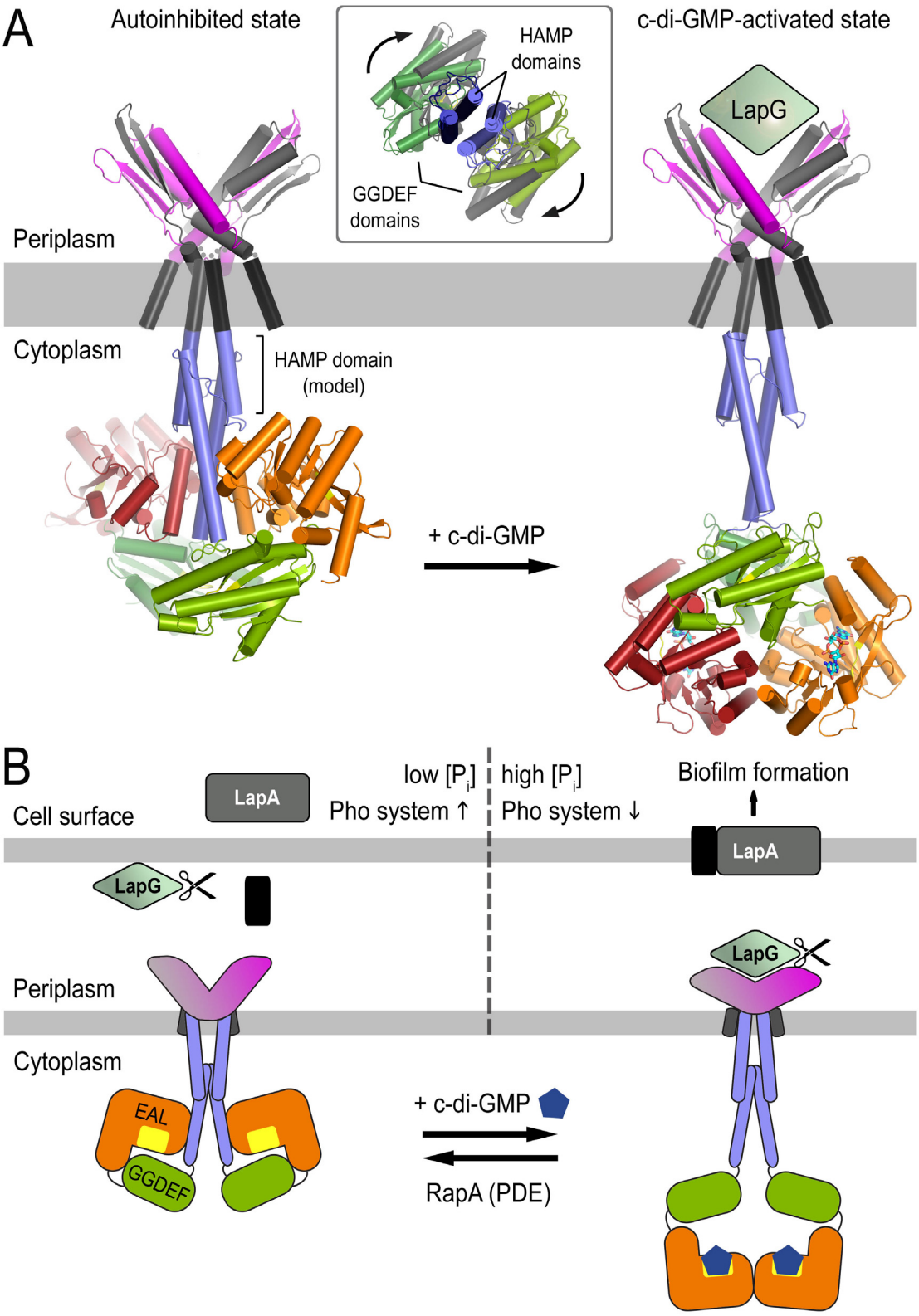
Our structural analyses of LapD revealed an autoinhibited conformation of the cytosolic domains in the absence of c-di-GMP, a dimeric state of c-di-GMP-bound EAL domains in the active state, and a domain-swapped dimer of the periplasmic output domain that is competent for LapG binding. The HAMP domain was modeled based on available structural information for this relay module, with the S helix forming a continuous extension of the HAMP domain's second helix (*46, 47*). In conjunction with the biochemical and genetic analyses described in an accompanying manuscript, we propose the following model for the activation of LapD and its mechanism of inside-out signaling across the inner bacterial membrane (Fig. 3.7): The S helix and GGDEF domain function as a physical lock gating access of c-di-GMP to the EAL domain. In this conformation, LapD's output domain is held in a LapG binding-incompetent state, and hence LapG gains access to and cleaves LapA, releasing this critical biofilm adhesin from the cell surface. An increase in the cellular c-di-GMP level will outcompete the inhibitory interactions in the cytoplasmic domains, likely accompanied by a large conformational change allowing EAL domain dimerization. Coupling between dimerization and c-di-GMP binding may further contribute to the efficiency of the activation switch, by preventing reversal to the autoinhibited state. Many mutations in the cytoplasmic module including the HAMP domain lead to aberrant, constitutive activation of LapD (Fig. 3.5) (*23*). These data suggest that intrinsic autoinhibitory interactions are indeed necessary to prevent the system from adopting a constitutively active conformation.

The HAMP domain is directly coupled to the GGDEF and EAL domains via the S helix. HAMP domains occur in a large number of predominantly transmembrane

sensor proteins that relay environmental signals to intracellular events (outside-in signaling). Rotation of the helices in HAMP dimers has been described as the main mechanism for signal transmission. It is conceivable that the release of the S helix from the EAL domain and associated events involving the GGDEF and EAL domains will trigger a rotation in the HAMP domain in a similar fashion, yielding a conformational change in the output domain and allowing it to sequester LapG.

What is the relevance for the unusual fold of LapD's output domain? Unlike CitA and related sensor proteins, which bind small molecules in the periplasm and relay this information to the inside of the cell, LapD sequesters a periplasmic protein upon receiving a cytosolic signal. We reason that a domain-swapped fold would respond more efficiently and precisely in coupling conformational changes in the cytosolic domains across the membrane, compared to canonical dimeric periplasmic domains. Given the functional importance and the particular position of W^{125}, we hypothesize that the output domain may act as a molecular ruler with the tryptophan residues forming the tips of a caliper. Varying the angle between the arms of the V-shaped fold upon c-di-GMP-triggered HAMP domain rotation could form the basis for modulating binding of LapG in the periplasm.

Figure 3.7: Structure-based model for LapD inhibition and activation. (A) Structural model of full-length LapD. We derived models for the autoinhibited and activated, c-di-GMP-bound state of LapD based on the crystal structures described here. Only the c-di-GMP-bound receptor is capable of LapG binding in the periplasm. The HAMP domains were modeled based on sequence-alignments and available structural information (*46, 47*). The inset shows the predicted HAMP (and GGDEF) domain rotation associated with the activation of LapD. **(B)** Model for LapD-mediated control of biofilm formation. The cartoon presents the current model for biofilm formation controlled by the c-di-GMP receptor LapD, based on our structural and functional analyses, previous results (*23, 25-27*).


A
Autoinhibited state
HAMP
domains
GGDEF
domains
c-di-GMP-activated state
LapG
Periplasm
Cytoplasm
HAMP domain
(model)
+ c-di-GMP
B
Cell surface
LapA
low [Pi]
Pho system ↑
high [Pi]
Pho system ↓
Biofilm formation
LapA
LapG
LapG
Periplasm
Cytoplasm
EAL
GGDEF
+ c-di-GMP
RapA (PDE)

Although competent for specific LapG binding, the isolated LapD output domain failed to compete for LapG sequestration with the full-length c-di-GMP bound receptor (Newell, unpublished data). It is likely that the intracellular and transmembrane domains facilitate the formation of a stable, high-affinity state. The observation that the isolated output domain can bind LapG is consistent with a model in which the nucleotide-free intracellular domains hold the receptor in an autoinhibited conformation that relaxes into a LapG-binding state upon activation. Consequently, removal of the regulatory domains would allow for the output domain to adopt the active, LapG-binding conformation. In addition, potential higher-order oligomerization of LapD into lattices may contribute to sequestering LapG over larger membrane surfaces and to the fine-tuning of the signaling system. Two crystal structures described here, of the output domain and the c-di-GMP-bound EAL domain, show some potentially relevant higher-order interactions (Fig. 3.14C and 3.14D). Further experiments will be required to determine the oligomeric state of full-length LapD in the absence and presence of c-di-GMP.

2.4. Conclusions

Here, we elucidated the molecular mechanism underlying *Pseudomonas fluorescens* LapD function and regulation. Based on sequence conservation, LapD homologs in other *Pseudomonas* strains, including *P. putida* and *P. aeruginosa*, are likely to function in a similar fashion (Fig. 3.8; Table S2) (*23, 24*). While LapD and LapG from *Pseudomonas aeruginosa* (PA1433 and PA1434, respectively) show a high degree of sequence conservation and functionally rescue deletions in these genes in P*seudomonas fluorescens*, no biofilm phenotype has been associated with this signaling system in their native strain (*22,* Newell- unpublished data), consistent with the absence of an obvious LapA homolog in this species. In contrast, we identified

similar effector systems and targets in more distant genera including *Legionella* and various *Vibrio* strains. In all these bacteria, *lapD* and *lapG* homologs with conserved, functionally important residues exist within the same operon (Fig. 3.8; Table S2). LapD from *Vibrio cholerae El Tor* represents a special case since its EAL domain is encoded by a second gene, separated from the transmembrane receptor containing the output, HAMP and GGDEF domains. While the relevance of this finding requires further investigation, these genes have been found upregulated in rugose strains of *Vibrio. cholerae*, associated with increased biofilm formation (48).

In most cases, the bioinformatic analysis also detected the presence of associated ABC transporters, as in the case of *Pseudomonas fluorescens* LapD. More variation is observed with regard to the putative substrates of the cysteine protease LapG. Newell *et al.* identified the large adhesin LapA as a LapG substrate, involved in biofilm formation and stability in *Pseudomonas fluorescens*. Based on the cleavage site sequence, other LapA homologs were identified in a variety of strains. In addition, we predict that LapG homologs may have different substrates in systems for which no clear LapA-type proteins could be identified. Proteins containing regions with homology to the LapG-cleavage site of LapA have been spotted in RTX-like bacterial toxins, and for the majority of such candidate substrates, these proteins are encoded in close genetic proximity to *lapD* and *lapG* homologs.

The GGDEF-EAL domain-containing proteins described here are degenerate with respect to their active sites, lack catalytic activity, and function as c-di-GMP receptors. A similar system has been previously described in *Escherichia coli*. Unlike LapD, the transmembrane HAMP-GGDEF-EAL domain-containing protein CsrD regulates degradation of regulatory RNAs, but we speculate that the cytosolic module may be autoregulated in a similar fashion (*49*). Other proteins containing the tandem domain module with higher degree of conservation at the putative enzyme active sites

exist in association with a HAMP domain in some bacterial genomes. The mechanism described for LapD may also be applicable to these systems, in which the HAMP domain and S-helix could be regulatory features to control the phosphodiesterase and/or diguanylate cyclase activity in outside-in signaling mechanism, thus leading to changes in cellular c-di-GMP levels.

REFERENCES

1. Hall-Stoodley, L., J. W. Costerton, et al. (2004). "Bacterial biofilms: from the natural environment to infectious diseases." *Nat Rev Microbiol* **2**(2): 95-108.
2. O'Toole, G., H. B. Kaplan, et al. (2000). "Biofilm formation as microbial development." *Annu Rev Microbiol* **54**: 49-79.
3. Parsek, M. R. and P. K. Singh (2003). "Bacterial biofilms: an emerging link to disease pathogenesis." *Annu Rev Microbiol* **57**: 677-701.
4. Hengge, R. (2009). "Principles of c-di-GMP signalling in bacteria." *Nat Rev Microbiol* **7**(4): 263-73.
5. Weinhouse, H., S. Sapir, et al. (1997). "c-di-GMP-binding protein, a new factor regulating cellulose synthesis in Acetobacter xylinum." *FEBS Lett* **416**(2): 207-11.
6. Schirmer, T. and U. Jenal (2009). "Structural and mechanistic determinants of c-di-GMP signalling." *Nat Rev Microbiol* **7**(10): 724-35.
7. D'Argenio, D. A. and S. I. Miller (2004). "Cyclic di-GMP as a bacterial second messenger." *Microbiology* **150**(Pt 8): 2497-502.
8. Ryan, R. P., Y. Fouhy, et al. (2006). "Cell-cell signaling in Xanthomonas campestris involves an HD-GYP domain protein that functions in cyclic di-GMP turnover." *Proc Natl Acad Sci U S A* **103**(17): 6712-7.
9. Simm, R., M. Morr, et al. (2004). "GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility." *Mol Microbiol* **53**(4): 1123-34.
10. Tal, R., H. C. Wong, et al. (1998). "Three cdg operons control cellular turnover of cyclic di-GMP in Acetobacter xylinum: genetic organization and occurrence of conserved domains in isoenzymes." *J Bacteriol* **180**(17): 4416-

25.

11. Amikam, D. and M. Y. Galperin (2006). "PilZ domain is part of the bacterial c-di-GMP binding protein." *Bioinformatics* **22**(1): 3-6.

12. Ryjenkov, D. A., R. Simm, et al. (2006). "The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria." *J Biol Chem* **281**(41): 30310-4.

13. Krasteva, P. V., J. C. Fong, et al. (2010). "Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP." *Science* **327**(5967): 866-8.

14. Hickman, J. W. and C. S. Harwood (2008). "Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor." *Mol Microbiol* **69**(2): 376-89.

15. Leduc, J. L. and G. P. Roberts (2009). "Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of Xanthomonas axonopodis pv. citri." *J Bacteriol* **191**(22): 7121-2.

16. Chan, C., R. Paul, et al. (2004). "Structural basis of activity and allosteric control of diguanylate cyclase." *Proc Natl Acad Sci U S A* **101**(49): 17084-9.

17. De, N., M. Pirruccello, P.V. Krasteva et al. (2008). "Phosphorylation-independent regulation of the diguanylate cyclase WspR." *PLoS Biol* **6**(3): e67.

18. Lee, V. T., J. M. Matewish, et al. (2007). "A cyclic-di-GMP receptor required for bacterial exopolysaccharide production." *Mol Microbiol* **65**(6): 1474-84.

19. Beyhan, S., L. S. Odell, et al. (2008). "Identification and characterization of cyclic diguanylate signaling systems controlling rugosity in Vibrio cholerae." *J Bacteriol* **190**(22): 7392-405.

20. Galperin, M. Y., A. N. Nikolskaya, et al. (2001). "Novel domains of the

prokaryotic two-component signal transduction systems." *FEMS Microbiol Lett* **203**(1): 11-21.

21. Christen, M., B. Christen, et al. (2005). "Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP." *J Biol Chem* **280**(35): 30829-37.

22. Kulasakara, H., V. Lee, et al. (2006). "Analysis of Pseudomonas aeruginosa diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence." *Proc Natl Acad Sci U S A* **103**(8): 2839-44.

23. Newell, P. D., R. D. Monds, et al. (2009). "LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by Pseudomonas fluorescens Pf0-1." *Proc Natl Acad Sci U S A* **106**(9): 3461-6.

24. Gjermansen, M., M. Nilsson, et al. (2009). "Characterization of starvation-induced dispersion in Pseudomonas putida biofilms: genetic elements and molecular mechanisms." *Mol Microbiol.*

25. Hinsa, S. M., M. Espinosa-Urgel, et al. (2003). "Transition from reversible to irreversible attachment during biofilm formation by Pseudomonas fluorescens WCS365 requires an ABC transporter and a large secreted protein." *Mol Microbiol* **49**(4): 905-18.

