# STRUCTURAL AND BIOCHEMICAL ELUCIDATION OF THE MECHANISM OF C-DI-GMP MEDIATED INSIDE-OUT SIGNALING CONTROLLING PERIPLASMIC PROTEOLYSIS.

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

Debashree Chatterjee

August 2013

# STRUCTURAL AND BIOCHEMICAL ELUCIDATION OF THE MECHANISM OF C-DI-GMP MEDIATED INSIDE-OUT SIGNALING CONTROLLING PERIPLASMIC PROTEOLYSIS.

DEBASHREE CHATTERJEE, Ph. D.

Cornell University 2013

Bacteria have developed multiple strategies to adapt to diverse ecological niches and hostile environments. One such strategy involves formation and maintenance of multicellular communities known as biofilms. In these microbial aggregates, sessile bacterial cells are encased in an extracellular matrix. It has now been established that c-di-GMP, a ubiquitous bacterial second messenger, is a central regulator of this developmental process in bacteria. It exerts its effects on transcriptional, translational and post-translational levels. While diguanylate cyclases and phosphodiesterases with conserved GGDEF and EAL (and HD-GYP) domains are responsible for the production and degradation of the dinucleotide, respectively, the receptors form a more diverse group with degenerate, catalytically inactive GGDEF-EAL domain-containing proteins representing a major subfamily. One such protein, LapD from *Pseudomonas fluorescens*, uses an inside-out signaling mechanism to relay intracellular c-di-GMP concentration to control the localization of an outer-membrane anchored large adhesin protein LapA at the cell surface, by sequestering a periplasmic cysteine protease, LapG. When free, LapG cleaves the N-terminus of LapA, releasing it from the cell surface and ultimately leading to biofilm dispersal.

Based on our structure-function analysis, here we propose a mechanism for the c-di-GMP-mediated, regulation of periplasmic proteolysis by LapD. We first elucidate the molecular basis of signal recognition and relay by *P. fluorescens* LapD and identify orthologous systems in multiple other bacteria including many pathogens such as *Legionella pneumophila*. This is followed by our work on *L. pneumophila* LapG, which provides us with the first atomic models of a bacterial protease of the DUF920 family and we are able to identify a highly conserved $Ca^{2+}$-binding motif integral to its function. We then characterize the LapD-ortholog CdgS9 from *L. pneumophila* which confirms a common molecular mechanism. The crystal structure of the periplasmic output domain module reveals novel conformations and sheds new light on the mode of activation of the receptor. We finally report the structure of a complex between $CdgS9^{output}$ and *P. fluorescens* LapG which led to the discovery of the pharmacologically relevant binding interface between the output domain and LapG.

# BIOGRAPHICAL SKETCH

The author was born in 1984 in West Bengal, India. She finished her schooling from Loreto Convent, Asansol and Assembly of God Church School, Sodepur. In the year of 2002, she was accepted into Lady Brabourne College affiliated under Calcutta University, to pursue a Bachelor's degree in Chemistry. After graduating with honors in 2005, she joined the Indian Institute of Technology, Delhi to accomplish a Master's degree in Chemistry, where she worked under the expert guidance of Dr. S.K. Khare and completed a Master's thesis titled, "Biochemical characterization of extracellular proteases and lipases from *Pseudomonas aeruginosa*". In 2007, she came to Cornell University as a graduate student in the field of Chemistry and Chemical biology, where she initially joined the research group of Dr. Tadhg P. Begley. Following his decision to relocate to Texas A&M University, she shifted to Dr. Holger Sondermann's laboratory to pursue her doctoral dissertation. During this fruitful odyssey she devoted her efforts towards structural and functional characterization of proteins integral to bacterial biofilm development.

Dedicated to my loving parents Debiprasad Chatterjee and Suvra Chatterjee

# ACKNOWLEDGMENTS

This thesis and my journey from a graduate student to a PhD candidate would not have been possible without the help and support of a large number of people and I will be forever indebted to them for their support.

First and foremost, I would like to thank my advisor Dr. Holger Sondermann for his constant encouragement, guidance and support, while granting me the freedom to work independently and at my own pace. During my five years of research in his laboratory, Holger has always been superbly patient and extremely approachable and has helped me become an objective research scientist in a truly constructive manner. I am grateful to him for providing me with a nurturing environment in lab, one that has helped me perform to my full potential. I would also like to thank my other committee members, Dr. Richard A. Cerione and Dr. Hening Lin, for their valuable input and constant support. I am also extremely grateful to my collaborator George O'Toole and his lab members Pete and Chelsea.

I am immensely thankful to everyone who has helped me in some way towards the advancement of my graduate research. Special thanks to Jeremiah Hanes, a former postdoctoral fellow in the Begley lab, from whom I learned the basics of protein purification and enzymology, and Marcos V.A.S. Navarro and Nabanita De, former members of the Sondermann lab, for helping me out in my initial days and for patiently teaching me the practical aspects of molecular biology, large-scale protein purification and macromolecular crystallography. In addition, it has always been a great pleasure to work with Marcos and my colleague Petya Krasteva on the LapD protein project and I sincerely thank them for their respective contributions. My experience in Holger's laboratory would not however have been this enjoyable without the presence of Laura Byrnes, Lynda Goh, Qi Wang, John O'Donnell, Krista

Giglio and Rick Cooley. Thanks a lot everyone!

Spending five years in Ithaca requires a special kind of dedication and support system and I am extremely glad that I was blessed with an awesome group of friends such as Anandarup, Soumya, Debashree, Snad, Anupam, Aritro, Deepti, Trina, Rachna, Siddarth, Anand and Amit. Ithaca will not be the same without them and thanks for being there all the time.

Of course, no amount of thanks would be enough for my husband, Parag, for being there during the best and especially the worst times.

Finally, I would love to thank my parents for their blessings and constant encouragement without which none of this would have ever been possible.

TABLE OF CONTENTS

*CHAPTER III*

**Structural characterization of a conserved, calcium-dependent periplasmic protease from *Legionella pneumophila***

LIST OF FIGURES

*CHAPTER I*



*CHAPTER II*

*CHAPTER III*

*CHAPTER IV*

LIST OF TABLES

# CHAPTER 1

# Introduction

## 1.1. Biofilms

Bacteria can exist either as independent, planktonic (free-swimming) cells or as members of surface anchored communities known as biofilms, which are ubiquitous not only in nature, but also in industrial and clinical settings (1). A biofilm is a densely populated microbial community in which the bacterial cells are embedded in a self-produced, extracellular matrix mainly composed of exopolysaccharides (EPS), proteins and nucleic acids (Fig.1.1). This matrix confers multiple advantages for the bacterial population. For example, it functions as a scaffold that holds the cells together, provides a medium for intercellular signaling as well as protects against environmental assaults, predators and antibiotic treatment (1, 9). In fact pathogenic, biofilm dwelling bacteria are often much less susceptible to antimicrobial agents than planktonic cells, mostly because the biofilm matrix tends to resist the penetration of the antibiotics into the embedded cells, rendering many commonly used antibiotics ineffective (2). These bacterial aggregates have also been associated with nosocomial infections, owing to their propensity to develop on various implants, like catheter lines and heart valves (3). Recently, it has been established that bacterial biofilms can also contribute to non-implant diseases like otitis media (inflammation of the middle ear), periodontitis (infection of gums and bones around the teeth) and cystic fibriosis (chronic lung disease in children and adolescents) (4-7). Since bacteria in biofilms are able to withstand antibiotic treatment, such infections and contaminations are

extremely difficult to eradicate. The magnitude of the problem is further amplified by the steady decline in the number of new FDA approved antimicrobial agents coupled with a rapid emergence of new, multi-drug resistant bacterial strains (8). Since bacterial biofilms have such a huge impact on human society, it has motivated a great deal of research to unravel the genetic, biochemical and physiological basis of this developmental process. We need to attain a fundamental understanding of the components and molecular mechanisms that initiate and maintain these bacterial communities, so that we are able to come up with effective strategies to manipulate these components and pathways, in order to prevent and/or disrupt bacterial biofilms.

Biofilm development in bacteria has been proposed to occur in a series of highly regulated steps (9, 10). This transition in lifestyle is triggered in response to environmental and physiological cues like bacterial cell density, nutrient availability, cellular stress etc. Firstly, flagellar or pilus-mediated motility might be required for the cells to approach a surface to initiate reversible or transient attachment. In this phase, the microbial cells associate with the surface mostly through the poles of the cell, as a result of which this interaction is relatively weak. A subpopulation of these reversibly attached bacteria can then become irreversibly attached through the longitudinal axis of the cell (11). This shift from reversible to irreversible attachment is often brought about by cell surface associated adhesion factors (12). It is typically also accompanied by a change in the transcription profile leading to an up-regulation of genes essential for biofilm development and maintenance (synthesis and secretion of EPS and proteinaceous appendages) and down-regulation of genes responsible for motility and virulence (9-12). This leads to the formation of a monolayer of cells which can then

form micro-colonies. Over time, these micro-colonies develop into macro-colonies and eventually a mature biofilm is formed. Like biofilm growth, biofilm dispersal is also a tightly controlled process and is often set off under conditions of limited nutrient availability or upon accumulation of toxic waste, to release the sessile cells into the environment to colonize new surfaces (90, 91). For many pathogenic bacteria, biofilm dispersal plays an important role in the transmission of bacteria from environmental reservoirs to human hosts, and in the aggravation and spread of infection within a host. The mechanistic details of bacterial biofilm dispersal are not yet well understood but it is an interesting area of research that might pave the way to the development of novel antimicrobial agents that can either inhibit formation or promote dispersal of biofilms.