26. Hinsa, S. M. and G. A. O'Toole (2006). "Biofilm formation by Pseudomonas fluorescens WCS365: a role for LapD." *Microbiology* **152**(Pt 5): 1375-83.

27. Newell, P. D., C. D. Boyd, et al. “C-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage” (*submitted*).

28. Anantharaman, V., S. Balaji, et al. (2006). "The signaling helix: a common functional theme in diverse signaling proteins." *Biol Direct* **1**: 25.

29. Falke, J. J. and G. L. Hazelbauer (2001). "Transmembrane signaling in bacterial chemoreceptors." *Trends Biochem Sci* **26**(4): 257-65.

30. Stewart, V. and L. L. Chen (2010). "The S helix mediates signal transmission as a HAMP domain coiled-coil extension in the NarX nitrate sensor from Escherichia coli K-12." *J Bacteriol* **192**(3): 734-45.

31. Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." *J Mol Biol* **372**(3): 774-97.

32. Barends, T. R., E. Hartmann, et al. (2009). "Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase." *Nature* **459**(7249): 1015-8.

33. Rao, F., Y. Qi, et al. (2009). "The functional role of a conserved loop in EAL domain-based cyclic di-GMP-specific phosphodiesterase." *J Bacteriol* **191**(15): 4722-31.

34. De, N., M. V. Navarro, et al. (2009). "Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR." *J Mol Biol* **393**(3): 619-33.

35. Wassmann, P., C. Chan, et al. (2007). "Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition." *Structure* **15**(8): 915-27.

36. Minasov, G., S. Padavattan, et al. (2009). "Crystal structures of YkuI and its complex with second messenger cyclic Di-GMP suggest catalytic mechanism of phosphodiester bond cleavage by EAL domains." *J Biol Chem* **284**(19): 13174-84.

37. Navarro, M. V., N. De, et al. (2009). "Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX." *Structure* **17**(8): 1104-16.

38. De, N., M. V. Navarro, et al. (2010). "Biophysical assays for protein interactions in the Wsp sensory system and biofilm formation." *Methods in Enzymology* **471**: 161-184.

39. Lee, B. and F. M. Richards (1971). "The interpretation of protein structures: estimation of static accessibility." *J Mol Biol* **55**(3): 379-400.

40. Monds, R. D., P. D. Newell, et al. (2007). "Phosphate-dependent modulation of c-di-GMP levels regulates Pseudomonas fluorescens Pf0-1 biofilm formation by controlling secretion of the adhesin LapA." *Mol Microbiol* **63**(3): 656-79.

41. Holm, L., S. Kaariainen, et al. (2008). "Searching protein structure databases with DaliLite v.3." *Bioinformatics* **24**(23): 2780-1.

42. Reinelt, S., E. Hofmann, et al. (2003). "The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain." *J Biol Chem* **278**(40): 39189-96.

43. Cheung, J., C. A. Bingman, et al. (2008). "Crystal structure of a functional dimer of the PhoQ sensor domain." *J Biol Chem* **283**(20): 13762-70.

44. Chang, C., C. Tesar, et al. (2010). "Extracytoplasmic PAS-like domains are common in signal transduction proteins." *J Bacteriol* **192**(4): 1156-9.

45. Sevvana, M., V. Vijayan, et al. (2008). "A ligand-induced switch in the periplasmic domain of sensor histidine kinase CitA." *J Mol Biol* **377**(2): 512-23.

46. Airola, M. V., K. J. Watts, et al. "Structure of concatenated HAMP domains provides a mechanism for signal transduction." *Structure* **18**(4): 436-48.

47. Hulko, M., F. Berndt, et al. (2006). "The HAMP domain structure implies helix rotation in transmembrane signaling." *Cell* **126**(5): 929-40.

48. Beyhan, S., K. Bilecen, et al. (2007). "Regulation of rugosity and biofilm formation in Vibrio cholerae: comparison of VpsT and VpsR regulons and epistasis analysis of vpsT, vpsR, and hapR." *J Bacteriol* **189**(2): 388-402.

49. Suzuki, K., P. Babitzke, et al. (2006). "Identification of a novel regulatory

protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E." *Genes Dev* **20**(18): 2605-17.

APPENDIX

Supplemental Information

S3.1. Materials and methods

Protein expression and purification

For *in vivo* experiments, LapD protein variants were expressed from an arabinose-inducible vector (pMQ72) in a Δ*lapD* strain as previously described (*S1, S2*). Proteins used in crystallization and *in vitro* studies were expressed and purified as follows.

The coding regions corresponding to the GGDEF-EAL dual domain module ($LapD^{dual}$; residues 220-648) and to the isolated EAL domain ($LapD^{EAL}$; residues 399-648) of *Pseudomonas fluorescens* Pf0-1 LapD were amplified by standard PCR and cloned into a modified pProExHtb expression vector (Invitrogen), where the TEV protease cleavage site was engineered into a PreScission Protease site for removal of the N-terminally fused hexahistidine (His_6) moiety. The coding region corresponding to the periplasmic output domain of LapD ($LapD^{output}$; residues 22-151) was PCR-amplified and cloned into a modified pET28a expression plasmid (Novagen) yielding an N-terminally His_6-tagged SUMO fusion protein. The His_6-tagged SUMO moiety was cleavable using the yeast protease Ulp-1. Finally, LapG protein was expressed as a C-terminally His_6-tagged version after cloning and expression from a pET21a expression vector (Novagen).

Native and selenomethionine-derivatized proteins were overexpressed in *E. coli*, as previously described (*S3*). Briefly, native proteins were expressed in T7 Express cells (NEB), grown at 37˚C in Terrific Broth (TB) media supplemented with

50 μg/ml kanamycin for the pET28a vector, or 100 μg/ml ampicillin for the pProEx and pET21 vectors. At an optical density corresponding to 0.8-1.2 absorbance at 600 nm (OD_{600}), the temperature was reduced to 18˚C and protein expression was induced with 1mM IPTG. Selenomethionine-derivatized proteins were expressed in T7 Crystal Express cells (NEB), grown in M9 minimal media supplemented with the appropriate antibiotic, vitamins (1 μg/ml thiamin and 1 μg/ml biotin), carbon source (0.4% glucose), trace elements, and amino acids (40 μg/ml of each of the 20 amino acids with selenomethionine substituting for methionine). Protein expression in minimal medium was induced at cell densities corresponding to OD_{600} of 0.4-0.5. After 16 hours of expression at 18 degrees, cells were harvested by centrifugation, resuspended in NiNTA buffer A (25 mM Tris-HCl, pH 8.4, 500 mM NaCl, and 20 mM Imidazole), and flash-frozen in liquid nitrogen.

After thawing and cell lysis by sonication, cellular debris were removed by centrifugation and clear lysates were incubated with NiNTA resin (Qiagen) equilibrated in NiNTA buffer A. The resin was washed excessively with buffer A and proteins were eluted in a single step of NiNTA buffer A supplemented with 500 mM Imidazole. Proteins were buffer exchanged into desalting buffer (25 mM Tris-HCl, pH 8.4, 300 mM NaCl, and 5mM β-mercaptoethanol) and, where applicable, affinity tags were removed by incubation with the yeast protease Ulp-1 or PreScission Protease at 4˚C overnight. Cleaved proteins were collected in the flow-through during a second step of NiNTA affinity chromatography (HisTrap; GE Healthcare). As final step of protein purification, proteins were concentrated and subjected to size-exclusion chromatography on a Superdex200 column (GE Healthcare) equilibrated with gel filtration buffer (25 mM Tris-HCl, pH 8.4, and 250-300 mM NaCl). Proteins were concentrated on a Centricon ultracentrifugation device with an appropriate molecular weight cut-off (Millipore) to final concentrations in the low milimolar range. Protein

aliquots were flash frozen in liquid nitrogen and stored at -80˚C.

Crystallization, X-ray data collection, and structure solving

All crystals were obtained by hanging drop vapor diffusion after mixing equal volumes of protein (10-30 mg/ml) and reservoir solution. $LapD^{dual}$ was crystallized without cleavage of the N-terminally fused hexahistidine tag. Native and selenomethionine-derivatized proteins yielded single crystals grown at 20˚C with reservoir solution containing 14% PEG 4000 and 0.1 M MES, pH 6.0 (space group $P3_2$). In addition, $LapD^{dual}$ yielded crystals with I32 space group symmetry after mixing with reservoir solution of 0.2 M Ammonium acetate, 0.1 M Sodium citrate, pH 5.6, and 15% PEG 4000. For cryoprotection in either case, crystals were soaked in reservoir solution supplemented with 30% Ethylene glycol prior to freezing.

For crystallization of the isolated, untagged EAL domain in the presence of c-di-GMP, the protein solution was supplemented with 1mM purified nucleotide prior to setting up the crystallization trials. Diffraction-quality crystals with $P6_522$ space group symmetry grew after incubation for 7-10 days at 4˚C with well solution containing 0.1 M Bis-Tris, pH 6.5, and 1.5 M Ammonium sulfate. Cyclic di-GMP bound EAL domain crystals with $C222_1$ symmetry grew at 20˚C in a crystallization condition containing 0.1 M Bis-Tris, pH 5.5, 0.2 M Ammonium sulfate, and 24% PEG 3350. Prior to freezing, the crystals were soaked in their respective reservoir solutions supplemented with 25% of the cryoprotectant xylitol.

$LapD^{output}$ crystals used for data collection were grown at 4˚C after mixing with a reservoir solution containing 22% PEG monomethyl ether 2,000 and 0.15 M Potassium bromide. For crystal freezing, the mother liquor was supplemented with 20% xylitol to ensure cryoprotection.

Cryo-preserved crystals for all protein constructs were flash-frozen and stored

in liquid nitrogen. Data were collected on frozen crystals at 100K using synchrotron radiation at the Cornell High Energy Synchrotron Source (CHESS).

Data reduction was carried out with the software package HKL2000 (*S4*) and XDS (*S5*). Experimental phases for $LapD^{dual}$ and $LapD^{output}$ crystals were obtained from Single Anomalous Diffraction (SAD) experiments on crystals grown from selenomethionine-derivatized proteins. Heavy atom positions and solvent flattening was carried out by using the software package PHENIX (*S6*). For $LapD^{dual}$, initial phases were extended by using the software DM (*S7*), and the first model was built into the electron density map automatically by using the software Buccaneer (*S8*). The structure of nucleotide-bound $LapD^{EAL}$ was determined by molecular replacement in PHENIX with the EAL domain of $LapD^{dual}$ as the search model. Refinement in PHENIX and COOT yielded the final models (*S6, S9*). Data collection and refinement statistics are summarized in Table S1. Illustrations were made in Pymol (DeLano Scientific).

Size exclusion chromatography-coupled static multi-angle light scattering (MALS)

For MALS measurements, purified proteins (250 μM; injected concentration) were subjected to size exclusion chromatography (SEC) using a Shodex KW-803 column (JM Science) equilibrated in gel filtration buffer (25 mM Tris-HCl pH 8.4 and 250 mM NaCl). Where specified, wild-type or mutant LapD protein variants were incubated with c-di-GMP (500 μM), produced enzymatically (see below), for 30 minutes at room temperature prior to SEC. The SEC system was coupled to a three-angle light scattering detector and a refractive index detector (miniDAWN TREOS and Optilab rEX, respectively, Wyatt Technology). Data were collected every second at a flow rate of 0.5 ml/min and analyzed with the software ASTRA, yielding the molecular weight and mass distribution (polydispersity) of the samples. For data

quality control and normalization of the light scattering detectors, monomeric bovine serum albumin (Sigma) was used.

Enzymatic production of c-di-GMP

Cyclic di-GMP used in crystallization trials and c-di-GMP binding assays was synthesized enzymatically using a constitutively active WspR mutant (PA3702 $R^{242}A$) and GTP as a substrate (*S10*). Following purification by preparative reverse-phase HPLC and lyophilization, the nucleotide product was enzymatically tested as a substrate for phosphodiesterases (data not shown). Cyclic di-GMP purity and concentration were determined based on absorbance at 254 nm in comparison with commercially obtained standard (Biolog Life Science Institute).

Semi-quantitative c-di-GMP binding assay

Proteins (250 μM) were preincubated with excess c-di-GMP (500 μM) and separated from unbound nucleotide via SEC. SEC-eluted protein peaks were collected, concentrated to a final concentration of 200 μM to normalize for protein content, heat denatured and filtered through Microcon Centrigugal Filter Units (Millipore, 10 kDa cut-off). Nucleotide content in the resulting samples was analyzed on a C18 reverse-phase HPLC column by using a methanol-phosphate gradient (buffer A: 100mM monobasic potassium phosphate pH 6.0; buffer B: 70% buffer A, 30% methanol) (*S10*). Purified nucleotides were used for standardization. Integrated areas of the c-di-GMP peaks from three independent experiments were plotted as relative to those for the wild-type $LapD^{dual}$ and $LapD^{EAL}$ protein constructs, respectively.

Protein pull-down assay

Hexahistidine (His_6)-tagged LapG was incubated with NiNTA superflow resin (Qiagen) in low-salt binding buffer (25 mM Tris-HCl pH 8.4, 75 mM NaCl, 25 mM KCl, and 40 mM Imidazole). After removal of any unbound protein in consecutive wash steps, untagged LapD output domain variants were added to the reaction and incubated for 1 hour at 4ºC under nutation. The resin was extensively washed in low-salt binding buffer. The remaining affinity-bound proteins or protein complexes were eluted from the slurry in elution buffer (25 mM Tris-HCl pH 8.4, 500 mM NaCl, 300 mM Imidazole) and visualized using standard denaturing gel electrophoresis (SDS-PAGE).

Strains and growth conditions

Routine culturing of *P. fluorescens* Pf0-1 and *E. coli* was done in lysogeny broth (LB) at 30°C and 37°C, respectively. When appropriate, antibiotics were added to the medium at the following concentrations: *E. coli*- 10 μg/ml Gentamycin; *P. fluorescens*- 20 μg/ml Gentamycin. Plasmids were introduced into *P. fluorescens* by electroporation as described (*S11*). K10T media for biofilm assays was prepared as described (*S12*): K10T-π: 50 mM Tris-HCl pH 7.4, 0.2% (wt/vol) Bacto tryptone, 0.15% (vol/vol) glycerol, and 0.61 mM Mg_2SO_4. K10T-1 medium is K10T-π amended with 1 mM K_2HPO_4. A list of strains and plasmids used in the cell-based assays is provided as Supplemental Table 3.

Construction of LapD variants

For *in vitro* studies, *ladD* point mutants were generated using QuikChange site directed mutagenesis kit (Stratagene) following the manufacturer's instructions. For cell-based studies, *lapD* alleles were generated by recombination cloning (*S2*). Briefly, *lapD* was amplified in two pieces, the 5' fragment ending just upstream of the

codon to be changed and the 3' fragment beginning just downstream. The new codon and 21 complimentary bases were included in each internal primer such that recombination in yeast with the parent pMQ72 vector would yield a contiguous *lapD* ORF with the desired codon change.