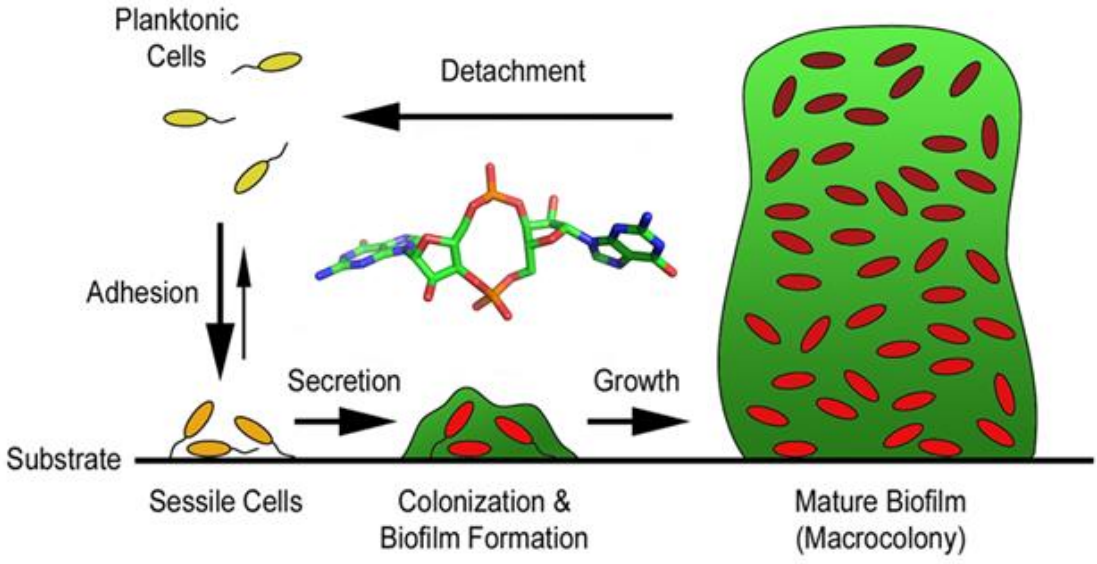



**Figure 1.1: Overview of biofilm formation** (figure adapted from [61]).
Bacteria can reversibly attach to surfaces followed by loss of motility and synthesis of exopolysaccharides and cellular adhesins. Mature biofilms are characterized by a complex architecture and collaborative behavior between functionally differentiated members. Finally, dispersal processes are governed by the secretion of glycolipid surfactants, proteases and nucleases, thereby leading to the release of highly motile planktonic cells, capable of colonizing new surfaces. Several of these processes are regulated by a bacterial second messenger, c-di-GMP (center).

## 1.2. Cyclic dimeric GMP (c-di-GMP)

The bacterial second messenger cyclic dimeric (3'→5') guanylic acid (cyclic di-GMP or c-di-GMP) is a monocyclic RNA dinucleotide that has emerged as a central molecule that orchestrates the transition in bacterial life-style between the motile, planktonic state and the sessile, biofilm state (13, 14, 15). Besides regulating biofilm development, c-di-GMP has also been shown to modulate a plethora of other cellular and physiological processes like motility, virulence, cell cycle progression and differentiation (13-18). It has recently been discovered that c-di-GMP is also recognized by mammalian immune systems as a uniquely bacterial molecule. Thus it is now being considered as a promising vaccine adjuvant (19-22).

Cyclic di-GMP was first discovered in 1987 by Benziman and colleagues, as an allosteric activator of *Gluconacetobacter xylinus* cellulose synthase (23). In the last 25 years since its discovery, enzymes for generation (diguanylate cyclases or DGCs) and degradation (phosphodiesterases or PDEs) of c-di-GMP have been identified in all major bacterial phyla, thereby giving it the status of an universal bacterial second messenger. In addition to the monomeric form, c-di-GMP can also form a stable dimer with stacked self-intercalated guanine units and both forms have been found in crystal structures of c-di-GMP metabolizing and binding proteins (24-33). It is generally accepted that once synthesized, c-di-GMP exists as a sequestered, rather than a general, diffusive signal (34, 35). Quantification of cytosolic c-di-GMP levels in several bacterial species indicates concentrations in the micromolar range or lower. Considering the fact that most c-di-GMP receptors and PDEs that have been identified

so far have lower dissoaciation constants for the dinucleotide, it is understood that cellular c-di-GMP exists primarily in a protein bound form (34, 36).

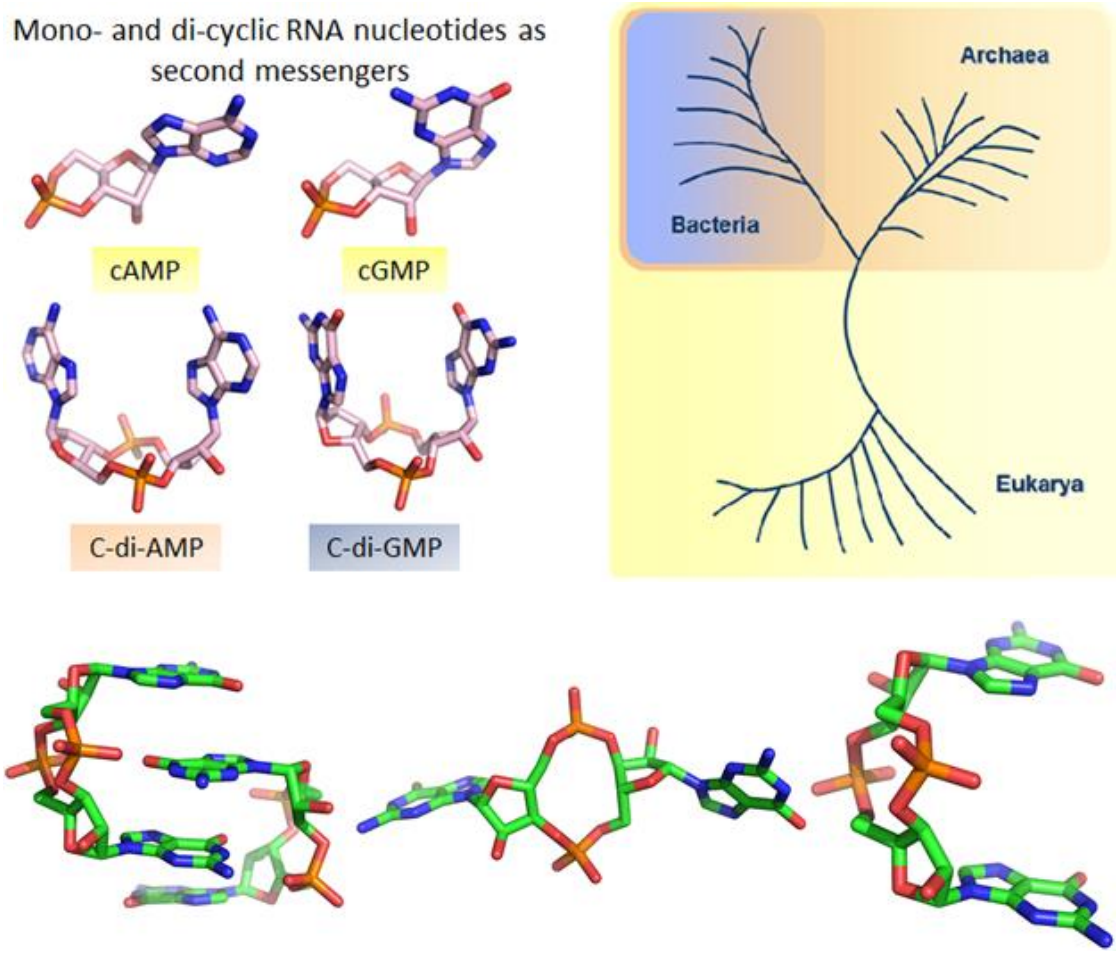



**Figure 1.2: Cyclic di-GMP as a bacterial second messenger.** (Figure adapted from [61] and [93]).
**(A)** Phylogenetic distribution of prevalent RNA-based second messengers: Signal transduction networks that rely on the monocyclic nucleotide second messenger cAMP and cGMP modulate a diverse range of 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 is utilized by both bacteria and archaea, while c-di-GMP is largely bacterial, but it has also been identified in the eukaryote *Dictyostelium*, a social amoeba.
**(B)** Several distinct conformations of protein bound c-di-GMP.

### 1.3. Cyclic di-GMP metabolism

Cyclic di-GMP is generated from two molecules of GTP by enzymes known as diguanylate cyclases (DGCs) harboring the GGDEF domain (18, 33, 37, 38, 40), and hydrolyzed into either linear-di-GMP (pGpG) or GMP by enzymes known as phosphodiesterases (PDEs), with EAL (39, 40) or HD-GYP domains (40-42) (Fig. 1.4 A-C). These c-di-GMP turnover domains are eponymous with the highly conserved, signature amino acid motifs of their active sites, and are often found in association with a wide variety of signal input domains implying that a broad range of environmental and cellular inputs can be integrated into the c-di-GMP signaling network. The GGDEF domain also often contains a c-di-GMP-binding site (I-site) distal to the active site, which is involved in an allosteric negative feedback regulation (product-inhibition). Genomic and bioinformatics analyses have shown that these c-di-GMP metabolizing domains can exist in nature either as stand-alone protein domains or in tandem as a part of the same polypeptide chain. The co-existence of two domains with opposing enzymatic activities in the same protein however represents an "enzymatic conundrum", raising interesting questions about their mode of regulation and final enzymatic output (43, 44).

Theoretically, three possibilities exist that might explain this "enzymatic conundrum". One scenario is where both domains are catalytically active, but they are differentially regulated by environmental or cellular signals such that only one activity is prevalent at any given point. Examples of such bifunctional enzymes include proteins BphG1 from *Rhodobacter sphaeroides*, with a PAS-GAF-PHY-GGDEF-EAL module (45), MSDGC-1 from *Mycobacterium smegmatis, with a* GAF-GGDEF-EAL

domain architecture (92) and Lpl0329 from *Legionella pneumophila*, which contains a REC-GGDEF-EAL module (46). A second scenario is possible where one of the two tandem domains is catalytically inactive (47, 48) and plays an allosteric role in modulating the output of the enzymatically active domain, either by binding (but not processing) the substrate, e.g. GTP-binding by inactive GGDEF domains (CC3396 protein from *Caulobacter crescentus;* 47) and c-di-GMP binding by inactive EAL domains (32, 49, 50), or by mediating protein-protein or protein-RNA interactions. Finally, a third scenario is possible where both the domains have completely lost their enzymatically activity but can function as receptors of c-di-GMP. The transmembrane protein LapD from *P. fluorescens* is an example of this type of tandem GGDEF-EAL domain proteins (51, 52).

GGDEF and EAL/HD-GYP domains are found to be among the most abundant protein domains encoded in bacterial genomes, which interestingly concurs with the ability of bacteria to survive and flourish in diverse ecological niches. The surprisingly large number of predicted c-di-GMP signaling proteins in some genomes (there are 43 in one of our study organisms, *P. fluorescens* Pf0-1) raises the possibility that c-di-GMP signaling networks are extraordinarily complex. Recent studies provide strong evidence that c-di-GMP signaling proteins (DGCs or PDEs) can impact distinct outputs through differential subcellular localization (49, 53-55). Other studies have shown that c-di-GMP signaling proteins in the same pathway physically interact with one another or their targets (56, 57). These data, combined with the well-established role for oligomerization in regulating DGC and PDE activities (37, 58, 59), points to protein-protein interactions and multi-protein complexes as a likely means for

compartmentalizing specific c-di-GMP signals. Understanding how c-di-GMP receptors participate in generation of biological outputs regulated by this nucleotide is essential to understanding a critical component of the bacterial lifestyle, that is, the transition between motile and sessile states.

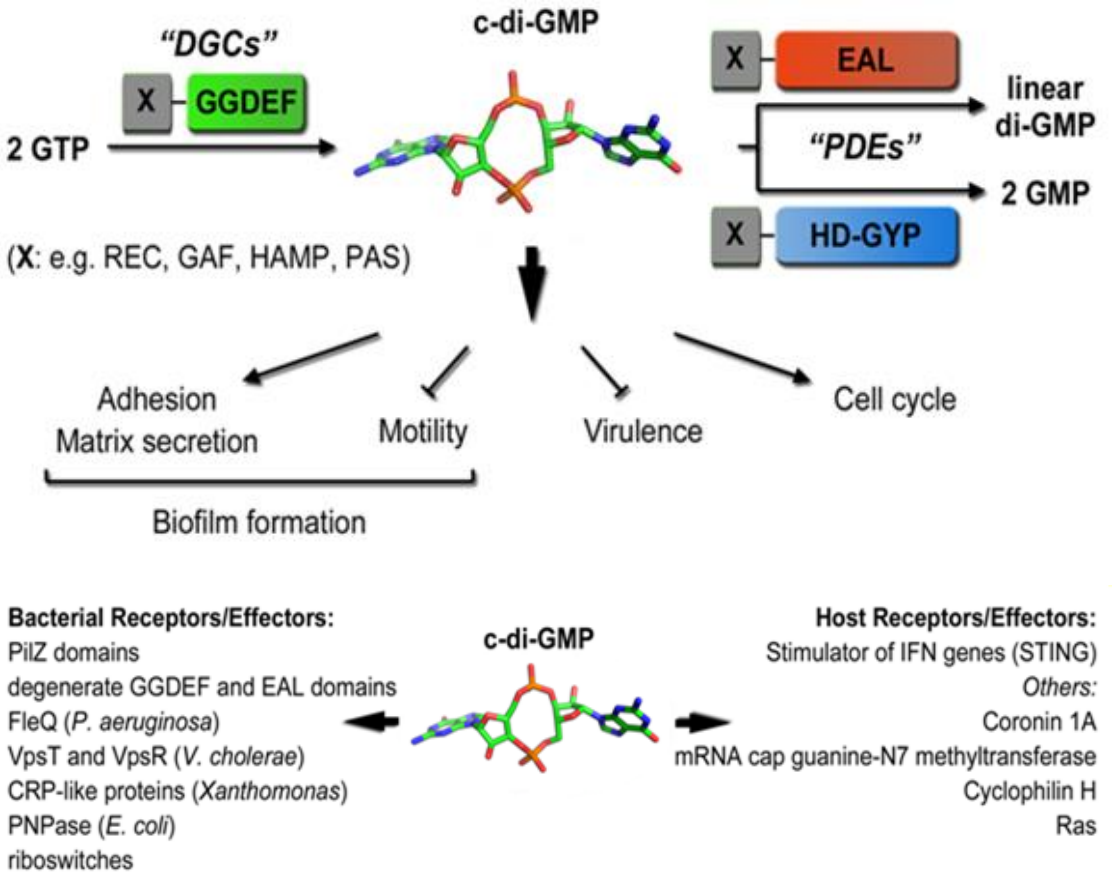



**Figure 1.3: Cyclic di-GMP signaling module.** (Figure adapted from [61]).
**(A)** Cyclic di-GMP signal generation, degradation and recognition: While protein domains catalyzing c-di-GMP synthesis and hydrolysis have been fairly well characterized, effectors of the dinucleotide are only beginning to emerge.
**(B)** Effectors or receptors of c-di-GMP: Modules and proteins for c-di-GMP sensing are diverse. While several bacterial effectors have been studied extensively, only a few well-validated targets on host cells have been discovered till date. Others await further characterization.

### 1.4. Bacterial c-di-GMP receptors

Unlike c-di-GMP turnover enzymes which are readily identifiable based on primary sequence information, c-di-GMP effector components are much more diverse and cannot be represented by any one single domain or c-di-GMP binding site. At present, although our knowledge about c-di-GMP receptors and their downstream targets is rather limited, this situation is rapidly changing and new classes of c-di-GMP effectors are beginning to emerge (as summarized in recent reviews [60-62]).

One of the most well characterized classes of c-di-GMP receptors comprises proteins containing PilZ domains (e.g. PA2960 from *P. aeruginosa*, YcgR from *Escherichia coli* [63], BcsA from *G. xylinus* [64], DgrA protein from *C. crescentus* [65], and PlzC and PlzD from *Vibrio cholerae* [66]). The PilZ domain can either occur as a separate protein domain or be present downstream to an EAL or GGDEF-EAL composite unit. Interestingly, the PilZ domain proteins can bind c-di-GMP either as a closed monomer (67) or as a self-intercalated dimer (30, 64) and can also adopt different oligomeric states from monomeric to tetrameric (30, 68, 69), implying potentially different modes of downstream signal propagation.

Enzymatically inactive GGDEF domains with their allosteric c-di-GMP-binding I-sites (RxxD motif) form a second class of c-di-GMP receptors (24, 70). Some examples of this category of receptors include the response regulator PopA from *C. crescentus*, the hybrid histidine kinase SgmT from *Myxococcus xanthus* and PelD, which regulates c-di-GMP-dependent polysaccharide biosynthesis in *P. aeruginosa* (71, 72). Likewise, catalytically incompetent EAL domains that have retained their ability to bind the dinucleotide can also function as c-di-GMP receptors, e.g. FimX, a

protein that governs twitching motility in *P. aeruginosa* (32), YkuI from *Bacillus subtilis* and LapD, an inner-membrane localized protein that modulates biofilm development in *P. fluorescens* in response to exogenous inorganic phosphate availability (73).

Bacterial transcriptional regulators that modulate gene expression upon binding cytosolic c-di-GMP constitute another interesting class of c-di-GMP receptors. Enhancer-binding protein FleQ from *P. aeruginosa* with an $AAA^{+}$/ATPase σ54-interaction domain, the matrix production and motility regulation protein VpsT from *V. cholerae,* which harbors a non-canonical receiver domain, and the CRP/FNR-type transcriptional activators from *Xanthomonas* and *Burkholderia* (57, 74–76), are a few examples of this category.

The discovery of c-di-GMP-specific riboswitches as a novel class of c-di-GMP receptors by Breaker and colleagues has further diversified the gamut of physiological outputs regulated by this prokaryotic dinucleotide signaling molecule. Riboswitches are RNA aptamers that undergo changes in their secondary structure conformation upon binding small molecule ligands thereby modulating the transcription and/or translation of downstream genes (77-81). Two different types of these short c-di-GMP-binding RNA molecules are currently known, c-di-GMP-I and c-di-GMP-II, the former being more prevalent in the bacterial kingdom.

### 1.5. Mammalian c-di-GMP receptors

Eukaryotic immune systems have previously been shown to recognize c-di-GMP as an exclusive bacterial molecule, but the molecular basis of immunostimulation by c-di-GMP has become clear with the identification of the transmembrane protein STING as a direct sensor of the dinucleotide signal (82, 83). STING binds c-di-GMP *in vitro* with a dissociation constant ($k_d$) of 5 μM and triggers generation of type I interferons by signaling through the TANK binding kinase 1 (TBK1) and the interferon regulatory transcription factor 3 (IRF3). Recently, a second mammalian c-di-GMP sensing protein has been identified, the DEAD-box helicase DDX41 (84) which appears to have a slightly higher affinity for dinucleotide binding. However, it is still not completely understood as to how this bacterial intracellular second messenger makes its way into the cytosol of macrophages or whether its concentration ever reaches micromolar levels.

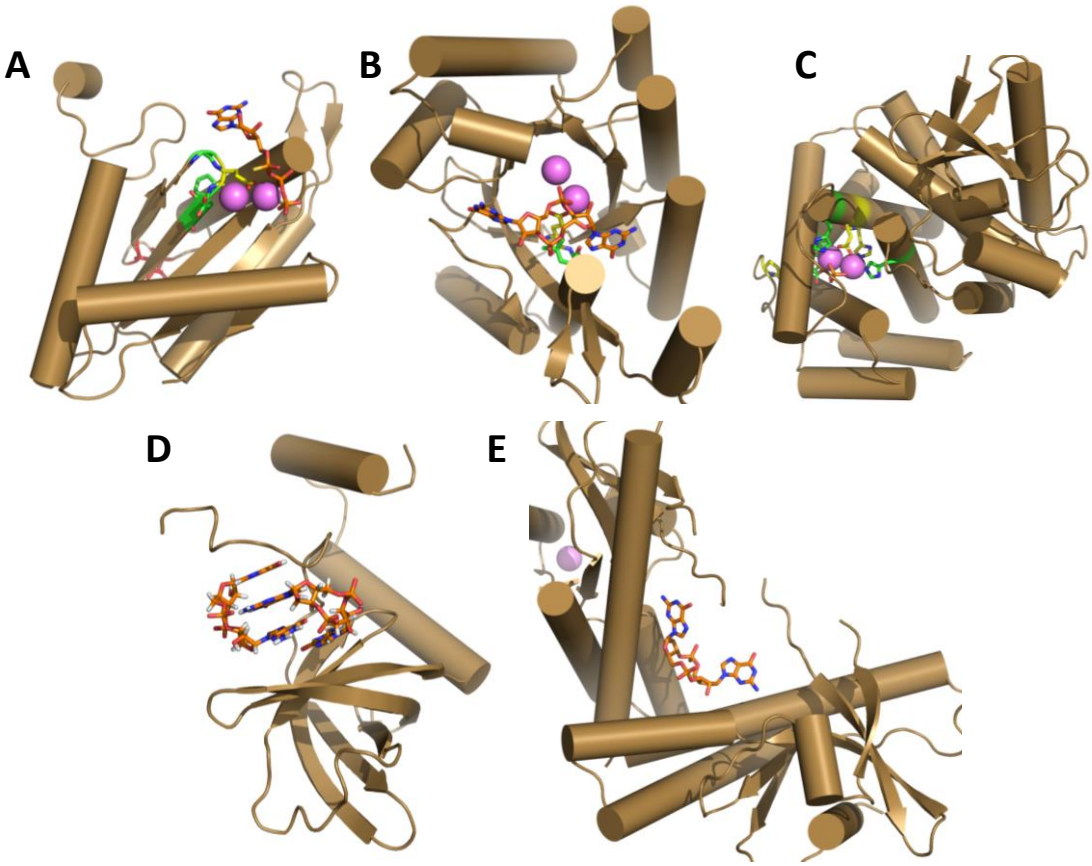



**Figure 1.4: Structural organization cyclic di-GMP-related proteins.** (Figure modified from [44]).
The upper row shows enzymes of c-di-GMP metabolism, and the lower row shows two types of c-di-GMP-binding proteins.
**(A)** GGDEF domain of PleD from *C. crescentus* with the bound substrate analog GTPαS. Bound $Mg^{2+}$ ions are shown as violet spheres. Active site glutamate is shown in yellow (86) (PDB entry 2V0N).
**(B)** EAL domain of Tbd1265 with bound c-di-GMP. Bound $Mg^{2+}$ ions are shown as violet spheres. Active site glutamate is shown in yellow (65) (PDB entry 3N3T).
**(C)** HD-GYP domain of Bd1817. Bound Fe atoms are shown as violet spheres. HDxxP motif is shown in yellow (129) (PDB entry 3TM8).
**(D)** C-di-GMP binding site of the PilZ domain of PA4608 (82) (PDB entry 2L74).
**(F)** C-di-GMP bound to the stimulator of interferon genes STING (74) (PDB entry 4EMT).