Quantitative biofilm formation and surface attachment assays

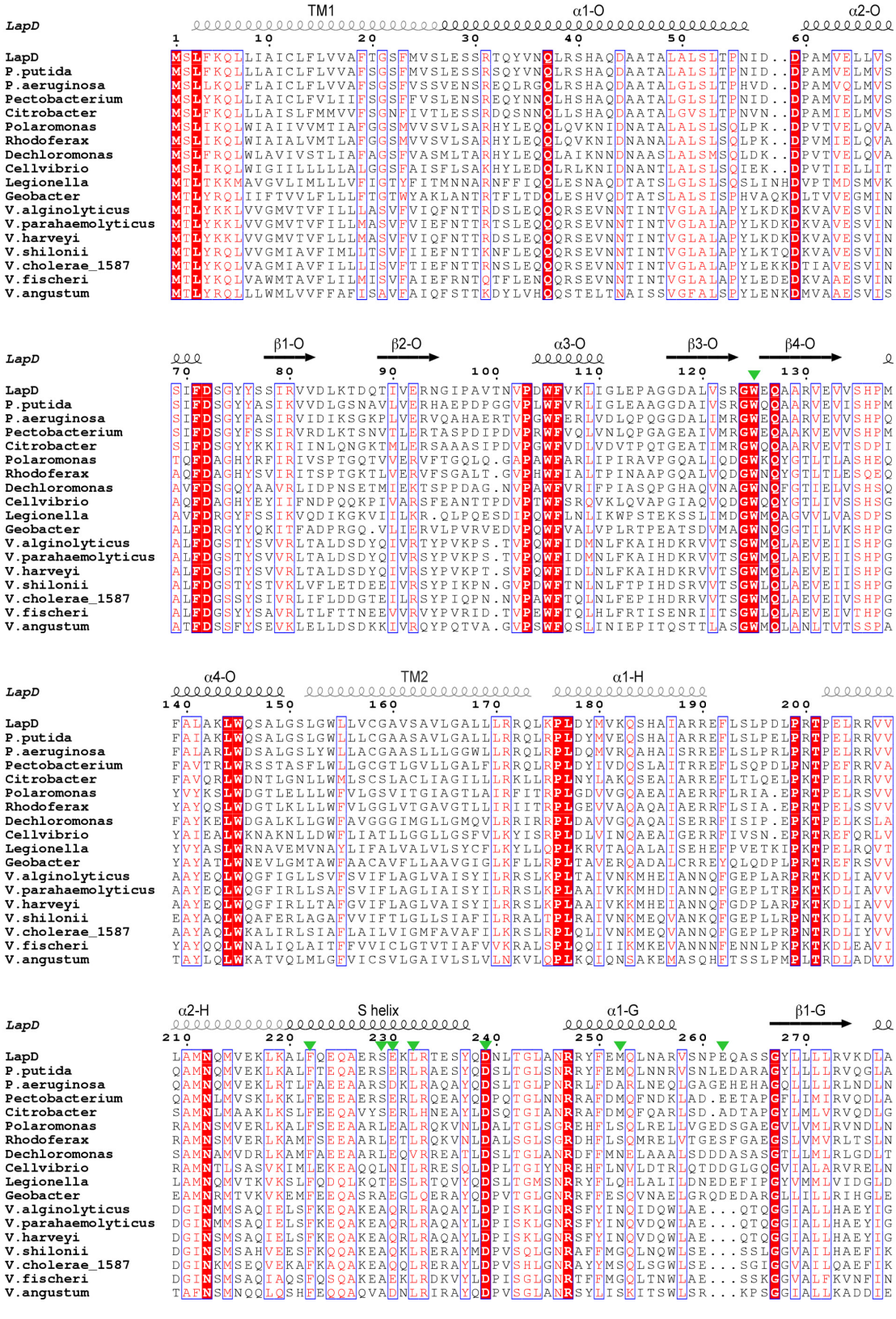
To quantify biofilm formation, strains were grown statically for 6 hours in K10T-1 medium as described previously (*S1*). Biofilm biomass was stained with 0.1% crystal violet for 15 min, and stain dissolved and quantified by spectrophotometry, measuring O.D. at 550nm. We analyzed the effects of P_i starvation on attachment by comparing biofilm levels in high P_i (K10T-1) and low P_i (K10T-π) media over time, as done previously (*S1*).

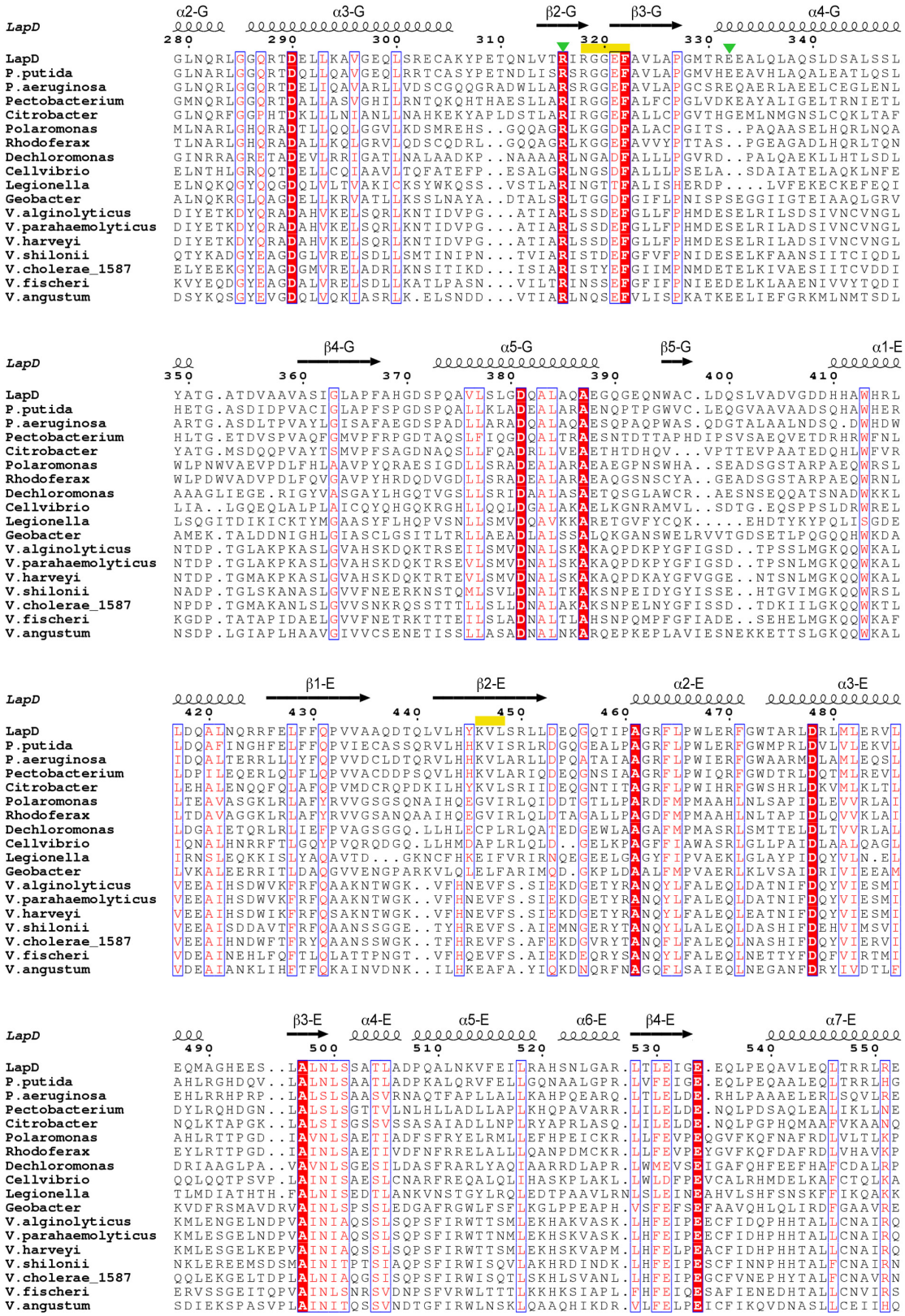
Assessment of LapD protein levels by Western blot

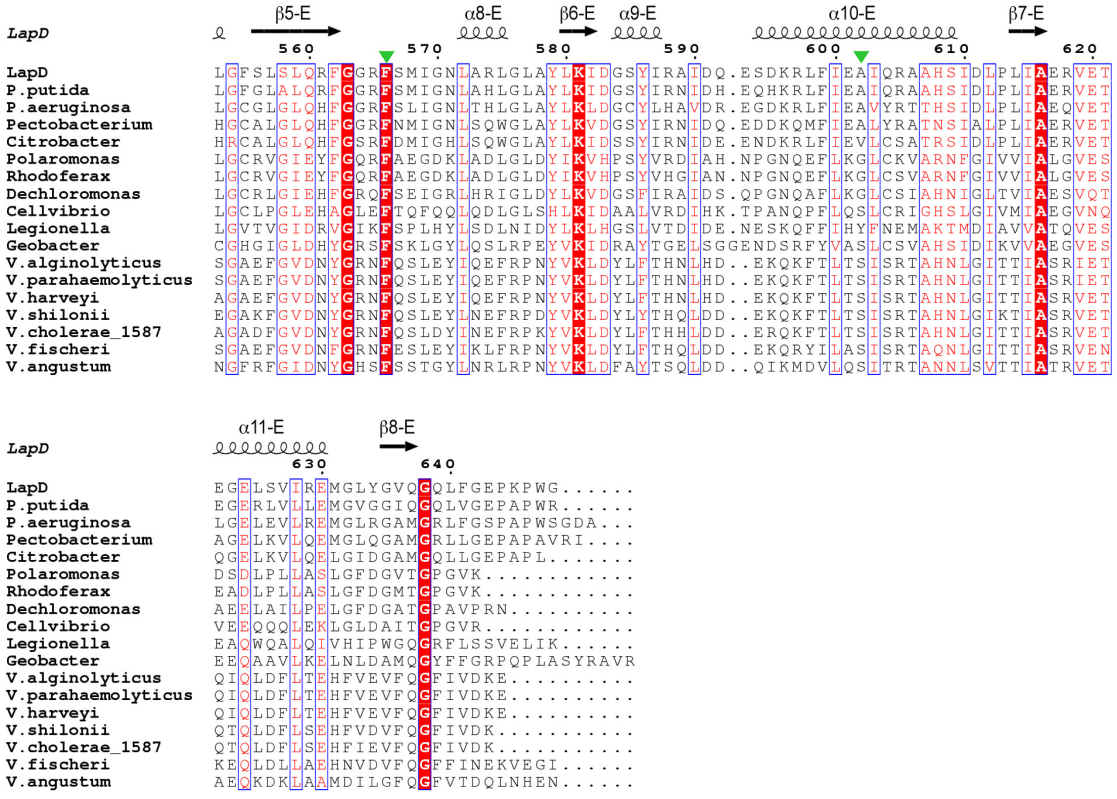
LapD proteins expressed in *P. fluorescens* Pf0-1 were visualized by Western blot as described previously (*S1*), with the following modifications. Blots were probed for the His_6-epitope with a rabbit anti-His_6 antibody (Genscript). Samples consisted of clarified cell lysates prepared by harvesting cells from 3 ml of overnight culture, sonicating 3 x 10s in 500 µl of buffer (20 mM Tris pH 8, 10 mM $MgCl_2$), and pelleting debris at 15,000 x g for 12 min. Samples were normalized to protein concentration using the BCA kit (Pierce).

S3.2. Supplemental Figures

Figure 3.8: Sequence alignment of LapD homologs.
A sequence alignment of LapD homologs from various species was generated with ClustalW2 (*S13*) and formatted with ESPript (*S14*). Key residues discussed in the manuscript are marked with closed, green arrows. The degenerate GGDEF and EAL signature motifs (RGGEF and KVL, respectively) are marked with yellow bars. Secondary structure elements are shown based on the crystallographic data and secondary structure predictions for the transmembrane and HAMP domains. The following sequences were used to generate the alignment: *Pseudomonas fluorescens* Pf0-1 (LapD, YP_345864), *Pseudomonas putida* KT2440 (NP_742334), *Pseudomonas aeruginosa* PA01 (NP_250124), *Pectobacterium carotovorum* subsp. *brasiliensis* PBR1692 (ZP_03826388), *Citrobacter sp.* ATCC 29220 (ZP_06355256), *Polaromonas sp.* JS666 (YP_547171), *Rhodoferax ferrireducens* T118 (YP_524995), *Dechloromonas aromatica* RCB (YP_286553), *Cellvibrio japonicus* Ueda107 (YP_001981887), *Legionella pneumophila* str. *Lens* (YP_126219), *Geobacter sp.* M18 (ZP_05313414), *Vibrio alginolyticus* 12G01 (ZP_01258281), *Vibrio parahaemolyticus* AQ3810 (ZP_01990882), *Vibrio harveyi* HY01 (ZP_01986262), *Vibrio shilonii* AK1 (ZP_01866121), *Vibrio cholerae* 1587 (ZP_01950486), *Vibrio fischeri* ES114 (YP_207124), *Vibrio angustum* S14 (ZP_01233947).


LapD
TM1
α1-O
α2-O
β1-O
β2-O
α3-O
β3-O
β4-O
α4-O
TM2
α1-H
α2-H
S helix
α1-G
β1-G
LapD
P.putida
P.aeruginosa
Pectobacterium
Citrobacter
Polaromonas
Rhodoferax
Dechloromonas
Cellvibrio
Legionella
Geobacter
V.alginolyticus
V.parahaemolyticus
V.harveyi
V.shilonii
V.cholerae_1587
V.fischeri
V.angustum