### 1.6. Protein targets of interest

Arguably, the most well understood role of a c-di-GMP-mediated signaling cascade in bacteria is in regulation of biofilm development. Cyclic di-GMP has been shown to impact biofilm growth in multiple layers including control at the transcriptional, translational and post-translational levels. There is evidence that c-di-GMP can regulate the various extracellular matrix components that contribute to biofilm structure, including biosynthesis and secretion of exopolysaccharides, proteinaceous fimbriae, adhesive pili and extracellular DNA (85).

Using gram-negative species of bacteria such as *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Legionella pneumophila* and *Vibrio cholerae*, we attempted to unravel the structural and mechanistic aspects of c-di-GMP mediated control of biofilm growth, maintenance and dispersal and its possible role in disease pathogenesis. Based on recent reports as well as personal communication with our collaborators, we chose several putative and known genes as targets of our studies. My efforts were primarily directed towards understanding the molecular basis of c-di-GMP mediated inside-out signaling and its role in regulating periplasmic proteolysis, a project I pursued in close collaboration with Prof. George O'Toole's group at Dartmouth.

Over the past decade the O'Toole lab has explored the mechanism whereby *P. fluorescens* forms a biofilm and studied how a key environmental nutrient, inorganic phosphate, regulates the formation of biofilms by this soil pseudomonad. We next wanted to use a combination of genetic, biochemical and structural studies to elucidate the mechanistic basis of this process.

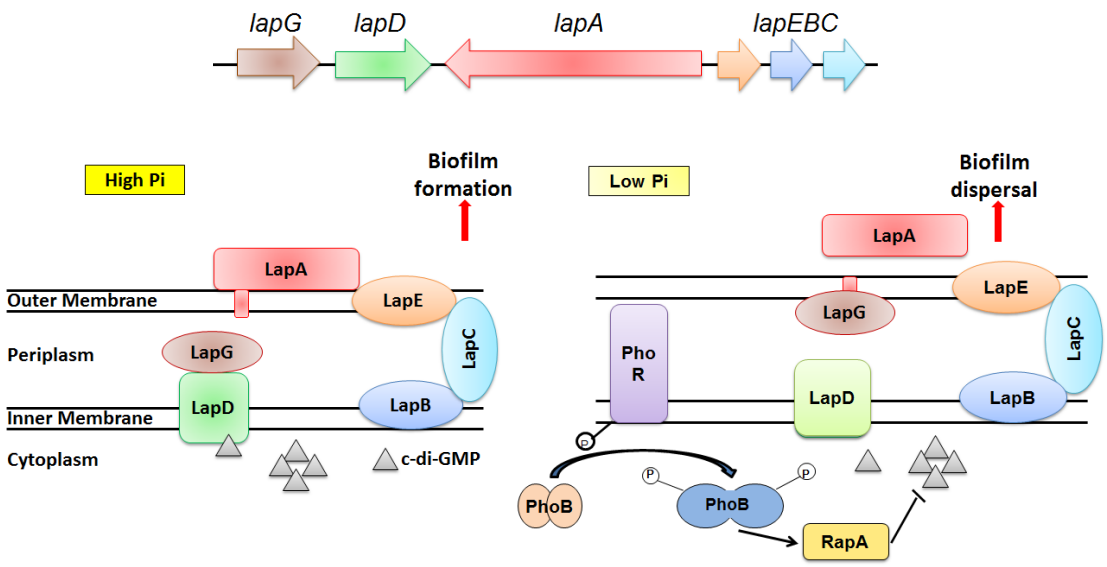



**Figure 1.5: Lap operon and LapD/LapG functional cycle.**
Current model summarizing how $P_i$ level controls biofilm formation by *P. fluorescens*. *Right*: Under conditions of low $P_i$, activation of the PhoR kinase leads to phosphorylation of PhoB. PhoB~P dimer binds to the Pho Box sequence upstream of the *rapA* gene, activating its transcription (86). The RapA protein cleaves c-di-GMP to form pGpG through its PDE activity, depleting cellular c-di-GMP pools, which in turn leads to the dissociation of the nucleotide from the c-di-GMP effector LapD. This signal promotes the egress of LapA to the culture supernatant. The loss of LapA requires the functional LapG periplasmic protease. In low $P_i$ conditions, when c-di-GMP is low, the periplasmic domain of LapD does not bind to LapG, and LapG is free to cleave the N-terminus of LapA. Cells cannot maintain stable surface attachments in low $P_i$ due to loss of LapA. This causes biofilm dispersal.
*Left*: Under high $P_i$ conditions RapA is not expressed, and c-di-GMP accumulates in the cell. LapD binds c-di-GMP and sends a signal across the inner membrane promoting the maintenance of LapA on the cell surface. LapD promotes LapA-mediated cell-surface adherence because it binds LapG and sequesters LapG from processing the N-terminus of LapA. Cells with full-length LapA on their cell-surface can form irreversible attachments to the substratum and form a biofilm (87).

One pathway c-di-GMP regulates biofilm development is through the cell attachment machinery of *P. fluorescens*, with the Lap operon at its center. This operon constitutes a set of genes which are absolutely essential for biofilm formation by *P. fluorescence* in response to nutrient availability. It consists of the large secreted protein LapA, an ABC transporter (LapEBC) for its delivery to the cell surface, a periplasmic protease LapG and the c-di-GMP receptor LapD (11, 51, 87-89). LapA, a large adhesin (predicted protein of ~520 kDa), is localized to the cell surface and plays a critical role in the establishment of biofilms by *P. fluorescens*, in particular, during an early step called "irreversible attachment", whereby bacterial cells attach themselves to the surface via their long axis. Secretion of this adhesin requires the LapEBC, an ATP binding cassette type transporter. The key environmental nutrient, inorganic phosphate ($P_i$) regulates biofilm formation via the well characterized Pst-PhoRB regulatory system. Pst-PhoRB controls biofilm formation in response to phosphate via the transcriptional control of RapA, a c-di-GMP-specific PDE, which when expressed, depletes intracellular levels of this signaling molecule (88, 89) (Fig. 1.5). Reduced cytosolic levels of c-di-GMP in turn results in the cleavage and release of LapA from the cell surface (88).

LapD is an inner membrane localized c-di-GMP-binding protein that controls biofilm formation by controlling localization of LapA on the cell surface in response to changing pools of c-di-GMP in the cytoplasm. LapD contains degenerate and enzymatically inactive GGDEF (DGC) and EAL (c-di-GMP-specific PDE) domains, and is the first protein shown to bind c-di-GMP through a degenerate EAL domain

(51). LapD mediates LapA localization in response to $P_i$-mediated changes in cytosolic c-di-GMP levels by regulating the activity of a periplasmic protease LapG (Fig 1.5).

LapG, which is encoded by a gene adjacent to *lapD*, is a periplasmically-localized cysteine protease required to proteolyze LapA between residues of a conserved Ala-Ala motif (residues 108-109), resulting in the release of LapA from the cell surface (87). Thus LapD utilizes a novel 'inside-out' signaling mechanism: binding c-di-GMP in the cytoplasm and communicating this signal to the periplasm via LapD's HAMP and periplasmic domains to connect environmental modulation of intracellular c-di-GMP levels by $P_i$ to regulation of LapA localization, and thus surface commitment by *P. fluorescens* (51).

In the following chapters we will present detailed structural and functional analyses of the transmembrane c-di-GMP sensing protein LapD, the periplasmic cysteine protease LapG and the LapG-LapD$^{output}$ complex. In addition, we will propose an array of future experiments to help further characterize c-di-GMP mediated pathways that regulate social behavior in bacteria.

In the first chapter, I contributed to our first attempts to elucidate the structure and regulation of LapD. In particular, as part of this work, I determined the first crystal structure of *P. fluorescens* LapD's periplasmic output domain. Through sequence conservation mapping onto the three-dimensional structure, I identified an invariable residue at the distal tip of the output domain that is critical for its interaction with LapG. I also established an *in vitro* binding assay to monitor LapG's interaction with the periplasmic output domain of LapD, and could show that the aforementioned conserved residue was crucial for the interaction. In the second chapter, I describe the

first structural characterization of a bacterial transglutaminase like cysteine protease (BTLCP). My work on *L. pneumophila* LapG has provided the first atomic models of a protease belonging to this family. I have also been able to identify a strictly conserved, calcium-binding motif in LapG that is crucial for its proteolysis activity. $Ca^{2+}$-binding by LapG has been validated and quantified in solution using isothermal titration calorimetry. I have also shown that *L. pneumophila* LapG can cleave the corresponding substrate from *P. fluorescens*, thereby suggesting a common mechanism of substrate recognition and proteolytic cleavage. In the third chapter of my thesis, I have described the structural analysis of the periplasmic output domain of the LapD ortholog from *L. pneumophila* as well as its complex with *P. fluorescens* LapG. On the basis of these structures, I have successfully identified the binding interface between these two proteins as well as the conformational changes induced by ligand binding. The final chapter summarizes my work and describes future goals.

References


1. Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat. Rev. Microbiol. 2:95–108.

2. Mah TF, O'Toole GA. 2001. Mechanisms of biofilm resistance to antimicrobial agents. Trends Microbiol. 9(1):34-9.

3. Høiby N, Ciofu O, Johansen HK, Song Z, Moser C, Jensen PØ, Molin S, Givskov M, Tolker-Nielsen T, Bjarnsholt T. 2011.The clinical impact of bacterial biofilms. Int. J. Oral Sci. 3(2):55-65.

4. Bakaletz LO. 2012. Bacterial biofilms in the upper airway - evidence for role in pathology and implications for treatment of otitis media. Paediatr. Respir. Rev. 13(3):154-9.

5. Schaudinn C, Gorur A, Keller D, Sedghizadeh PP, Costerton JW. 2009. Periodontitis: an archetypical biofilm disease. J. Am. Dent. Assoc. 140(8):978-86.

6. Høiby N, Ciofu O, Bjarnsholt T. 2010. Pseudomonas aeruginosa biofilms in cystic fibrosis. Future Microbiol. 5(11):1663-74.

7. Moreau-Marquis S, Stanton BA, O'Toole GA. Pseudomonas aeruginosa biofilm formation in the cystic fibriosis airway. 2008. Pulm. Pharmacol. Ther. 21(4):595-9.

8. Spellberg B, Powers JH, Brass EP, Miller LG, Edwards JE. 2004. Trends in antimicrobial drug development: Implications for the future. Antimicrobial Research and Development • CID. 38:1279–86.

9. Monds RD, O'Toole GA. 2009. The developmental model of microbial biofilms: ten years of a paradigm up for review. Trends Microbiol. 17(2):73-87.

10. Nadell CD, Xavier JB, Foster KR.The sociobiology of biofilms. 2009. FEMS Microbiol Rev. 33(1):206-24.

11. Hinsa SM, Espinosa-Urgel M, Ramos JL, O'Toole GA. 2003. Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. Molecular Microbiology. 49(4), 905–918.

12. Caiazza NC, O'Toole GA. 2004. SadB is required for the transition from reversible to irreversible attachment during biofilm formation by Pseudomonas aeruginosa PA14. J. Bacteriol. 186(14):4476-85.

13. D'Argenio DA, Miller SI. 2004. Cyclic di-GMP as a bacterial second messenger. Microbiology. 150:2497–2502.

14. Jenal U. 2004. Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria? Curr. Opin. Microbiol. 7:185–191.

15. Jenal U, Malone J. 2006.Mechanisms of cyclic-di-GMP signaling in bacteria. Annu. Rev. Genet. 40:385– 407.

16. Römling U, Gomelsky M, Galperin MY. 2005. c-di-GMP: the dawning of a novel bacterial signalling system. Mol. Microbiol. 57:629–639.

17. Römling U, Amikam D. 2006.Cyclic di-GMP as a second messenger. Curr. Opin. Microbiol. 9:218 –228.

18. Tamayo R, Pratt JT, Camilli A. 2007.Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. Annu. Rev. Microbiol. 61:131–148.

19. Ebensen T, Schulze K, Riese P, Morr M, Guzman CA. 2007. The bacterial second messenger cdiGMP exhibits promising activity as a mucosal adjuvant. Clin. Vaccine Immunol. 14:952–958.

20. Karaolis DK, Means TK, Yang D, Takahashi M, Yoshimura T, Muraille E, Philpott D, Schroeder JT, Hyodo M, Hayakawa Y, Talbot BG, Brouillette E, Malouin F. 2007. Bacterial c-di-GMP is an immunostimulatory molecule. J. Immunol. 178:2171–2181.

21. Karaolis DK, Newstead MW, Zeng X, Hyodo M, Hayakawa Y, Bhan U, Liang H, Standiford TJ. 2007. Cyclic di-GMP stimulates protective innate immunity in bacterial pneumonia. Infect. Immun.75:4942– 4950.

22. Ogunniyi AD, Paton JC, Kirby AC, McCullers JA, Cook J, Hyodo M, Hayakawa Y, Karaolis DK. 2008. c-di-GMP is an effective immunomodulator and vaccine adjuvant against pneumococcal infection. Vaccine 26:4676–4685.

23. Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R, Braun S, de Vroom E, van der Marel GA, van Boom JH, Benziman M. 1987.Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. Nature 325:279 –281.

24. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T. 2004.Structural basis of activity and allosteric control of diguanylate cyclase. Proc. Natl. Acad. Sci. U. S. A. 101:17084 –17089.

25. Barends TR, Hartmann E, Griese JJ, Beitlich T, Kirienko NV, Ryjenkov DA, Reinstein J, Shoeman RI, Gomelsky M, Schlichting I. 2009.Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. Nature 459:1015–1018.

26. Minasov G, Padavattan S, Shuvalova L, Brunzelle JS, Miller DJ, Basle A, Massa C, Collart FR, Schirmer T, Anderson WF. 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. 284:13174 –13184.

27. Tchigvintsev A, Xu X, Singer A, Chang C, Brown G, Proudfoot M, Cui H, Flick R, Anderson WF, Joachimiak A, Galperin MY, Savchenko A, Yakunin AF. 2010. Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases. J. Mol. Biol. 402:524–538.

28. Smith KD, Lipchock SV, Ames TD, Wang J, Breaker RR, Strobel SA. 2009. Structural basis of ligand binding by a c-di-GMP riboswitch. Nat. Struct. Mol.Biol.16:1218 –1223.

29. Habazettl J, Allan MG, Jenal U, Grzesiek S. 2011. Solution structure of the PilZ domain protein PA4608 complex with cyclic di-GMP identifies charge clustering as molecular readout. J. Biol. Chem. 286:14304 –14314.

30. Ko J, Ryu KS, Kim H, Shin JS, Lee JO, Cheong C, Choi BS. 2010. Structure of PP4397 reveals the molecular basis for different c-di-GMP binding modes by Pilz domain proteins. J. Mol. Biol. 398:97–110.

31. Krasteva PV, Fong JC, Shikuma NJ, Beyhan S, Navarro MV, Yildiz FH, Sondermann H. 2010. *Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. Science 327:866–868.

32. Navarro MV, De N, Bae N, Wang Q, Sondermann H. 2009.Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX. Structure 17:1104 –1116.

33. Wassmann P, Chan C, Paul R, Beck A, Heerklotz H, Jenal U, Schirmer T. 2007. Structure of $BeF_3^-$-modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. Structure 15:915–927.

34. Weinhouse H, Sapir S, Amikam D, Shilo Y, Volman G, Ohana P, Benziman M. 1997. c-di-GMP-binding protein, a new factor regulating cellulose synthesis in Acetobacter xylinum. FEBS Lett. 416(2):207-11.

35. Jenal U, Malone J. 2006. Mechanisms of cyclic-di-GMP signaling in bacteria. Annu, Rev. Genet. 40:385-407.

36. Simm R, Morr M, Kader A, Nimtz M, Römling U. 2004. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. Mol Microbiol. 53(4):1123-34.

37. De N, Pirruccello M, Krasteva PV, Bae N, Raghavan RV, Sondermann H. 2008. Phosphorylation-independent regulation of the diguanylate cyclase WspR. PLoS Biol. 6(3):e67.

38. De N, Navarro MV, Raghavan RV, Sondermann H. 2009. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. J Mol Biol. 393(3):619-33.

39. Wu Q, Gardner KH. 2009. Structure and insight into blue light-induced changes in the BlrP1 BLUF domain. Biochemistry. 48(12):2620-9.

40. Schirmer T, Jenal U. 2009. Structural and mechanistic determinants of c-di-GMP signaling. Nat. Rev. Microbiol. 724-35.

41. Dow JM, Fouhy Y, Lucey JF, Ryan RP. 2006. The HD-GYP domain, cyclic di-GMP signaling, and bacterial virulence to plants. Mol Plant Microbe Interact. 19(12):1378-84.

42. Lovering AL, Capeness MJ, Lambert C, Hobley L, Sockett RE. 2011. The structure of an unconventional HD-GYP protein from Bdellovibrio reveals the roles of conserved residues in this class of cyclic-di-GMP phosphodiesterases. MBio. 2(5).

43. Seshasayee AS, Fraser GM, Luscombe NM. 2010. Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity. Nucleic Acids Res. 38(18):5970-81.

44. Römling U, Galperin MY, Gomelsky M. 2013. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol. Mol. Biol. Rev. 77(1):1-52.

45. Tarutina M, Ryjenkov DA, Gomelsky M. 2006. An unorthodox bacteriophytochrome from *Rhodobacter sphaeroides* involved in turnover of the second messenger c-di-GMP. J. Biol. Chem. 281:34751–34758.

46. Levet-Paulo M, Lazzaroni JC, Gilbert C, Atlan D, Doublet P, Vianney A. 2011. The atypical two-component sensor kinase Lpl0330 from *Legionella pneumophila* controls the bifunctional diguanylate cyclasephosphodiesterase Lpl0329 to modulate bis(3',5')- cyclic dimeric GMP synthesis. J. Biol. Chem. 286:31136 –31144.

47. Christen M, Christen B, Folcher M, Schauerte A, Jenal U. 2005.Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. J. Biol. Chem. 280: 30829–30837.

48. Schmidt AJ, Ryjenkov DA, Gomelsky M. 2005. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. J. Bacteriol. 187:4774–4781.

49. Kazmierczak BI, Lebron MB, Murray TS. 2006. Analysis of FimX, a phosphodiesterase that governs twitching motility in *Pseudomonas aeruginosa*. Mol. Microbiol. 60:1026 –1043.

50. Qi Y, Chuah ML, Dong X, Xie K, Luo Z, Tang K, Liang ZX. 2011. Binding of cyclic diguanylate in the non-catalytic EAL domain of FimX induces a long-range conformational change. J. Biol. Chem. 286:2910–2917.

51. Newell PD, Monds RD, O'Toole GA. 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. 106:3461–3466.

52. Navarro MV, Newell PD, Krasteva PV, Chatterjee D, Madden DR, O'Toole GA, Sondermann H. 2011. Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. PLoS Biol. 9(2):e1000588.

53. Guvener, Z. T., and C. S. Harwood. 2007. Subcellular location characteristics of the *Pseudomonas aeruginosa* GGDEF protein, WspR, indicate that it produces cyclic-di-GMP in response to growth on surfaces. Mol Microbiol 66:1459-73.

54. Merritt JH, D.-G. Ha, K. N. Kowles, W. Lu, D. K. Morales, J. D. Rabinowitz, Z. Gitai, and G. A.O'Toole. 2010. Specific control of *Pseudomonas aeruginosa* surface-associated behaviorts by two cdi-GMP digunaylate cyclases. mBio 1:e00183-10.

55. Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B, and U. Jenal. 2004. Cell cycle dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. Genes & Dev 18:715-727.

56. Bobrov AG, Kirillina O, Forman S, Mack D, and Perry RD. 2008. Insights into *Yersinia pestis* biofilm development: topology and co-interaction of Hms inner membrane proteins involved in exopolysaccharide production. Environ Microbiol 10:1419-32.

57. Chin KH, Lee YC, Tu ZL, Chen CH, Tseng YH, Yang JM, Ryan RP, Y. McCarthy JM, Dow A, Wang H, and Chou SH. 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 396:646-62.

58. Barends TR, Hartmann E, Griese JJ, Beitlich T, Kirienko NV, Ryjenkov DA, Reinstein J, Shoeman RL, Gomelsky M, and Schlichting I. 2009. Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. Nature 459:1015-8.

59. Paul R, Abel S, Wassmann P, Beck A, Heerklotz H, and Jenal U. 2007. Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. J Biol Chem 282:29170-7.

60. Sondermann H, Shikuma NJ, Yildiz FH. 2012. You've come a long way: c-di-GMP signaling. Curr. Opin. Microbiol. 15:140 –146.