α2-G α3-G β2-G β3-G α4-G
LapD
280 290 300 310 320 330 340
LapD GLNQRLGGQRTDELLKAVGEQLSRECAKYPETQNLVTRIRGGEFAVLAPGMTREEALQLAQSLDSALSSL
P.putida GLNARLGGQRTDQLLQAVGEQLRRTCASYPETNDLISRSRGGEFAVLAPGMVHEEAVHLAQALEATLQSL
P.aeruginosa GLNQRLGGQRTDELIQAVARLLVDSCGQQGRADWLLARSRGGEFAVLAPGCSREQAERLAEELCEGLENL
Pectobacterium GMNQRLGGQRTDALLASVGHILRNTQKQHTHAESLLARIRGGEFALFCPGLVDKEAYALIGELTRNIETL
Citrobacter GLNQRFGGPHTDKLLLNIANLLNAHKEKYAPLDSTLARIRGGEFALLCPGVTHGEMLNMGNSLCQKLTAF
Polaromonas MLNARLGHQRADTLLQQLGGVLKDSMREHS..GQQAGRLKGGDFALACPGITS..PAQAASELHQRLNQA
Rhodoferax TLNARLGHQRADALLKQLGRVLQDSCQDRL..GQQAGRLKGGEFAVVYPTTAS..PGEAGADLHQRLTQN
Dechloromonas GINRRAGRETADEVLRRIGATLNALAADKP..NAAAARLNGADFALLLPGVRD..PALQAEKLLHTLSDL
Cellvibrio ELNTHLGRQQTDELLCQIAAVLTQFATEFP..ESALGRLNGSDFALLLPSELA..SDAIATELAQKLNFE
Legionella ELNQKQGYQQGDQLVLTVAKICKSYWKQSS..VSTLARINGTTFALISHERDP....LVFEKECKEFEQI
Geobacter ALNQKRGLQAGDELLKRVATLLKSSLNAYA..DTALSRLTGGDFGIFLPNISPSEGGIIGTEIAAQLGRV
V.alginolyticus DIYETKDYQRADAHVKELSQRLKNTIDVPG...ATIARLSSDEFGLLFPHMDESELRILSDSIVNCVNGL
V.parahaemolyticus DIYETKDYQRADAHVKELSQRLKNTIDVPG...ATIARLSSDEFGLLFPHMDESELRILADSIVNCVNGL
V.harveyi DIYETKDYQRADAHVKELSQRLKNTIDVPG...ATIARLSSDEFGLLFPHMDESELRILADSIVNCVNGL
V.shilonii QTYKADGYEAGDGLVRELSDLLSMTINIPN...TVIARISTDEFGFVFPNIDESELKFAANSIITCIQDL
V.cholerae_1587 ELYEEKGYEAGDGMVRELADRLKNSITIKD...ISIARISTYEFGIIMPNMDETELKIVAESIITCVDDI
V.fischeri KVYEQDGYEAGDALVRELSDLLKATLPASN...VILTRINSSEFGFIFPNIEEDELKLAAENIVYYTQDI
V.angustum DSYKQSGYEVGDQLVQKIASRLLK.ELSNDD...VTIARLNQSEFVLISPKATKEELIEFGRKMLNMTSDL
β4-G α5-G β5-G α1-E
LapD
350 360 370 380 390 400 410
LapD YATG.ATDVAAVASIGLAPFAHGDSPQAVLSLGDQAIAQAEGQGEQNWAC.LDQSLVADVGDDHHAWHRL
P.putida HETG.ASDIDPVACIGLAPFSPGDSPQALLKLADEALARAENQPTPGWVC.LEQGVAAVAADSQHAWHER
P.aeruginosa ARTG.ASDLTPVAYLGISAFAEGDSPADLLARADQALAQAESQPAQPWAS.QDGTALAALNDSQ.DWHDW
Pectobacterium HLTG.ETDVSPVAQFGMVPFRPGDTAQSLFIQGDQALTRAESNTDTTAPHDIPSVSAEQVETDRHRWFNL
Citrobacter YATG.MSDQQPVAYTSMVPFSAGDNAQSLLFQADRLLVEAETHTDHQV...VPTTEVPAATEDQHLWFVR
Polaromonas WLPNWVAEVPDLFHLAAVPYQRAESIGDLLSRADEALARAEAEGPNSWHA..SEADSGSTARPAEQWRSL
Rhodoferax WLPDWVADVPDLFQVGAVPYHRDQDVGDLLSRADEALARAEAQGSNSCYA..GEADSGSTARPAEQWRNL
Dechloromonas AAAGLIEGE.RIGYVASGAYLHGQTVGSLLSRIDAALASAETQSGLAWCR..AESNSEQQATSNADWKKL
Cellvibrio LIA..LGQEQLALPLAICQYQHGQKRGHLLQQLDGALAKAELKGNRAMVL..SDTG.EQSPPSLDRWREL
Legionella LSQGITDIKICKTYMGAASYFLHQPVSNLLSMVDQAVKKARETGVFYCQK.....EHDTYKYPQLISGDE
Geobacter AMEK.TALDDNIGHLGIASCLGSITLTRLLAEADLALSSALQKGANSWELRVVTGDSETLPQGQQHWKDA
V.alginolyticus NTDP.TGLAKPKASLGVAHSKDQKTRSEILSMVDNALSKAKAQPDKPYGFIGSD..TPSSLMGKQQWKAL
V.parahaemolyticus NTDP.TGLAKPKASLGVAHSKDQKTRSEVLSMVDNALSKAKAQPDKPYGFIGSD..TPSNLMGKQQWKAL
V.harveyi NTDP.TGMAKPKASLGVAHSKDQKTRTEVLSMVDNALSKAKAQPDKAYGFVGGE..NTSNLMGKQQWKAL
V.shilonii NADP.TGLSKANASLGVVFNEERKNSTQMLSVLDNALTKAKSNPEIDYGYISSE..HTGVIMGKQQWRSL
V.cholerae_1587 NPDP.TGMAKANLSLGVVSNKRQSSTTTLLSLLDNALAKAKSNPELNYGFISSD..TDKIILGKQQWKTL
V.fischeri KGDP.TATAPIDAELGVVFNETRKTTTEILSLADNALTLAHSNPQMPFGFIADE..SEHELMGKQQWKAF
V.angustum NSDP.LGIAPLHAAVGIVVCSENETISSLLASADNALNKARQEPKEPLAVIESNEKKETTSLGKQQWKAL
β1-E β2-E α2-E α3-E
LapD
420 430 440 450 460 470 480
LapD LDQALNQRRFELFFQPVVAAQDTQLVLHYKVLSRLLDEQGQTIPAGRFLPWLERFGWTARLDRLMLERVL
P.putida LDQAFINGHFELFFQPVIECASSQRVLHHKVISRLRDGQGEALPAGRFLPWLERFGWMPRLDVLVLEKVL
P.aeruginosa IDQALTERRLLLYFQPVVDCLDTQRVLHHKVLARLLDPQATAIAAGRFLPWIERFGWAARMDLAMLEQSL
Pectobacterium LDPILEQERLQLFLQPVVACDDPSQVLHHKVLARIQDEQGNSIAAGRFLPWIQRFGWDTRLDQTMLREVL
Citrobacter LEHALENQQFQLAFQPVMDCRQPDKILHYKVLSRIIDEQGNTITAGRFLPWIHRFGWSHRLDKVMLKLTL
Polaromonas LTEAVASGKLRLAFYRVVGSGSQNAIHQEGVIRLQIDDTGTLLPARDFMPMAAHLNLSAPIDLEVVRLAI
Rhodoferax LTDAVAGGKLRLAFYRVVGSANQAAIHQEGVIRLQLDLTAGALLPAGDFMPMAAHLNLTAPIDLQVVKLAI
Dechloromonas LDGAIETQRLRLIEFPVAGSGGQ.LLHLECPLRLQATEDGEWLAAGAFMPMASRLSMTTELDLTVVRLAL
Cellvibrio IQNALHNRRFTLGQYPVQRQDGQ.LLHMDAPLRLQLD..GELKPAGFFIAWASRLGLLPAIDLAALQAGL
Legionella IRNSLEQKKISLYAQAVTD...GKNCFHKEIFVRIRNQEGEELGAGYFIPVAEKLGLAYPIDQYVLN.EL
Geobacter LVKALEERRITLDAQGVVENGPARKVLQLELFARIMQD.GKPLDAALFMPVAERLKLVSAIDRIVIEEAM
V.alginolyticus VEEAIHSDWVKFRFQAAKNTWGK..VFHNEVFS.SIEKDGETYRANQYLFALEQLDATNIFDQYVIESMI
V.parahaemolyticus VEEAIHSDWVKFRFQAAKNTWGK..VFHNEVFS.SIEKDGETYRANQYLFALEQLDATNIFDQYVIESMI
V.harveyi VEEAIHSDWIKFRFQSAKNTWGK..VFHNEVFS.SIEKDGETYRANQYLFALEQLEATNIFDQYVIESMI
V.shilonii VEEAISDDAVTFRFQAANSSGGE..TYHREVFS.AIEMNGERYTANQYLLALEQLDASHIFDEHVIMSVI
V.cholerae_1587 VEEAIHNDWFTFRYQAANSSWGK..TFHREVFS.AFEKDGVRYTANQFLFALEQLNASHIFDQYVIERVI
V.fischeri VDEAINEHLFQFTLQLATTPNGT..VFHQEVFS.AIEKDEQRYSANQYLFALEQLNETTYFDQFVIRTMI
V.angustum VDEAIANKLIHFTFQKAINVDNK..ILHKEAFA.YIQKDNQRFNAGQFLSAIEQLNEGANFDRYIVDTLF
β3-E α4-E α5-E α6-E β4-E α7-E
LapD
490 500 510 520 530 540 550
LapD EQMAGHEES..LALNLSSATLADPQALNKVFEILRAHSNLGAR.LTLEIGE.EQLPEQAVLEQLTRRLRE
P.putida AHLRGHDQV..LALNLSAATLADPKALQRVFELLGQNAALGPR.LVFEIGE.EQLPEQAALEQLTRRLHG
P.aeruginosa EHLRRHPRP..LALSLSAASVRNAQTFAPLLALLKAHPQEARQ.LTLELDE.RHLPAAAELERLSQVLRE
Pectobacterium DYLRQHDGN..LALSLSGTTVLNLHLLADLLAPLKHQPAVARR.LILELDE.NQLPDSAQLEALIKLLNE
Citrobacter NQLKTAPGK..LALSISGSSVSSASAIADLLNPLRYAPRLASQ.LILELDE.NQLPGPHQMAAFVKAANQ
Polaromonas AHLRTTPGD..IAVNLSAETIADFSFRYELRMLLEFHPEICKR.LLFEVPEQGVFKQFNAFRDLVLTLKP
Rhodoferax EYLRTTPGD..IAINLSAETIVDFNFRRELALLLQANPDMCKR.LLFEVPEYGVFKQFDAFRDLVHAVKP
Dechloromonas DRIAAGLPA..VAVNLSGESILDASFRARLYAQIAARRDLAPR.LWMEVSEIGAFQHFEEFHAFCDALRP
Cellvibrio QQLQQTPSVP.LAINISAESLCNARFREQALQLIHASKPLAKL.LWLDFPEVCALRHMDELKAFCTQLKA
Legionella TLMDIATHTH.FALNISEDTLANKVNSTGYLRQLEDTPAAVLRNLSLEINEAHVLSHFSNSKFFIKQAKK
Geobacter KVDFRSMAVDRVALNLSPSSLEDGAFRGWLFSFLKGLPPEAPH.VSFEFSEFAAVQHLQLIRDFGAAVRE
V.alginolyticus KMLENGELNDPVAINIAQSSLSQPSFIRWTTSMLEKHAKVASK.LHFEIPEECFIDQPHHTALLCNAIRQ
V.parahaemolyticus KMLESGELNDPVAINIAQSSLSQPSFIRWTTNMLEKHSKVASK.LHFEIPEECFIDHPHHTALLCNAIRQ
V.harveyi KMLESGELKEPVAINIAQSSLSQPSFIRWTTSMLEKHSKVAPM.LHFELPEACFIDHPHHTALLCNAIRQ
V.shilonii NKLEREEMSDSMAINITPTSIAQPSFIRWISQVLAKHRDINDK.LHFEIPESCFINNPHHTALFCNAIRN
V.cholerae_1587 QQLEKGELTDPLALNIAQGSISQPSFIRWISQTLSKHLSVANL.LHFEIPEGCFVNEPHYTALFCNAVRN
V.fischeri ERVSSGEITQPVAINLSNRSVDNPSFVRWLTTTLKKHKSIAPL.FHFEIQESAFIENEDHTALLCNVIRS
V.angustum SDIEKSPASVPLAINITQSSVNDTGFIRWLNSKLQAAQHIKDR.VLFELPEICFIKQVDNASLLCEIIHQ


LapD
β5-E
α8-E
β6-E
α9-E
α10-E
β7-E
560
570
580
590
600
610
620
LapD LGFSLSLQRFGGRFSMIGNLARLGLAYLKIDGSYIRAIDQ.ESDKRLFIEAIQRAAHSIDLPLIAERVET
P.putida LGFGLALQRFGGRFSMIGNLAHLGLAYLKIDGSYIRNIDH.EQHKRLFIEAIQRAAHSIDLPLIAERVET
P.aeruginosa LGCGLGLQHFGGRFSLIGNLTHLGLAYLKLDGCYLHAVDR.EGDKRLFIEAVYRTTHSIDLPLIAEQVET
Pectobacterium HGCALGLQHFGGRFNMIGNLSQWGLAYLKVDGSYIRNIDQ.EDDKQMFIEALYRATNSIALPLIAERVET
Citrobacter HRCALGLQHFGSRFDMIGHLSQWGLAYLKIDSSYIRNIDE.ENDKRLFIEVLCSATRSIDLPLIAERVET
Polaromonas LGCRVGIEYFGQRFAEGDKLADLGLDYIKVHPSYVRDIAH.NPGNQEFLKGLCKVARNFGIVVIALGVES
Rhodoferax LGCRVGIEYFGQRFAEGDKLADLGLDYIKVHPSYVHGIAN.NPGNQEFLKGLCSVARNFGIVVIALGVES
Dechloromonas LGCRLGIEHFGRQFSEIGRLHRIGLDYLKVDGSFIRAIDS.QPGNQAFLKGLCSIAHNIGLTVIAESVQT
Cellvibrio LGCLPGLEHAGLETTQFQQLQDLGLSHIKIDAALVRDIHK.TPANQPFLQSLCRIGHSLGIVMIAEGVNQ
Legionella LGVTVGIDRVGIKFSPLHYLSDLNIDYIKLHGSLVTDIDE.NESKQFFIHYFNEMAKTMDIAVVATQVES
Geobacter CGHGIGLDHYGRSFSKLGYLQSLRPEYVKIDRAYTGELSGGENDSRFYVASLCSVAHSIDIKVVAEGVES
V.alginolyticus SGAEFGVDNYGRNFQSLEYIQEFRPNYVKLDYLFTHNLHD..EKQKFTLTSISRTAHNLGITTIASRIET
V.parahaemolyticus SGAEFGVDNYGRNFQSLEYIQEFRPNYVKLDYLFTHNLHD..EKQKFTLTSISRTAHNLGITTIASRIET
V.harveyi AGAEFGVDNYGRNFQSLEYIQEFRPNYVKLDYLFTHNLHD..EKQKFTLTSISRTAHNLGITTIASRVET
V.shilonii EGAKFGVDNYGRNFQSLEYLNEFRPDYVKLDYLYTHQLDD..EKQKFTLTSISRTAHNLGIKTIASRVET
V.cholerae_1587 AGADFGVDNYGRNFQSLDYINEFRPKYVKLDYLFTHHLDD..ERQKFTLTSISRTAHNLGITTIASRVET
V.fischeri SGAEFGVDNFGRNFESLEYIKLFRPNYVKLDYLFTHQLDD..EKQRYILASISRTAQNLGITTIASRVEN
V.angustum NGFRFGIDNYGHSFSSTGYLNRLRPNYVKLDFAYTSQLDD..QIKMDVLQSITRTANNLSVTTIATRVET
α11-E
β8-E
630
640
LapD EGELSVIREMGLYGVQGQLFGEPKPWG.....
P.putida EGERLVLLEMGVGGIQGQLVGEPAPWR.....
P.aeruginosa LGELEVLREMGLRGAMGRLFGSPAPWSGDA...
Pectobacterium AGELKVLQEMGLQGAMGRLLGEPAPAVRI....
Citrobacter QGELKVLQELGIDGAMGQLLGEPAPL.....
Polaromonas DSDLPLLASLGFDGVTGPGVK.....
Rhodoferax EADLPLLASLGFDGMTGPGVK.....
Dechloromonas AEELAILPELGFDGATGPAVPRN.....
Cellvibrio VEEQQQLEKLGLDAITGPGVR.....
Legionella EAQWQALQIVHIPWGQGRFLSSVELIK.....
Geobacter EEQAAVLKELNLDAMQGYFFGRPQPLASYRAVR
V.alginolyticus QIQLDFLTEHFVEVFQGFIVDKE.....
V.parahaemolyticus QIQLDFLTEHFVEVFQGFIVDKE.....
V.harveyi QIQLDFLTEHFVEVFQGFIVDKE.....
V.shilonii QTQLDFLSEHFVDVFQGFIVDK.....
V.cholerae_1587 QTQLDFLSEHFIEVFQGFIVDK.....
V.fischeri KEQLDLLAEHNVDVFQGFFINEKVEGI.....
V.angustum AEQKDKLAAMDILGFQGFVTDQLNHEN.....