61. Krasteva PV, Giglio KM, Sondermann H. 2012. Sensing the messenger: the diverse ways that bacteria signal through c-di-GMP. Protein Sci. 21: 929–948.

62. Ryan RP, Tolker-Nielsen T, Dow JM. 2012. When the PilZ don't work: effectors for cyclic di-GMP action in bacteria. Trends Microbiol. 20:235–242.

63. Alm RA, Bodero AJ, Free PD, Mattick JS. 1996. Identification of a novel gene, *pilZ*, essential for type 4 fimbrial biogenesis in *Pseudomonas aeruginosa*. J. Bacteriol. 178:46 –53.

64. Ryjenkov DA, Simm R, Römling U, Gomelsky M. 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. 281:30310 –30314.

65. Christen M, Christen B, Allan MG, Folcher M, Jeno P, Grzesiek S, Jenal U. 2007. DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*. Proc. Natl. Acad. Sci. U.S.A. 104:4112– 4117.

66. Pratt JT, Tamayo R, Tischler AD, Camilli A. 2007. PilZ domain proteins bind cyclic diguanylate and regulate diverse processes in *Vibrio cholerae*. J. Biol. Chem. 282:12860 –12870.

67. Benach J, Swaminathan SS, Tamayo R, Handelman SK, Folta-Stogniew E, Ramos JE, Forouhar F, Neely H, Seetharaman J, Camilli A, Hunt JF. 2007. The structural basis of cyclic diguanylate signal transduction by PilZ domains. EMBO J. 26:5153–5166.

68. Gentner M, Allan MG, Zaehringer F, Schirmer T, Grzesiek S. 2012. Oligomer formation of the bacterial second messenger c-di-GMP: reaction rates and equilibrium constants indicate a monomeric state at physiological concentrations. J. Am. Chem. Soc. 134:1019 –1029.

69. Li TN, Chin KH, Fung KM, Yang MT, Wang AH, Chou SH. 2011. A novel tetrameric PilZ domain structure from xanthomonads. PLoS One.6:e22036.

70. Christen B, Christen M, Paul R, Schmid F, Folcher M, Jenoe P, Meuwly M, Jenal U. 2006. Allosteric control of cyclic di-GMP signaling. J. Biol. Chem. 281:32015–32024.

71. Petters T, Zhang X, Nesper J, Treuner-Lange A, Gomez-Santos N, Hoppert M, Jenal U, Søgaard-Andersen L. 2012. The orphan histidine protein kinase SgmT is a c-di-GMP receptor and regulates composition of the extracellular matrix together with the orphan DNA binding response regulator DigR in *Myxococcus xanthus*. Mol. Microbiol. 84:147–165.

72. Whitney JC, Colvin KM, Marmont LS, Robinson H, Parsek MR, Howell PL. 2012. Structure of the cytoplasmic region of PelD, a degenerate diguanylate cyclase receptor that regulates exopolysaccharide production in *Pseudomonas aeruginosa*. J. Biol. Chem. 287:23582–23593.

73. Minasov G, Padavattan S, Shuvalova L, Brunzelle JS, Miller DJ, Basle A, Massa C, Collart FR, Schirmer T, Anderson WF. 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. 284:13174 –13184.

74. Fazli M, O'Connell A, Nilsson M, Niehaus K, Dow JM, Givskov M, Ryan RP, Tolker-Nielsen T. 2011. The CRP/FNR family protein Bcam1349 is a c-di-GMP effector that regulates biofilm formation in the respiratory pathogen *Burkholderia cenocepacia*. Mol. Microbiol. 82:327–341.

75. Gomelsky M. 2009. Cyclic-di-GMP-binding CRP-like protein: a spectacular new role for a veteran signal transduction actor. J. Bacteriol. 191:6785–6787.

76. Leduc JL, Roberts GP. 2009. Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of *Xanthomonas axonopodis* pv. *citri*. J. Bacteriol.191:7121–7122.

77. Barrick JE, Breaker RR. 2007. The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol. 8:R239.

78. Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR. 2008. Riboswitches in eubacteria sense the second messenger cyclic-di-GMP. Science 321:411– 413.

79. Kulshina N, Baird NJ, Ferre-D'Amare AR. 2009. Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. Nat. Struct. Mol. Biol. 16:1212–1217.

80. Smith KD, Lipchock SV, Ames TD, Wang J, Breaker RR, Strobel SA. 2009. Structural basis of ligand binding by a c-di-GMP riboswitch. Nat. Struct. Mol. Biol. 16:1218 –1223.

81. Smith KD, Shanahan CA, Moore EL, Simon AC, Strobel SA. 2011. Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. Proc. Natl. Acad. Sci. U. S. A. 108:7757–7762.

82. Burdette DL, Monroe KM, Sotelo-Troha K, Iwig JS, Eckert B, Hyodo M, Hayakawa Y, Vance RE. 2011. STING is a direct innate immunesensor of cyclic di-GMP. Nature 478:515–518.

83. Huang YH, Liu XY, Du XX, Jiang ZF, Su XD. 2012. The structural basis for the sensing and binding of cyclic di-GMP by STING. Nat. Struct. Mol. Biol. 19:728 –730.

84. Parvatiyar K, Zhang Z, Teles RM, Ouyang S, Jiang Y, Iyer SS, Zaver SA, Schenk M, Zeng S, Zhong W, Liu ZJ, Modlin RL, Liu YJ, Cheng G. 2012. The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. Nat. Immunol. 13:1155–1161.

85. Römling U. 2012. Cyclic di-GMP, an established secondary messenger still speeding up. Environ. Microbiol. 14:1817–1829.

86. Tiscler AD, Camilli A. 2004. Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation. Mol. Microbiol. 53:857-869.

87. Newell PD, Boyd CD, Sondermann H, O'Toole GA. 2011. c-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage. PLoS Biol 9(2):e1000587.

88. Monds RD, Newell PD, Gross RH, O'Toole GA. 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 63:656-79.

89. Monds RD, Newell PD, Schwartzman JA, O'Toole GA. 2006. Conservation of the Pho regulon in *Pseudomonas fluorescens* Pf0-1. Appl Environ Microbiol 72:1910-24.

90. Kaplan JB. 2010. Biofilm dispersal: mechanisms, clinical implications, and potential therapeutic uses. J Dent Res. 89(3):205-18.

91. McDougald D, Rice SA, Barraud N, Steinberg PD, Kjelleberg 2011. S. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. Nat. Rev. Microbiol. 10(1):39-50.

92. Bharati BK, Sharma IM, Kasetty S, Kumar M, Mukherjee R, Chatterji D. 2012. A full-length bifunctional protein involved in c-di-GMP turnover is required for long term survival under nutrient starvation in *Mycobacterium smegmatis*. Microbiology 158:1415–1427.

93. "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 V. Krasteva, August 2010.

# CHAPTER 2

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

### 2.1. Abstract

The bacterial second messenger c-di-GMP has emerged as a central regulator of biofilm formation. Increased cellular c-di-GMP levels lead to stable cell attachment, which in *Pseudomonas fluorescens* requires the widely conserved, transmembrane receptor LapD. C-di-GMP binding to its degenerate phosphodiesterase domain is communicated via a HAMP relay to the periplasmic domain, triggering sequestration of the protease LapG, thus preventing cleavage of the surface adhesin LapA. Here, we elucidate the molecular mechanism of autoinhibition and activation of LapD. In the absence of c-di-GMP, the intracellular module assumes an inactive conformation that is incompetent for dinucleotide binding. C-di-GMP binding to the phosphodiesterase domain disrupts the inactive state, forming a dimer interface between adjacent phosphodiesterase domains via interactions conserved in c-di-GMP-degrading enzymes. Efficient mechanical coupling of the conformational changes across the membrane is realized through an extensively domain-swapped, unique periplasmic fold. Our analyses identified a conserved system for the regulation of periplasmic proteases in a wide variety of bacteria, including many free-living and pathogenic species.

*Reproduced with permission from [Marcos V.A.S. Navarro, Peter D. Newell, Petya V. Krasteva, Debashree Chatterjee, George A. O'Toole, and Holger Sondermann, (2011). PLos Biology, 9(2):e1000588]. doi:10.1371/journal.pbio.1000588.

## 2.2. Introduction

Bacterial biofilms arise from planktonic microbial cells that attach to surfaces and form sessile multicellular communities, a process relevant to their 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 events including changes in motility, cell adhesion, and secretion are among the many processes driving bacterial biofilm formation. Such a plethora of physiological responses inevitably poses 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 containing either EAL or HD-GYP protein domains (7–10). The majority of cellular c-di-GMP appears to be bound to protein, eliciting localized, rather than diffusive signals (5). To date, only a few c-di-GMP receptors have been identified. Protein domains involved in c-di-GMP signal recognition include PilZ domains (12, 13), a non-canonical receiver domain in VpsT of *Vibrio cholerae* (14), the AAA σ54 interaction domain–containing transcription factor FleQ of *Pseudomonas aeruginosa* (15), and the cyclic nucleotide monophosphate–binding domain in Clp of

*Xanthomonas campestris* (16). In other cases, c-di-GMP turnover domains can also serve as sensors for the dinucleotide. For example, in GGDEF domain–containing proteins, an RxxD motif can serve as a c-di-GMP-binding inhibitory site either to regulate the activity of active enzymes (e.g., PleD of *Caulobacter crescentus* and WspR of *P. aeruginosa*) (17, 18) or to mediate protein–protein interactions in degenerate homologs (e.g., PelD of *P. aeruginosa* and CdgG of *V. cholerae*) (19, 20).

Bacterial proteins that mediate c-di-GMP turnover and signal transduction are often composed of multiple domains, allowing for a variety of regulatory inputs, signaling events, and/or physiological responses (21). For example, a large number of these 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 exhibits a regulatory function; and proteins in which both domains are degenerate and likely to work as c-di-GMP receptors (22, 23). Despite the frequent occurrence of this signaling module in bacterial genomes, 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 (24). LapD is required for stable cell attachment and biofilm formation in *P. fluorescens* and *P. putida* (25–27). It responds to changes in cellular c-di-GMP levels modulated by the availability of inorganic phosphate, an essential nutrient that is limiting in many ecosystems (24, 28).

Under phosphate starvation conditions, the expression of the phosphodiesterase RapA is upregulated, reducing cellular c-di-GMP levels and cell attachment. Increased phosphate availability yields an inactive Pho regulon, reduced RapA expression, and, as a consequence, a rise in cellular c-di-GMP concentration. As c-di-GMP levels change LapD switches between two states: the dinucleotide-unbound "off" state that retards stable biofilm formation by facilitating the secretion of the cell surface adhesin LapA, and the c-di-GMP-bound "on" state that supports cell adhesion by preventing the release of LapA from the outer membrane (24, 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 (24).