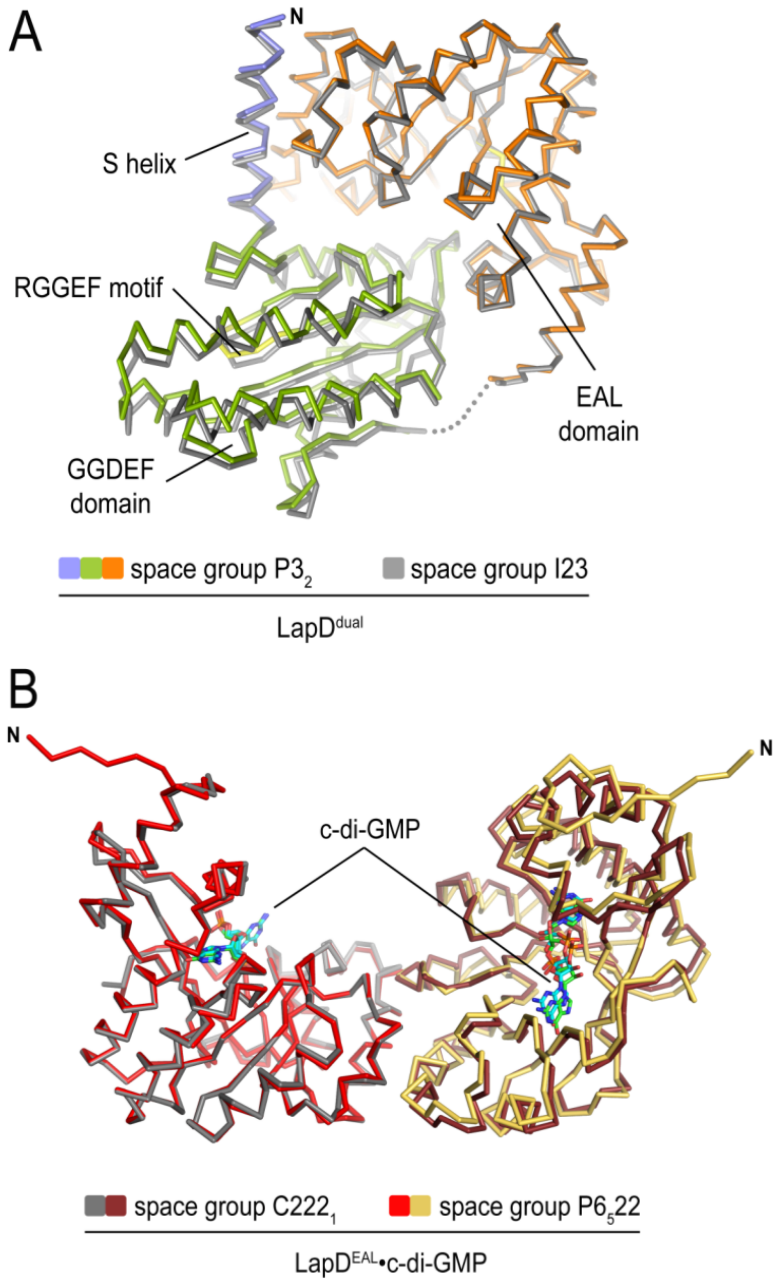


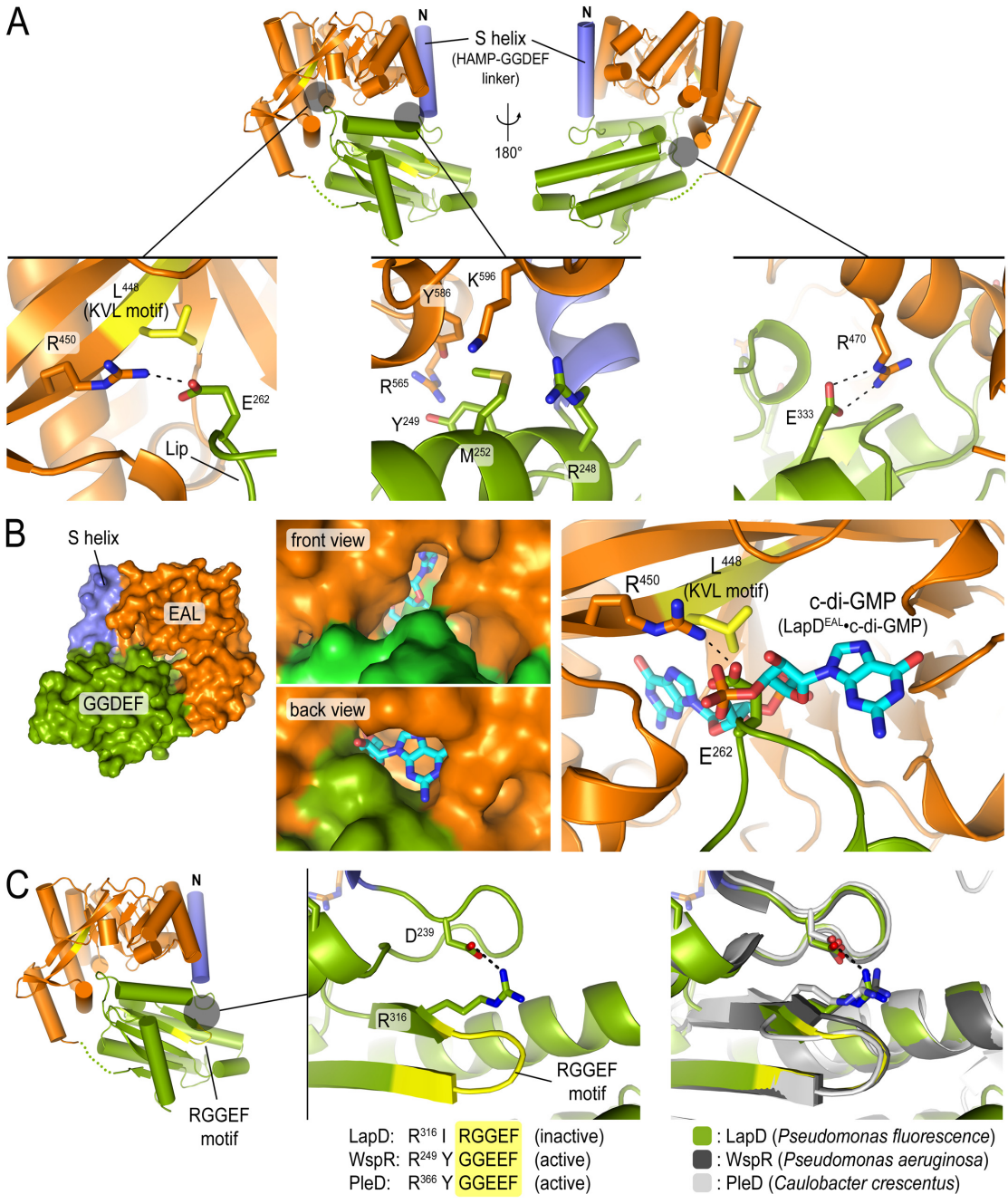
Figure 3.9: Crystal forms of LapD^{dual} and LapD^{EAL}•c-di-GMP.
(A) LapD^{dual}. Two independent crystal forms were obtained for LapD^{dual}. The resulting structures were superimposed on the EAL domain and shown as protein backbone traces (Data courtesy of Marcos V.A.S. Navarro).
(B) Cyclic di-GMP-bound LapD^{EAL}. Two independent crystal forms were obtained for LapD^{EAL}. Both crystal lattices show the same dimeric assembly of EAL domains. Dimers were superimposed on one EAL domain and shown as protein backbone traces.

Figure 3.10: GGDEF-EAL domain interactions and S helix-GGDEF domain linker conformation observed in apo-LapD^{dual} (Data courtesy of Marcos V.A.S. Navarro).

(A) GGDEF-EAL domain interaction. Close-up views are shown for regions of direct contact between the GGDEF and EAL domains in the autoinhibited structure of LapD^{dual}. The GGDEF and EAL domain is colored in green and orange, respectively. The S helix is colored in blue.

(B) Nucleotide binding pocket in apo-LapD^{dual}. A close-up view of the c-di-GMP binding pocket of LapD is shown (right panel). Cyclic di-GMP is shown in stick presentation after superimposing the crystal structure of LapD^{EAL}•c-di-GMP onto the EAL domain of apo-LapD^{dual}. The interacting residue pair R^{450}/E^{262} in LapD is incompatible with c-di-GMP binding. The left panels show surface presentations of apo-LapD^{dual}. The middle panel shows accessibility of the c-di-GMP binding site, with c-di-GMP taken from LapD^{EAL}•c-di-GMP after superimposition.

(C) S helix-GGDEF connector. The S helix and the GGDEF domain are connected via a short loop that forms a tight turn. The loop conformation is conserved in other GGDEF domain-containing proteins, and is stabilized by the interaction between two residues, D^{239} and R^{316}, which are strictly conserved in many GGDEF domain-containing proteins (*S10, S15-S17*). The arginine residue is directly preceding the GGDEF domain signature motif (GGDEF or GGEEF in active cyclases; RGGEF in LapD), the aspartate residue is located at the N-terminus of the loop. Its strict sequence and conformational conservation suggest a functional importance of the connector loop, likely restricting the conformational freedom between adjacent domains.


A
N
S helix
(HAMP-GGDEF
linker)
180°
N
L448
(KVL motif)
R450
E262
Lip
Y586
K596
R565
Y249
M252
R248
R470
E333
B
S helix
EAL
GGDEF
front view
back view
R450
L448
(KVL motif)
c-di-GMP
(LapDEAL•c-di-GMP)
E262
C
N
RGGEF
motif
D239
R316
RGGEF
motif
LapD: R316 I RGGEF (inactive)
WspR: R249 Y GGEEF (active)
PleD: R366 Y GGEEF (active)
: LapD (Pseudomonas fluorescence)
: WspR (Pseudomonas aeruginosa)
: PleD (Caulobacter crescentus)

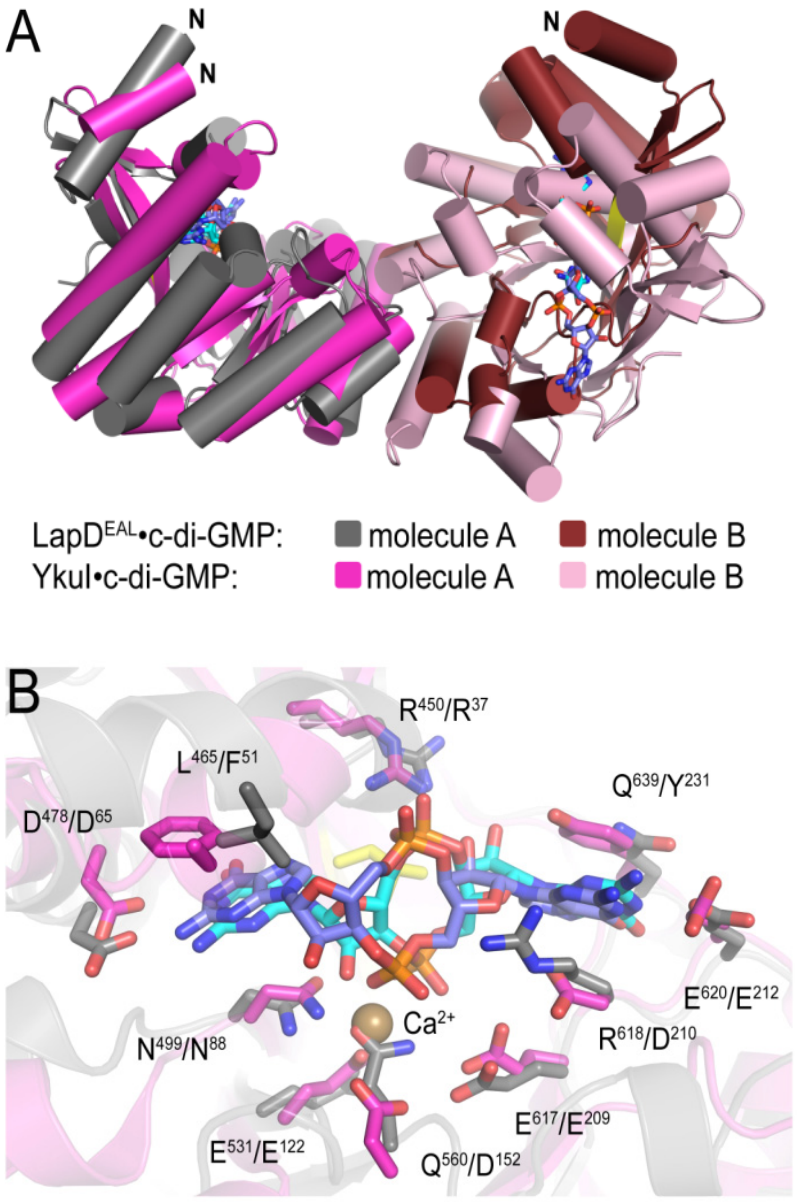


Figure 3.11: Comparison of c-di-GMP-bound LapD^{EAL} and YkuI dimers.
(A) Overview. Structures of EAL domain dimers of LapD and YkuI bound to c-di-GMP are shown in ribbon presentation (*S18*). Cyclic di-GMP is shown in stick presentation. Structures were superimposed on one of the EAL domain of the dimeric assemblies.
(B) Cyclic di-GMP binding site. A close-up view of the nucleotide binding pocket is shown. Residues involved in c-di-GMP (and, in the case of YkuI, divalent cation) coordination are shown in stick presentation.

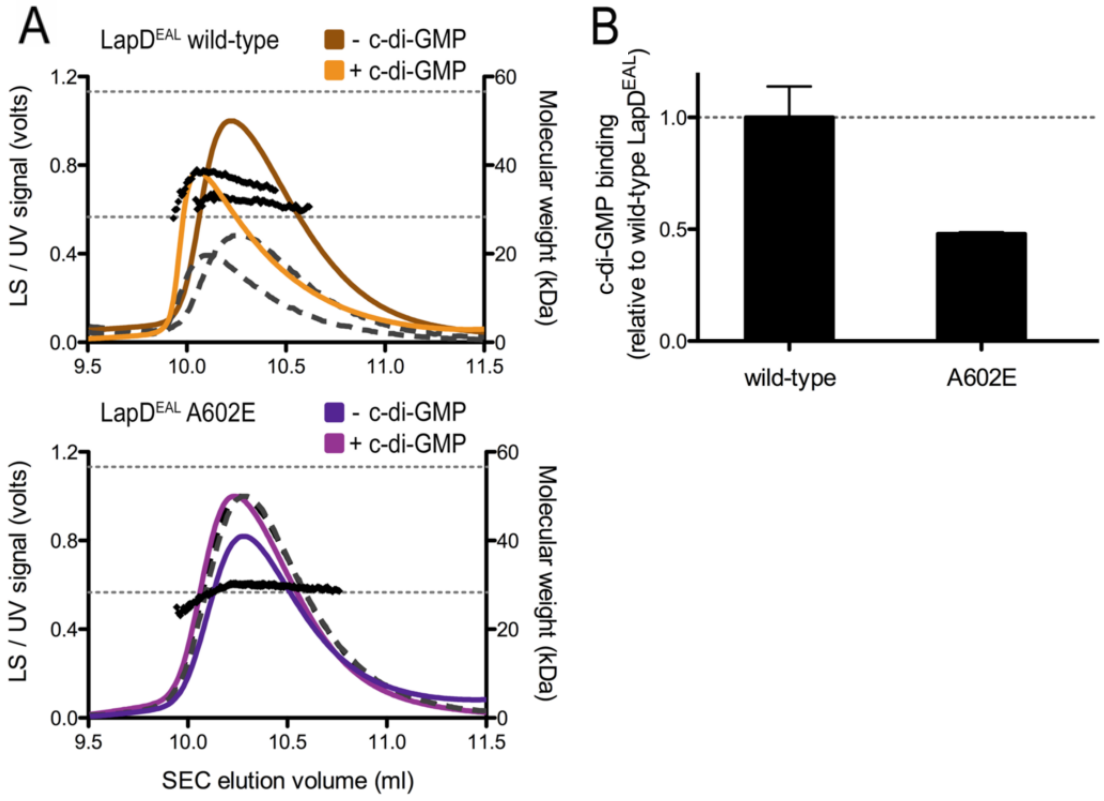


Figure 3.12: Cyclic di-GMP binding and quaternary structure of LapD^{EAL} in solution.