Recent work by Newell et al. (29) reveals the complete c-di-GMP signaling circuit by which LapD controls cell attachment in response to phosphate availability. 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 for 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 dinucleotide binding. However, it has remained unclear as to what prevents LapD from adopting an active conformation and how dinucleotide binding translates into an output signal.

Here, we present three crystal structures of LapD from *P. fluorescens* that provide models for the c-di-GMP-unbound cytoplasmic domain lacking only the HAMP domain, a c-di-GMP-bound EAL domain dimer, and the periplasmic domain.

Together these structures span almost the entire receptor and elucidate molecular mechanisms that regulate LapD function. The crystal structure of the cytoplasmic module containing the GGDEF–EAL tandem domains reveals the presence of an autoinhibitory motif formed by a helical extension of the HAMP domain. In this inactive state, the GGDEF domain restricts dinucleotide access to the EAL domain module. The crystal structure of dimeric, c-di-GMP-bound EAL domains provides insight into the conformational changes resulting from dinucleotide 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 LapD activity in inside-out signal transduction. Finally, our structural studies highlight many conserved features that allow us to identify similar signaling systems in a variety of bacterial strains including common pathogens such as *V. cholerae* and *Legionella pneumophila*.

## 2.3. Results and Discussion

### *Inactive State of the Intracellular Module of LapD*

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. 2.1). Based on secondary structure predictions, the linker forms a continuation of the second HAMP domain helix (Fig. 2.2). We will refer to this motif as the signaling helix (S helix). The structure of $LapD^{dual}$ (space group $P3_2$, one molecule in the asymmetric unit) was solved by single-wavelength anomalous dispersion phasing using selenomethione-substituted protein crystals (Table 2.1). We also obtained a second crystal form involving different crystal packing contacts (space group *I23*, one molecule in the asymmetric unit), yet the overall structure of $LapD^{dual}$ in the two crystals is identical (root mean square deviation [rmsd] of 0.9 Å over all atoms; Fig. 2.3A and 2.3C). In both cases, the biologically significant unit was predicted to be a monomer (33).

$LapD^{dual}$ adopts a compact, bilobal conformation (Fig. 2.1A). The GGDEF domain, comprising the N-terminal lobe, caps the dinucleotide-binding pocket of the EAL domain, which forms the C-terminal lobe of the tandem domain structure. The EAL domain buttresses the N-terminal S helix via predominantly hydrophobic interactions, burying 1,170 $Å^2$ (Fig. 2.1). The binding groove on the EAL domain, which accommodates the S helix, consists of the helix α6 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 and catalysis

(34, 40). In LapD, the consensus sequence of the loop determined for active phosphodiesterases is not conserved (40). This loop was referred to as loop 6 in SadR/RocR (40) and β5-α5 loop in the light-regulated phosphodiesterase BlrP1 (34). We will refer to this motif as the switch loop of LapD.

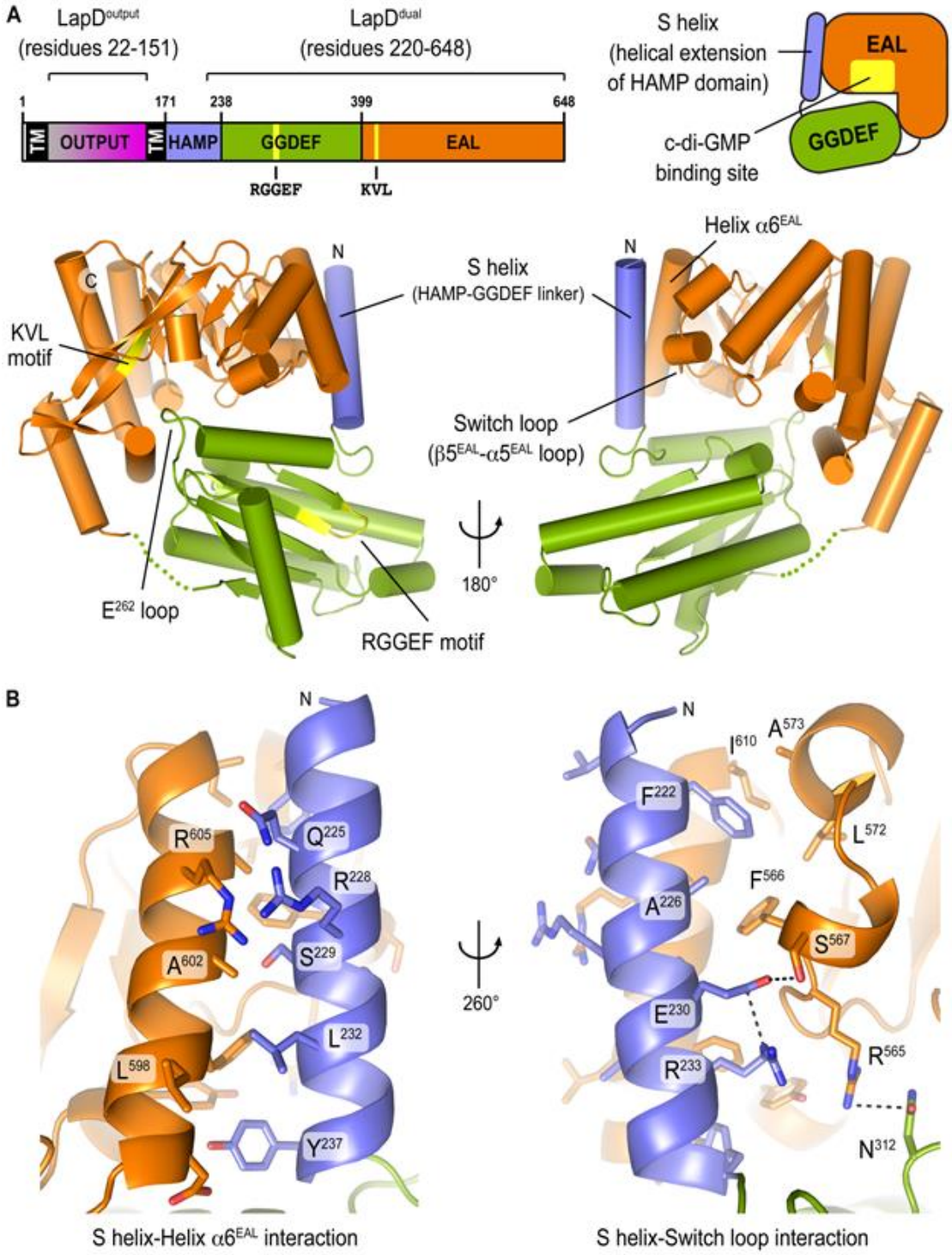
In addition to the S helix–EAL domain interaction, the GGDEF domain contacts the dinucleotide-binding surface of the EAL domain at multiple points, forming a loosely packed interface that buries 1,620 $Å^2$ of surface area (Fig. 2.1A, 2.4A, and 2.4B). One such contact, the salt bridge between an arginine residue ($R^{450}$) and a glutamate residue ($E^{262}$), forms a particularly close interaction (Fig. 2.4A). $R^{450}$ is located just downstream of the signature EAL motif (KVL in LapD) at the center of the c-di-GMP-binding site. $E^{262}$ is presented by a loop of the GGDEF domain. While $E^{262}$ directly occupies the dinucleotide-binding site, the loop itself is located at its periphery, partially blocking access of c-di-GMP to the EAL domain (Fig. 2.4B). Although the conformation of apo-LapD observed in the crystal structure is incompatible with c-di-GMP binding, the binding site is not completely occluded (Fig. 2.4B), and there may be a sufficient proportion of accessible EAL domains in solution to respond to increasing c-di-GMP concentrations, competing with the inhibitory interactions. In addition, there may be cooperative effects within the dimeric, full-length receptor that are not apparent from the structures of the isolated domains.

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. 2.4C). 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: $D^{239}$ in the loop and $R^{316}$ in the GGDEF domain of LapD, $D^{174}$ and $R^{249}$ in WspR, and $D^{292}$ and $R^{366}$ in PleD (17, 18, 38, 41). This interaction likely constrains the loop conformation, restricting the overall rotational freedom of the GGDEF domain relative to its associated regulatory module, the S helix in the case of LapD and the response receiver domain in the case of PleD and WspR.

In summary, the structural analysis of the cytoplasmic domain of LapD reveals that in the absence of c-di-GMP, the protein resides in a conformation incompatible with dinucleotide binding, with the GGDEF domain restricting access of c-di-GMP to the EAL domain. Dinucleotide binding would be accompanied by a major conformational change disrupting the conformation observed in the crystal structure.

**Figure 2.1: 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 domain diagrams (upper panel). The S helix forms an extension of the second HAMP domain helix. The switch loop is sensitive to the nucleotide-binding state of the EAL domain and is involved in dimerization and catalysis in active phosphodiesterases. Two views, separated by a 180° rotation, are shown.
**(B)** The S helix–EAL domain interface. A close-up view of the S helix–EAL domain interface is shown, with residues involved in direct, pairwise interactions shown as sticks. Two views, separated by a 260° rotation, are shown. Helix α6 and the switch loop form a surface buttressing the S helix.