(A) Oligomerization in solution. Size exclusion chromatography (SEC)-coupled multi-angle light scattering (MALS) analysis of wild-type and mutant LapD^{EAL} in the presence and absence of c-di-GMP are shown. The signal from the 90°-scattering detector is shown in color and the signal from the refractive index detector is shown as dashed line. Average molecular weights are plotted in black against the right Y-axis, as calculated every second across the protein elution peak. Theoretical molecular weights corresponding to those of a monomer and a dimer are indicated as horizontal dashed, grey lines. Injected protein and nucleotide concentrations were 250 μM and 500 μM, respectively.

(B) Cyclic di-GMP binding. Purified LapD^{EAL} (wild-type or A^{602}E) was incubated in the presence of c-di-GMP. Excess nucleotide was removed by gel filtration, and protein-bound c-di-GMP levels were assessed by reverse-phase HPLC after normalization for protein content and heat-denaturation. Data are expressed relative to the amount of c-di-GMP bound to wild-type LapD^{EAL}.

Figure 3.13: Phenotypic analyses of *lapD* mutants (Data courtesy of Peter D. Newell and George A. O'Toole).
(A) Biofilm phenotypes of double-mutants. Biofilm formation of Δ*lapD* cells expressing full-length, wild-type LapD or LapD point mutants was assessed. Quantification of crystal violet-stained biofilms is shown. Data are means -/+ SD of 8 replicates. Protein levels were determined by Western blotting using a primary antibody that recognizes the His_6-epitope at the C-terminus of LapD.
(B) Phosphate-regulated c-di-GMP signaling. Phosphate starvation leads to the expression the active phosphodiesterase RapA and a reduction in cellular c-di-GMP concentration (*S1*). LapD mutants were tested for their response to limiting phosphate concentration. Biofilm formation was monitored over 90 min after activation of the Pho system upon transfer to low-phosphate media, and compared to biofilm formation in phosphate-rich media. The mutant ΔH1 contains an activating deletion in the HAMP domain and has been described previously (*S1*). Mutants assessed for biofilm formation in Figure 5A but not included in Figure 5B are shown.

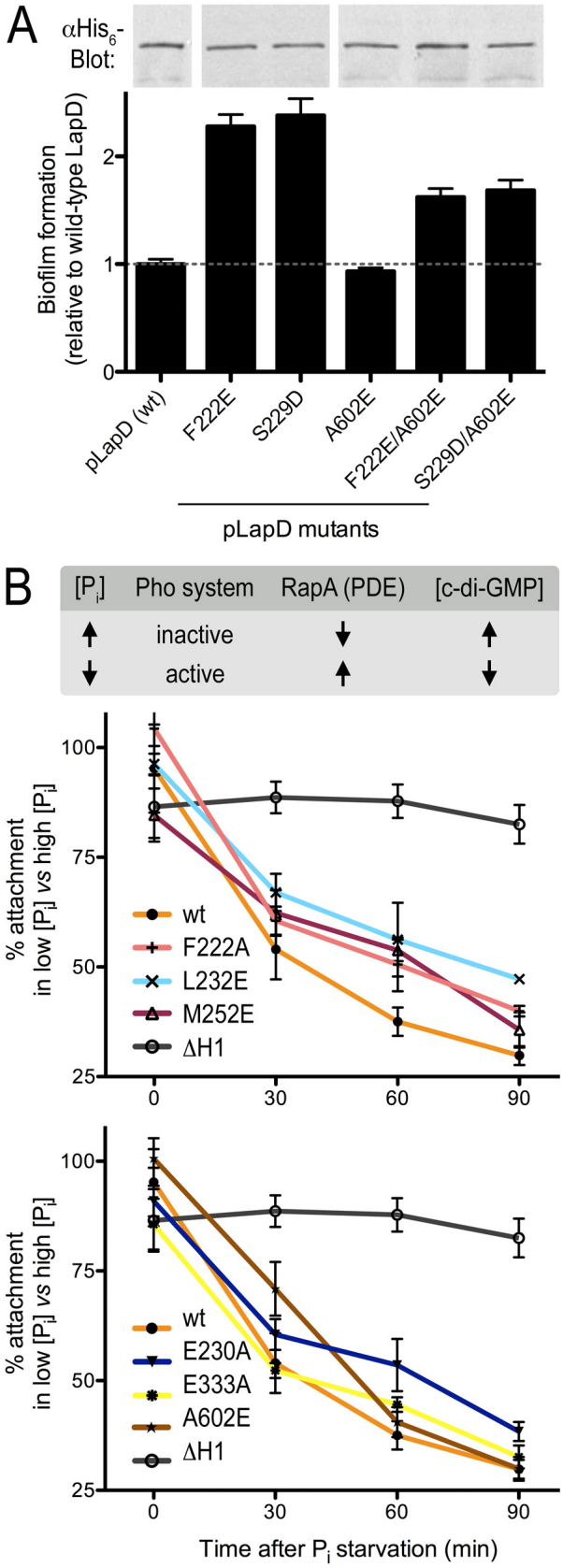
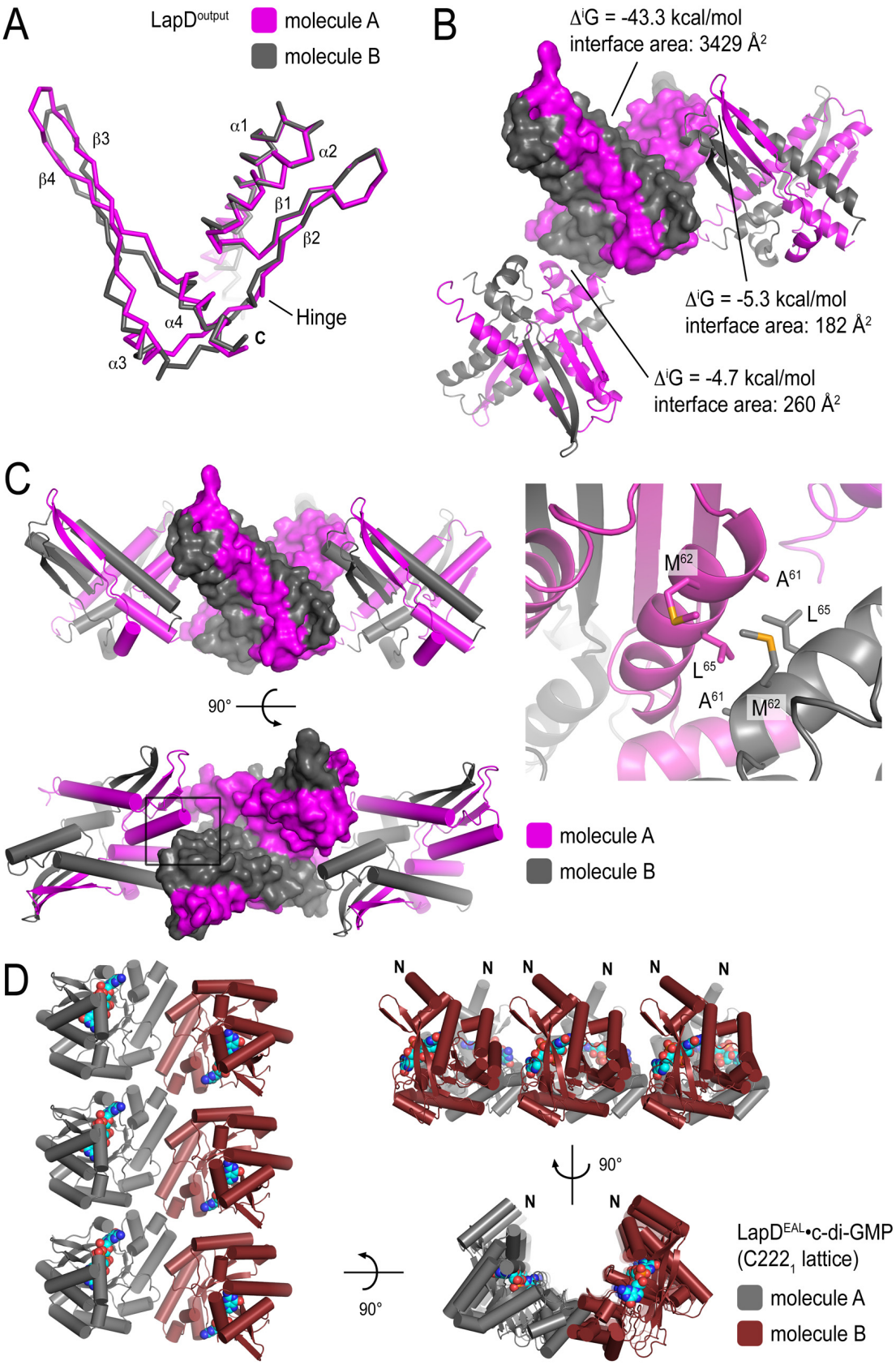

A
αHis6-
Blot:
Biofilm formation
(relative to wild-type LapD)
0
1
2
pLapD (wt)
F222E
S229D
A602E
F222E/A602E
S229D/A602E
pLapD mutants
B
[Pi]
Pho system
RapA (PDE)
[c-di-GMP]
inactive
active
% attachment
in low [Pi] vs high [Pi]
100
75
50
25
0
30
60
90
wt
F222A
L232E
M252E
ΔH1
wt
E230A
E333A
A602E
ΔH1
Time after Pi starvation (min)

Figure 3.14: Structural analysis of LapD^{output} and potential mechanisms for higher-order oligomerization of LapD.
(A) Comparison between LapD^{output} protomers. The periplasmic output domain of LapD crystallized with two molecules in the asymmetric unit. The protomers were superimposed on the first two helices of the fold, revealing a minor, rigid-body rotation of one half of the molecule relative to the other half between the two protomers. The rotation occurs at the connecting loop between β2 and α3 that forms the crossing-over point in the domain-swapped dimer.
(B) LapD^{output} crystal packing. Domain-swapped dimers of the output domain interact predominantly via two interfaces in the crystal lattices. One involves bottom-to-bottom interaction between LapD^{output} dimers via a conserved, hydrophobic patch coinciding with the putative membrane-interaction surface. The other interface involves hydrophobic interactions between the arms of the V-shaped output domain dimers.
(C) Potential higher-order oligomerization based on the structure of LapD^{output}. Crystal lattice contacts reveal a potential mode for higher-order assemblies of LapD. The close-up view (right panel) shows the hydrophobic contacts between output domain dimers.
(D) Potential higher-order oligomerization based on the structure of LapD^{EAL}•c-di-GMP. In the $C222_1$ crystal lattice, EAL domains form higher-order lattices that may highlight a mode for receptor oligomerization in the membrane.


A
LapD^output
molecule A
molecule B
β3
β4
α1
α2
β1
β2
Hinge
α4
c
α3
B
ΔiG = -43.3 kcal/mol
interface area: 3429 Å2
ΔiG = -5.3 kcal/mol
interface area: 182 Å2
ΔiG = -4.7 kcal/mol
interface area: 260 Å2
C
90°
M62
A61
L65
L65
A61
M62
molecule A
molecule B
D
N
N
N
N
N
N
90°
90°
N
N
LapD^EAL•c-di-GMP
(C222, lattice)
molecule A
molecule B

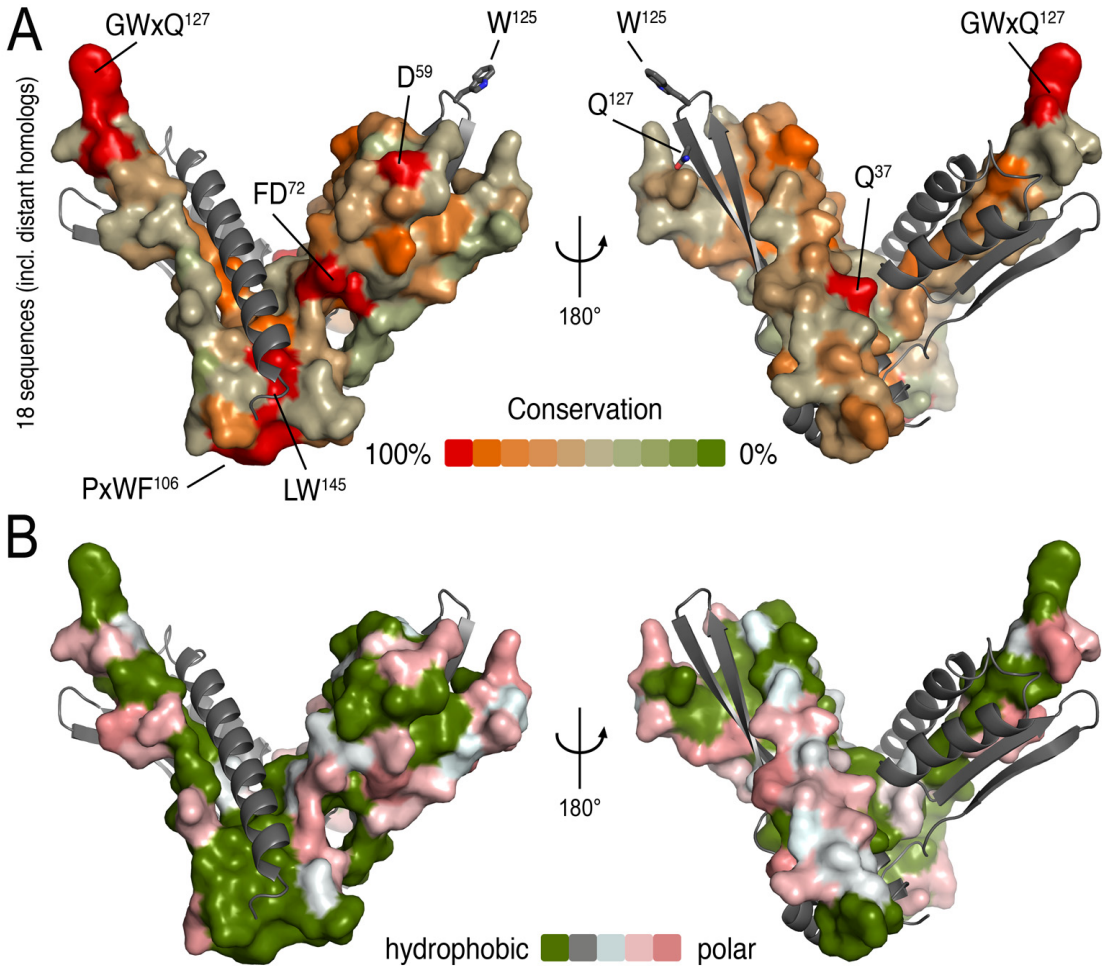


Figure 3.15: Surface conservation and hydrophobicity of LapD^{output}.
(A) Surface conservation. Based on an alignment of 18 sequences of LapD homologs (Fig. 3.8), the sequence conservation was mapped onto the solvent-accessible surface of the output domain. One protomer is shown in surface presentation, the other is shown in ribbon presentation. Conserved motifs and individual residues are highlighted. Two views, separated by a 180° rotation, are shown.
(B) Hydrophobicity mapped onto the molecular surface of LapD^{output}. The surface is colored according to the hydrophobicity of accessible residues. Hydrophobic residues are shown in green, polar and charged residues are in gray and pink, respectively.

S3.3. Supplemental Tables

Table S1: X-ray data collection and refinement statistics

	$LapD^{EAL}$•c-di-GMP (1)	$LapD^{EAL}$•c-di-GMP (2)	$LapD^{output}$ (Se-Met)	$LapD^{dual}$ (1)	$LapD^{dual}$ (2)	$LapD^{dual}$ (2) (Se-Met)
Data Collection						
X-ray source	CHESS, A1	CHESS, A1	CHESS, A1	CHESS, F2	CHESS, A1	CHESS, A1
Wavelength (Å)	0.9769	0.9769	0.9769	0.9793	0.9769	0.9769
Space group	$C222_1$	$P6_522$	$P2_12_12_1$	I32	$P3_2$	$P3_2$
Unit cell						
a, b, c (Å)	41.4, 204.8, 142.4	141.5, 141.5, 111.2	41.5, 71.4, 110.4	154.8, 154.8, 154.8	52.1, 52.1, 146.3	52.4, 52.4, 146.5
α, β, γ (°)	90, 90, 90	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 120	90, 90, 120
Resolution (Å)[a]	50-2.5 (2.59-2.5)	50-2.5 (2.59-2.5)	50-1.8 (1.83-1.77)	50-3.1 (3.29-3.1)	30-2.0 (2.07-2.0)	50-2.5 (2.59-2.5)
No. of reflections						
Total	172,322 (12,035)	568,367 (34,934)	433,383 (26,900)	123,220 (18,906)	188,981 (17,162)	105,992 (8,859)
Unique	20,943 (1,973)	23,191 (2,211)	32,517 (3,092)	11,246 (1,775)	29,997 (2,959)	31,174 (3,055)
Completeness (%)	99.1 (94.4)	99.7 (97.4)	99.5 (96.7)	98.9 (98.7)	99.7 (98.8)	99.8 (98.4)
Redundancy	8.2 (6.1)	24.5 (15.8)	13.3 (8.7)	10.9 (10.7)	6.3 (5.8)	3.4 (2.9)
I/σ(I)	19.8 (3.5)	34.4 (3.7)	30.2 (4.1)	29.4 (3.4)	42.6 (3.3)	24.0 (3.2)
R_{meas} (%)	8.7 (33.9)	7.5 (43.6)	6.5 (37.3)	7.8 (78.5)	4.4 (47.7)	4.7 (28.2)
Refinement						
R_{work} / R_{free} (%)	18.1 / 24.5	19.6 / 22.5	19.4 / 22.5	23.4 / 28.4	18.1 / 22.2	
rms deviations						
Bond length (Å)	0.009	0.008	0.006	0.011	0.011	
Bond angles (°)	1.271	1.239	0.994	1.405	1.233	
No. of atoms						
Protein	3866	2001	1979	3315	3315	
c-di-GMP	92	46	0	0	0	
Water	115	125	304	0	180	
Ave. B-factors (Å^2)						
Protein	37.2	54.2	32.6	83.8	54.0	
cyclic di-GMP	28.4	46.9	--	--	--	
Water	34.3	50.6	43.3	--	51.1	
Ramachandram (%)						
Favored	93.2	93.2	94.3	88.5	93.6	
Allowed	6.8	6.8	5.7	11.2	5.8	
Generously allowed	0.0	0.0	0.0	0.0	0.3	
Disallowed	0.0	0.0	0.0	0.3	0.3	

(a) Values in brackets are for the highest resolution bin.

Table S2: Conservation of the LapD/LapG signaling system

Organism	LapG-like	E-value	LapD-like neighbor	E-value	ABC transporter	Putative substrate	Description	Putative cleavage site
Pseudomonas fluorescens	PFL_0130	6E-146	PFL_0131	0	X	PFL_0133	LapA Homolog	QQAIAAGVDPTTALESTAAGPSAAGTGGAA
Pseudomonas putida	PP_0164	7E-95	PP_0165	0	X	PP_0168	LapA Homolog	QQAIAAGVDPTTELEATAAGPSSAGGGAL
Pseudomonas entomophilia	PSEEN0138	4E-93	PSEEN0139	0	X	PSEEN0141	LapA-like adhesin	QQAIAAGADPTTELEATAAGPAAAGGGSV
Erwinia carotovora atroseptica	ECA3263	3E-66	ECA3264	0	X	ECA3266	LapA-like adhesin	QDAIAQGADPTQVLEATAAGNDNTGEAG
Pseudomonas aeruginosa	PA1434	9E-66	PA1433	0	?	?	?	?
Desulfotalea psychrophila	DP0518	1E-47	DP0517	1E-83	X	DP0516	Large RTX protein	QALAEGKSIDDVLEKTAAGTEGSGGSYDF
Bordetella parapertussis	BPP0973	3E-45	BPP0972	3E-56	X	BPP0974	Hemagglutinin-like protein	LAALQDGRDPFDELDPTAAVIGGSGDSAGS
Chromobacterium violaceum	CV_0309	3E-45	CV_0310	7E-56	X	CV_0311	LapA-like adhesin	QQILAALDNPSANPGNPFDNLDPAAAGLND
Vibrio cholerae EI Tor	VC_A1081	7E-43	VC_A1082/1083	2E-36/3E-14	X	VCA0849	Hemagglutinin-like protein	QQAILDGVDPTTALEAAAAGAGAGG
Bordetella bronchiseptica	BB1185	3E-45	BB1184	2E-56	X	BB1186	Hemagglutinin-like protein	QDGRDPFDELDPTAAVIGGSGDSAGS
Azoarcus sp EbN	ebA1785	4E-45	ebA1787	5E-83	X	ebA1795	Hypothetical protein	NQIIEAINEGANLDDVLEAPAAGLAGGGGG
Vibrio cholerae 1587	A55_A0980	8E-43	A55_A0980	7E-54	?	A55_A0690	TP1SS repeat 143	QQAILDGVDPTTALEAAAAGAGAGG
Rhodoferax ferrireducens	Rfer_3763	1E-42	Rfer_3764	3E-90	X	Rfer_3766	Hemolysin-type RTX protein	IQALERGTDLSTELEATAAGLGVGGG
Shewanella denitrificans	Sden_0379	1E-42	Sden_0378	2E-52	X	Sden_0384	Large RTX protein	QDLIASGEDPTEDLPETAAGTPTGNQGNS
Vibrio fischeri	VF_A1167	3E-42	VF_A1166	3E-56	X	VF_A1162	FrpC-like RTX protein	LNAIIDGEDPSLITEAPAAGEDSGS
Vibrio vulnificus CMCP6	VV2_1126	3E-41	VV2_1127	2E-54	X	VV1_2715	RTX protein	QQIFAALEEGQDPTQLGDDFATAAGETGG
Thiomicrospira crunogena	Tcr_0209	5E-41	Tcr_0208	2E-78	X	Tcr_0105	VCBS protein	EEIRKAIEAVQEGNFDALEETAAGQQNDL
Photobacterium profundum	PBPRB0581	5E-41	PBPRB0580	3E-56	X	PBPRB0585	Hypothetical protein	QNAILSGDDPTETLDAAAAGNEAQGSS
Polaromonas sp.	Bpro_0309	2E-40	Bpro_0308	3E-92	X	Bpro_0306	Large RTX protein	QALERGTDLNQSLEATAAGLVPGGG
Legionella pneumophila lens	lpl0859	3E-40	lpl0860	4E-81	?	lpl0681	RtxA/TP1SS repeat 143	QEAIAKGIDPSIILDVLGSAAAGAEAVGSG
Methylibium petroleiphilum	Mpe_A1879	4E-40	Mpe_A1878	3E-79	X	Mpe_A1877	Large RTX protein	QALNEGDPDAATAAGGPASGGEA
Vibrio parahaemolyticus RIMD	VPA1734	9E-40	VPA1735	3E-60	X	VPA0668	TP1SS repeat 143	QQAILEGADPTAILEATAAGGDASG

Table S3: Strains and plasmids

Strain or plasmid	Genotype or Description	Reference
Escherichia coli		
S17-1(λpir)	*thi pro hsdR- hsdM+* Δ*recA* RP4-2::TcMu-Km::Tn7	(Simon et al., 1983)
Saccharomyces cerevisiae		
InvSc1	uracil auxotroph	Invitrogen
Pseudomonas fluorescens		
Δ*lapD*	unmarked deletion of the *lapD* gene	(Newell et al., 2009)
Plasmids		
pMQ72	*2μ URA3; ori & rep* pR01600; *colE1*, *aac1* P_{BAD} *araC*	(Shanks et al., 2006)
pLapD	LapD with C-terminal 6 His tag expressed from P_{BAD}	(Newell et al., 2009)
pLapD ΔH1	pLapD with 7 amino acid deletion in first helix of HAMP domain	(Newell et al., 2009)
pLapD W125E	pLapD point mutation in the indicated codon or codons	This study
pLapD W125E,S229D	pLapD point mutation in the indicated codon or codons	This study
pLapD F222A	pLapD point mutation in the indicated codon or codons	This study
pLapD F222E	pLapD point mutation in the indicated codon or codons	This study
pLapD S229D	pLapD point mutation in the indicated codon or codons	This study
pLapD E230A	pLapD point mutation in the indicated codon or codons	This study
pLapD L232E	pLapD point mutation in the indicated codon or codons	This study
pLapD M252E	pLapD point mutation in the indicated codon or codons	This study
pLapD E262A	pLapD point mutation in the indicated codon or codons	This study
pLapD E333A	pLapD point mutation in the indicated codon or codons	This study
pLapD A602E	pLapD point mutation in the indicated codon or codons	This study
pLapD F222E,A602E	pLapD point mutation in the indicated codon or codons	This study
pLapD S229D,A602E	pLapD point mutation in the indicated codon or codons	This study

SUPPLEMENTAL REFERENCES

S1. Newell, P. D., R. D. Monds, et al. (2009). "LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by Pseudomonas fluorescens Pf0-1." *Proc Natl Acad Sci U S A* **106**(9): 3461-6.

S2. Shanks, R. M., N. C. Caiazza, et al. (2006a). "Saccharomyces cerevisiae-based molecular tool kit for manipulation of genes from gram-negative bacteria." *Appl Environ Microbiol* **72**(7): 5027-36.

S3. Krasteva, P. V., J. C. Fong, et al. (2010). "Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP." *Science* **327**(5967): 866-8.

S4. Otwinowski, Z., and W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." *Methods Enzymol* **276**: 307-326.

S5. Kabsch, W. "Xds." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 125-32.

S6. Adams, P. D., R. W. Grosse-Kunstleve, et al. (2002). "PHENIX: building new software for automated crystallographic structure determination." *Acta Crystallogr D Biol Crystallogr* **58**(Pt 11): 1948-54.

S7. Cowtan, K. (1994). 'dm': "An automated procedure for phase improvement by density modification." *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* ***31***: 34-38.

S8. Cowtan, K. (2006). "The Buccaneer software for automated model building. 1. Tracing protein chains." *Acta Crystallogr D Biol Crystallogr* **62**(Pt 9): 1002-11.

S9. Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-32.

S10. De, N., M. Pirruccello, P.V. Krasteva et al. (2008). "Phosphorylation-

independent regulation of the diguanylate cyclase WspR." *PLoS Biol* **6**(3): e67.

S11. Choi, K. H., A. Kumar, et al. (2006). "A 10-min method for preparation of highly electrocompetent Pseudomonas aeruginosa cells: application for DNA fragment transfer between chromosomes and plasmid transformation." *J Microbiol Methods* **64**(3): 391-7.

S12. Monds, R. D., P. D. Newell, et al. (2006). "Conservation of the Pho regulon in Pseudomonas fluorescens Pf0-1." *Appl Environ Microbiol* **72**(3): 1910-24.

S13. Larkin, M. A., G. Blackshields, et al. (2007). "Clustal W and Clustal X version 2.0." *Bioinformatics* **23**(21): 2947-8.

S14. Gouet, P., E. Courcelle, et al. (1999). "ESPript: analysis of multiple sequence alignments in PostScript." *Bioinformatics* **15**(4): 305-8.

S15. Chan, C., R. Paul, et al. (2004). "Structural basis of activity and allosteric control of diguanylate cyclase." *Proc Natl Acad Sci U S A* **101**(49): 17084-9.

S16. De, N., M. V. Navarro, et al. (2009). "Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR." *J Mol Biol* **393**(3): 619-33.

S17. Wassmann, P., C. Chan, et al. (2007). "Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition." *Structure* **15**(8): 915-27.

S18. Minasov, G., S. Padavattan, et al. (2009). "Crystal structures of YkuI and its complex with second messenger cyclic Di-GMP suggest catalytic mechanism of phosphodiester bond cleavage by EAL domains." *J Biol Chem* **284**(19): 13174-84.

ACKNOWLEDGMENTS

Before I continue to the next chapter, I would like to acknowledge my colleagues and collaborators for their respective contributions to the LapD protein project. First and foremost, I would like to thank my mentor, Dr. Holger Sondermann, for his help in research design, data analysis and manuscript writing. I am also grateful to George A. O'Toole and Peter D. Newell for designing and conducting the cell-based assays; to Marcos V. A. S. Navarro for crystallographic studies on the LapD dual-domain protein construct and for general help with crystallographic data analysis; and to Debashree Chatterjee for crystallization trials and *in vitro* interaction studies involving LapD's output periplasmic module.

CHAPTER IV

Conclusions

Bacterial pathogenesis involves collaborative group behaviors such as quorum sensing, swarming, and biofilm formation. While prevalent quorum sensing correlates with virulence gene expression and acute-phase infections, biofilm formation allows for the development of chronic infections and microbial survival in hostile environments (*1, 2*). It has been appreciated only recently that these social behaviors are highly regulated developmental processes involving strictly coordinated inter- and intracellular signaling mechanisms. Cyclic di-GMP, an intracellular second messenger unique to the bacterial world, has emerged as a key regulator of the various processes involved in the initiation, progression, and dispersal of bacterial biofilms (*3, 4*). Diguanylate cyclases containing GGDEF domains and phosphodiesterases containing either EAL or HD-GYP domains have been identified as the enzymes controlling cellular c-di-GMP levels, yet little is known regarding signal transmission and the sensory targets of c-di-GMP (*5, 6*).

Comparative genomic analyses have been instrumental for the characterization of the various c-di-GMP turnover domains in bacteria (*6-9*). The available complete sequences of scores of bacterial genomes– and therefore possible cross-genome sequence alignments– have helped identify and predict the metabolic function of a number of signaling modules based on conserved sequence motifs, predicted secondary structure or domain organization, inter- and intraoperon genetic environment, and phylogenetic patterns of cross-species evolution (*6*). GGDEF, EAL, and HD-GYP domains have thus been characterized as some of the most abundant

protein modules encoded by bacterial genomes, with their representation per species roughly correlating with the organisms' adaptational capacities (5, 6). The nomenclature of these c-di-GMP turnover domains is also a direct result of comparative sequence analyses: GGDEF, EAL, and HD-GYP correspond to characteristic conserved amino acid motifs found in the corresponding signaling modules (*6-9*).

Similar bioinformatic analysis identified PilZ domains as protein modules involved in various c-di-GMP signaling pathways and predicted that they might function as direct sensors relaying the second messenger's input (*10*). Further experimental evidence has corroborated c-di-GMP binding to several PilZ domain-containing proteins, as well as their direct involvement in biofilm formation or expression of virulence determinants (*11-15*).

Interestingly, many bacteria that utilize c-di-GMP mediated signaling for adaptation lack PilZ domains encoded in their genomes (*10*). In addition, in some organisms PilZ signaling modules seem to control some, but not all of the c-di-GMP dependent processes involved in biofilm formation and pathogenicity. For example, while PilZ domain-containing proteins in *Vibrio cholerae* are not essential for rugosity and exopolysaccharide production, some of them are required for efficient intestinal colonization during environment-to-host transition (*13, 16*).

It has thus become increasingly evident that the gamut of bacterial c-di-GMP receptors spreads far beyond the family of PilZ protein domains. Identified c-di-GMP sensors include bacterial riboswitches, transcription factors of various domain architectures, divergent c-di-GMP turnover domains, and allosteric sites on active or degenerate diguanylate cyclases (*16-25*). Interestingly, some of the identified c-di-GMP targets belong to protein families generally involved in sensing or metabolizing different nucleotide-based small molecules. For example, FleQ of *Pseudomonas*

aeruginosa belongs to the AAA+ superfamily of ATPases, but seems to function as a c-di-GMP signal effector independent of ATP binding or hydrolysis (*18*). Similarly, Clp of *Xanthomonas campestris* is classified as a catabolite activator-like protein, whose homolog in *Escherichia coli* regulates gene expression in a cAMP-dependent manner (*20, 26*). On one hand this highlights the functional complexity and rapid evolution of homologous proteins in bacterial signaling networks. On the other, it is somewhat expected based on the fact that c-di-GMP turnover domains are themselves homologous to protein domains involved in the synthesis or hydrolysis of different nucleic acid-based metabolites. For example, while GGDEF domains are evolutionary close to adenylate cyclases (*9*), HD-GYP domains belong to the larger HD superfamily of phosphohydrolases with nucleotide-based substrates such as dGTP or ppGpp (*6*). This raises the question whether there are a limited number of universal c-di-GMP binding motifs and/or protein folds, or individual c-di-GMP receptors have evolved specific modes of recognition to ensure signal isolation. The two scenarios are not mutually exclusive and bioinformatic analyses based on sequence alignments and available structural data has so far been able to identify some common c-di-GMP binding features. These include the abovementioned PilZ domains (*10*), a prevalent riboswitch class in bacterial messenger RNAs (*17*), the characteristic I-site RxxD motifs on degenerate (PelD of *Pseudomonas aeruginosa*, CdgG of *Vibrio cholerae*) or active (PleD of *Caulobacter crescentus*, WspR of *Pseudomonas aeruginosa*) GGDEF modules (*16*, *23-25*, *27*), as well as the potential for catalytically inactive EAL domains to serve as c-di-GMP sensors (FimX of *Pseudomonas aeruginosa,* LapD of *Pseudomonas fluorescens*) (*21, 22*). It is nevertheless important to note the following: first, not all members of the abovementioned classes of potential c-di-GMP receptors would serve as such *in vivo*; second, the diversity of protein domain architectures, putative domain-domain interactions, and adopted c-di-GMP conformations

complicates prediction-based approaches for the identification of novel c-di-GMP receptors; third, many of the identified c-di-GMP sensors are so-far stand-alone representatives with uncharacterized mode of nucleotide recognition (FleQ of *Pseudomonas aeruginosa*, Clp of *Xanthomonas campestris*) (*18, 20, 26*); and forth, other classes of c-di-GMP receptors are likely to exist in nature.

Here we set out to identify and provide detailed mechanistic analyses of several putative or known c-di-GMP receptors. We identified VpsT of *V. cholerae* as a novel c-di-GMP receptor and presented the crystal structures of the c-di-GMP•free and nucleotide-bound states (Chapter II). As a response regulator with an N-terminal receiver domain and a C-terminal helix-turn-helix DNA-binding module, VpsT differs from previously identified c-di-GMP receptors and underscores the diversity of the second messenger's targets. Our structural and functional studies provided comprehensive characterization of the protein's c-di-GMP•dependent oligomerization propensity and its relevance as a mechanism for signal integration and global transcription control. Based on our analyses, we defined a novel structural class of receiver domain proteins, likely utilizing a distinct structural feature to dimerize upon signaling inputs with or without concurrent phosphorylation events. In addition, we were able to propose a model for VpsT-mediated changes in DNA architecture, and outlined a number of future experiments that would prove or disprove our hypotheses (Chapter II, Appendices A and B). Thus, our collaborative interdisciplinary approach provided for the first time in the literature an in-depth structural and functional analysis of a global transcription regulator able to inversely control the production of extracellular matrix and flagellar motility under direct c-di-GMP signaling input.

Our initial interest in VpsT as a putative c-di-GMP sensor has been largely based on its homology to CsgD from *Escherichia coli* and *Salmonella typhimurium*, a protein whose expression and function are tightly intertwined with various c-di-GMP

signaling pathways (*28, 29*). In addition, CsgD has been postulated to be incompetent for phosphotransfer and to instead be dependent on small molecule binding for efficient gene expression regulation (*29, 30*). Based on the crystal structure of the VpsT • c-di-GMP complex and on nucleotide-binding studies on CsgD *in vitro*, we determined that CsgD contains a divergent c-di-GMP•binding site and is uncapable of c-di-GMP recognition (data not shown). Whether the protein binds c-di-GMP or another small molecule *in vivo* remains so-far unknown. A recent study has proposed that CsgD could be nevertheless regulated by phosphorylation, as the protein autophosphorylates in the presence of high concentrations (10mM) of acetyl phosphate *in vitro* (*31*). While such phosphorylation appears to inhibit protein binding to its cognate DNA, it remains unclear whether this is due to a distinct functional state or is an artifact of non-specific binding and reduced protein stability. In support of the latter, acetyl phosphate does not affect CsgD function *in vivo* and a loss-of-function asparagine substitution of the conserved phosphoreceiver aspartate ($D^{59}N$) preserves phosphate binding to the protein *in vitro* (*31*). In addition, a phosphomimetic aspartate to glutamate substitution ($D^{59}E$) appears to inactivate the protein, but the mutant is indeed characterized by markedly reduced protein stability (*31*). Based on our structural studies on VpsT (*19*), phosphoreceiver D^{59} is likely proximal to a dimerization interface, disruption of which might explain the mutant's altered stability and function and account for the observed effects *in vitro* and *in vivo*.

Althouhg data available on the regulation of CsgD *in vivo* remains scarce and inconclusive, it is clear that the FixJ/LuxR/CsgD family of response regulators has evolved an array of diverse regulatory sensitivities. Based on our structural studies and comparative sequence analysis, we were able to identify conserved structural features (α6-dimerization, c-di-GMP coordinating residues, residues connecting the phosphoreceiver site with the nucleotide-binding pocket) that could shed light on the

evolution from archetypal two-component phosphotransfer to small molecule recognition and facilitate the prediction of additional c-di-GMP targets. Such putative c-di-GMP receptors include VC0396, a VpsT homolog encoded by the *Vibrio cholerae* genome with a functional c-di-GMP binding motif (data not shown), and PA0533, a PAS domain-containing homolog in the opportunistic pathogen *Pseudomonas aeruginosa*.

Similarly to VpsT, FleQ of *Pseudomonas aeruginosa* is a transcription regulator which inversely regulates flagellar motility and exopolysaccharide production upon direct c-di-GMP recognition (*18*). Since structural data regarding the exact mechanism of this nucleotide-dependent switch is currently unavailable, we set out to determine the crystal structures of its c-di-GMP•bound and nucleotide-free states. To this end, we have crystallized a functional protein construct in the presence of c-di-GMP and have obtained an experimental phase solution at 3.33 Å resolution (Chapter II, Appendix B). While further model building and refinement will be necessary for the identification of functionally important residues and the ligand recognition motif, we are confident that our studies will help expand the list of putative c-di-GMP receptors and will shed light on the function of biologically important homologs. Among those is *Vibrio cholerae* VpsR, an essential partner of VpsT in the regulation of rugosity and biofilm formation (*32*).

Besides regulators of gene expression, c-di-GMP sensors include proteins with degenerate c-di-GMP turnover domains. One such protein is LapD of *Pseudomonas fluorescens* (*22*): a transmembrane receptor that contains a conserved HAMP-GGDEF-EAL domain intracellular module, which senses cellular c-di-GMP levels and controls the conformation of the periplasmic output domain. Interaction of c-di-GMP bound LapD with the periplasmic protease LapG is crucial for maintenance of the surface adhesin LapA and leads to stable biofilm formation (*33*). Based on three

novel structures, solution biophysical and biochemical studies, as well as cell based phenotypic assays, we presented a model for the autoinhibition and activation of LapD in response to its ligand availability (Chapter III). In addition, we identified many features and functionally important motifs to be conserved in a diverse range of free-living and pathogenic species, where homologous systems of inside-out signaling are likely to control biofilm formation or toxin expression.

In summary, by using structural biology methods we were able to determine the atomic resolution structures and conformational states of several c-di-GMP receptors, members of various protein families in terms of both conserved domain organization and level of functional impact. We then used these molecular snapshots to guide our research into the proteins' function as signal integrators in the big picture of intracellular signaling networks. We identified functional motifs and protein interfaces likely to apply to homologs in a variety of species and thus provided basis for future cross-genome comparative studies and targeted function assessments. We believe that the results of our studies provide valuable information about c-di-GMP mediated virulence and adaptation, as well as, in the long term, targets for the development of novel therapeutics against bacterial infections.

REFERENCES

1. Parsek, M. R. and P. K. Singh (2003). "Bacterial biofilms: an emerging link to disease pathogenesis." *Annu Rev Microbiol* **57**: 677-701.
2. Furukawa, S., S. L. Kuchma, et al. (2006). "Keeping their options open: acute versus persistent infections." *J Bacteriol* **188**(4): 1211-7.
3. O'Toole, G., H. B. Kaplan, et al. (2000). "Biofilm formation as microbial development." *Annu Rev Microbiol* **54**: 49-79.
4. Jenal, U. and J. Malone (2006). "Mechanisms of cyclic-di-GMP signaling in bacteria." *Annu Rev Genet* **40**: 385-407.
5. Romling, U., M. Gomelsky, et al. (2005). "C-di-GMP: the dawning of a novel bacterial signalling system." *Mol Microbiol* **57**(3): 629-39.
6. Galperin, M. Y., A. N. Nikolskaya, et al. (2001). "Novel domains of the prokaryotic two-component signal transduction systems." *FEMS Microbiol Lett* **203**(1): 11-21.
7. Galperin, M. Y., D. A. Natale, et al. (1999). "A specialized version of the HD hydrolase domain implicated in signal transduction." *J Mol Microbiol Biotechnol* **1**(2): 303-5.
8. Hecht, G. B. and A. Newton (1995). "Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in Caulobacter crescentus." *J Bacteriol* **177**(21): 6223-9.
9. Pei, J. and N. V. Grishin (2001). "GGDEF domain is homologous to adenylyl cyclase." *Proteins* **42**(2): 210-6.
10. Amikam, D. and M. Y. Galperin (2006). "PilZ domain is part of the bacterial c-di-GMP binding protein." *Bioinformatics* **22**(1): 3-6.
11. Ryjenkov, D. A., R. Simm, et al. (2006). "The PilZ domain is a receptor for

the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria." *J Biol Chem* **281**(41): 30310-4.

12. Benach, J., S. S. Swaminathan, et al. (2007). "The structural basis of cyclic diguanylate signal transduction by PilZ domains." *Embo J* **26**(24): 5153-66.
13. Pratt, J. T., R. Tamayo, et al. (2007). "PilZ domain proteins bind cyclic diguanylate and regulate diverse processes in Vibrio cholerae." *J Biol Chem* **282**(17): 12860-70.
14. Ramelot, T. A., A. Yee, et al. (2007). "NMR structure and binding studies confirm that PA4608 from Pseudomonas aeruginosa is a PilZ domain and a c-di-GMP binding protein." *Proteins* **66**(2): 266-71.
15. Ko, J., K. S. Ryu, et al. (2010). "Structure of PP4397 reveals the molecular basis for different c-di-GMP binding modes by Pilz domain proteins." *J Mol Biol* **398**(1): 97-110.
16. Beyhan, S., L. S. Odell, et al. (2008). "Identification and characterization of cyclic diguanylate signaling systems controlling rugosity in Vibrio cholerae." *J Bacteriol* **190**(22): 7392-405.
17. Sudarsan, N., E. R. Lee, et al. (2008). "Riboswitches in eubacteria sense the second messenger cyclic di-GMP." *Science* **321**(5887): 411-3.
18. Hickman, J. W. and C. S. Harwood (2008). "Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor." *Mol Microbiol* **69**(2): 376-89.
19. Krasteva, P. V., J. C. Fong, et al. (2010). "Vibrio cholerae VpsT regulates matrix production and motility by directly sensing cyclic di-GMP." *Science* **327**(5967): 866-8.
20. Chin, K. H., Y. C. Lee, et al. (2010) "The cAMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene

expression in Xanthomonas campestris." *J Mol Biol* **396**(3): 646-62.

21. Navarro, M. V., N. De, et al. (2009). "Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX." *Structure* **17**(8): 1104-16.

22. Newell, P. D., R. D. Monds, et al. (2009). "LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by Pseudomonas fluorescens Pf0-1." *Proc Natl Acad Sci U S A* **106**(9): 3461-6.

23. Chan, C., R. Paul, et al. (2004). "Structural basis of activity and allosteric control of diguanylate cyclase." *Proc Natl Acad Sci U S A* **101**(49): 17084-9.

24. De, N., M. Pirruccello, P. V. Krasteva et al. (2008). "Phosphorylation-independent regulation of the diguanylate cyclase WspR." *PLoS Biol* **6**(3): e67.

25. Lee, V. T., J. M. Matewish, et al. (2007). "A cyclic-di-GMP receptor required for bacterial exopolysaccharide production." *Mol Microbiol* **65**(6): 1474-84.

26. Leduc, J. L. and G. P. Roberts (2009). "Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of Xanthomonas axonopodis pv. citri." *J Bacteriol* **191**(22): 7121-2.

27. Christen, B., M. Christen, et al. (2006). "Allosteric control of cyclic di-GMP signaling." *J Biol Chem* **281**(42): 32015-24.

28. Kader, A., R. Simm, et al. (2006). "Hierarchical involvement of various GGDEF domain proteins in rdar morphotype development of Salmonella enterica serovar Typhimurium." *Mol Microbiol* **60**(3): 602-16.

29. Romling, U. (2005). "Characterization of the rdar morphotype, a multicellular behaviour in Enterobacteriaceae." *Cell Mol Life Sci* **62**(11): 1234-46.

30. Chirwa, N. T. and M. B. Herrington (2003). "CsgD, a regulator of curli and cellulose synthesis, also regulates serine hydroxymethyltransferase synthesis in Escherichia coli K-12." *Microbiology* **149**(Pt 2): 525-35.

31. Zakikhany, K., C. R. Harrington, et al. (2010) "Unphosphorylated CsgD controls biofilm formation in Salmonella enterica serovar Typhimurium." *Mol Microbiol* **77**(3): 771-86.

32. Beyhan, S., K. Bilecen, et al. (2007). "Regulation of rugosity and biofilm formation in Vibrio cholerae: comparison of VpsT and VpsR regulons and epistasis analysis of vpsT, vpsR, and hapR." *J Bacteriol* **189**(2): 388-402.

33. Newell, P. D., C. D. Boyd, et al. "C-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage" (*submitted*).