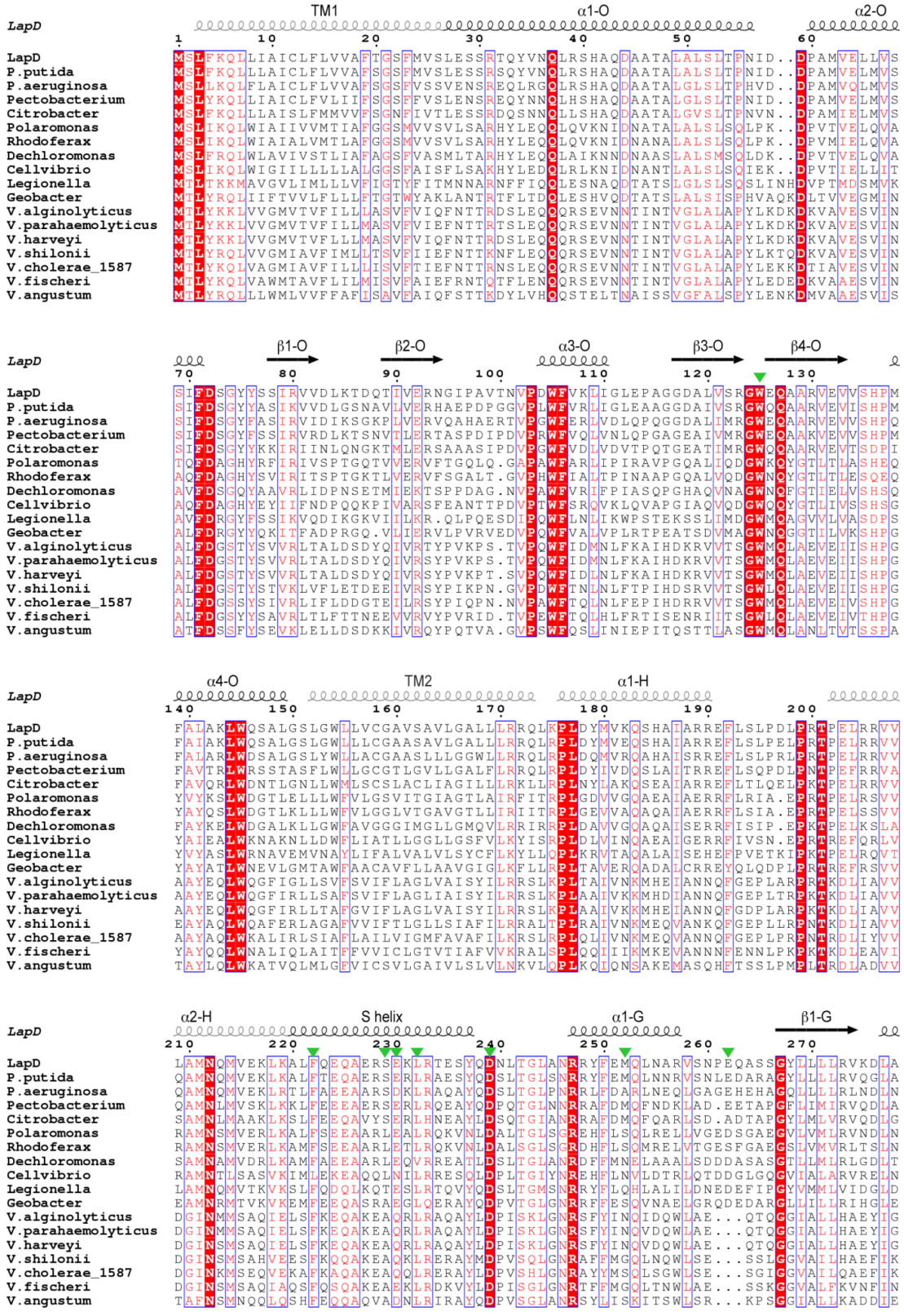

A
LapD^output
(residues 22-151)
LapD^dual
(residues 220-648)
1
171
238
399
648
TM
OUTPUT
TM
HAMP
GGDEF
EAL
RGGEF
KVL
S helix
(helical extension
of HAMP domain)
EAL
GGDEF
c-di-GMP
binding site
C
N
KVL
motif
S helix
(HAMP-GGDEF linker)
Helix α6^EAL
N
Switch loop
(β5^EAL-α5^EAL loop)
180°
E^262 loop
RGGEF motif
B
N
R^605
Q^225
R^228
S^229
A^602
L^232
L^598
Y^237
260°
N
I^610
A^573
F^222
L^572
F^566
A^226
S^567
E^230
R^565
R^233
N^312
S helix-Helix α6^EAL interaction
S helix-Switch loop interaction

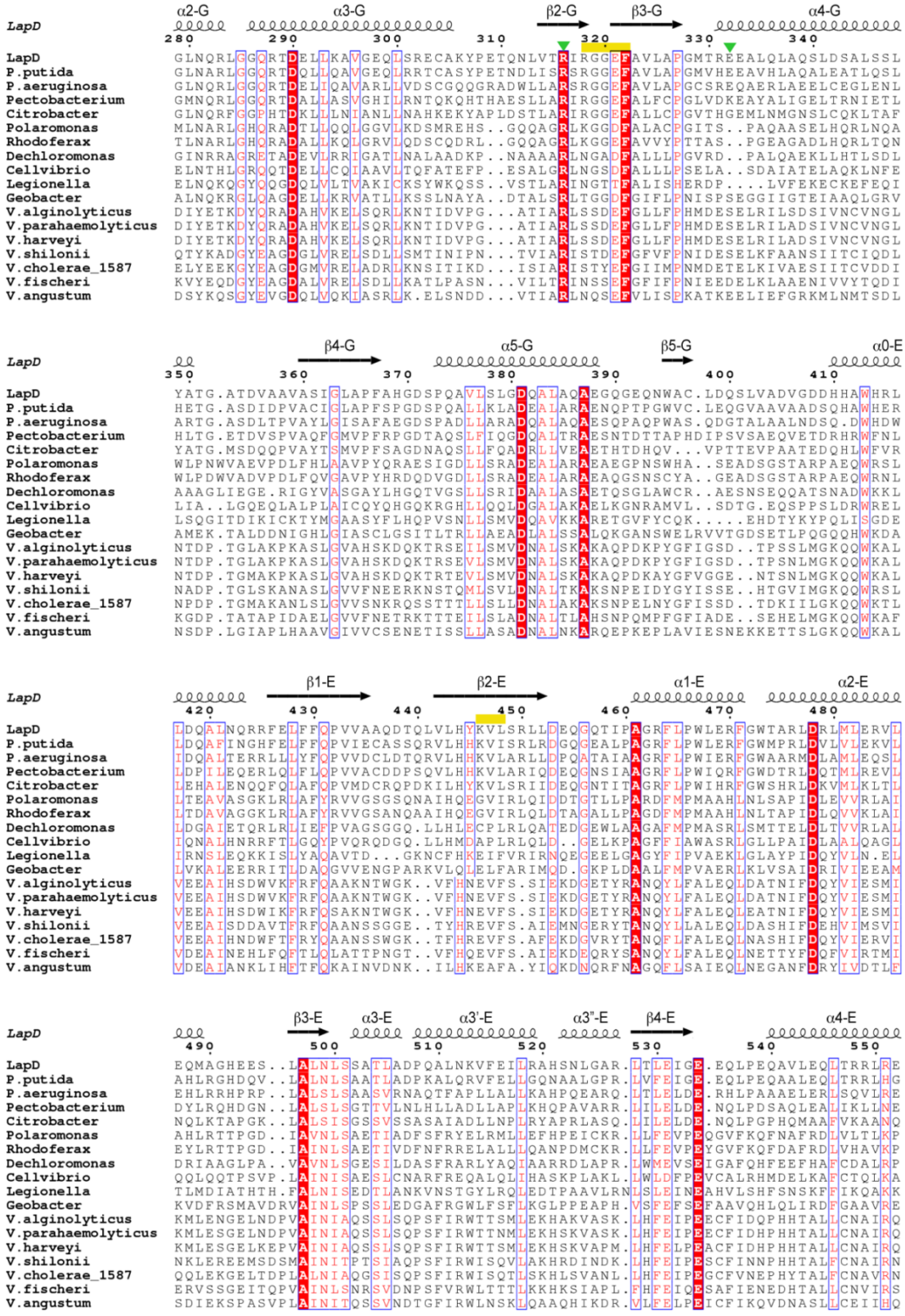
## Table 2.1: X-ray data collection and refinement statistics

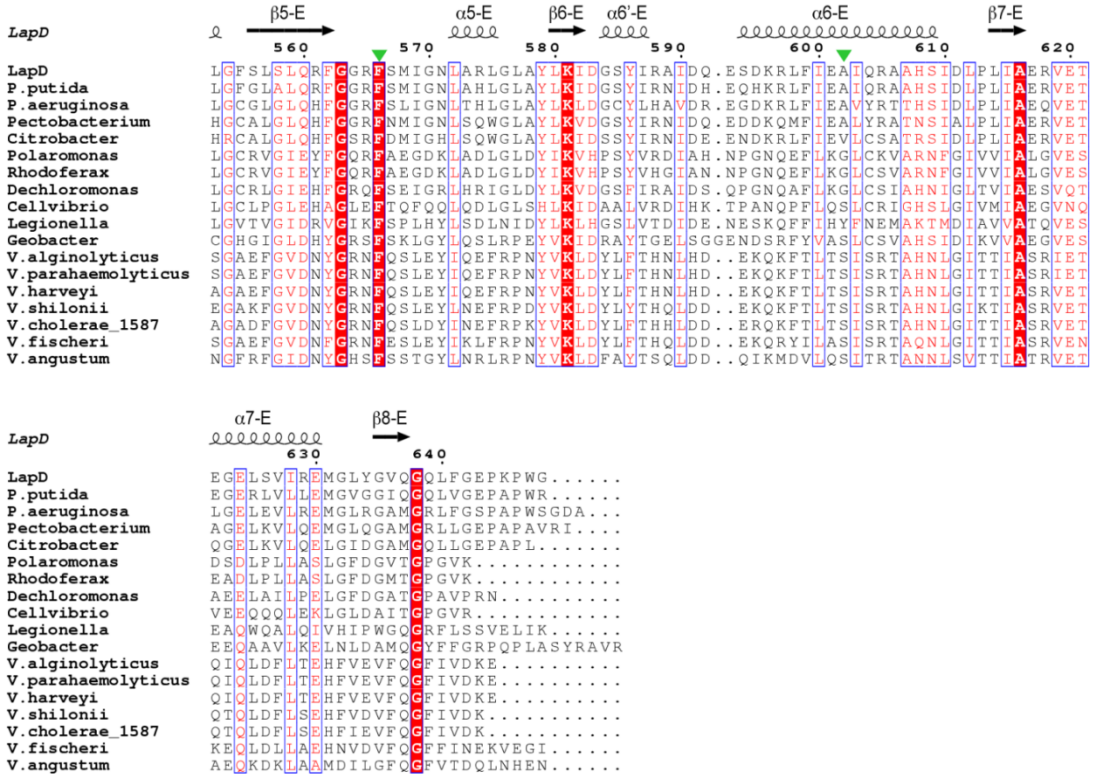
.

| | $LapD^{EAL}$•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) | 0.9771 | 0.9771 | 0.9771 | 0.9793 | 0.9771 | 0.9771 |
| Space group | $C222_1$ | $P6_522$ | $P2_12_12_1$ | I23 | $P3_2$ | $P3_2$ |
| 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 | |

**Figure 2.2: Sequence alignment of LapD homologs.**
A sequence alignment of LapD homologs from various species was generated with ClustalW2 (60) and formatted with ESPript (61). 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) and *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
β4-G α5-G β5-G α0-E
β1-E β2-E α1-E α2-E
β3-E α3-E α3'-E α3"-E β4-E α4-E
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


β5-E
α5-E
β6-E
α6'-E
α6-E
β7-E
LapD
560
570
580
590
600
610
620
LapD LGFSLSLQRPGGRFSMIGNLARLGLAYLKIDGSYIRATDQ.ESDKRLFIEAIQRAAHSIDLPLIAERVET
P.putida LGFGLALQRPGGRFSMIGNLAHLGLAYLKIDGSYIRNIDH.EQHKRLFIEAIQRAAHSIDLPLIAERVET
P.aeruginosa LGCGLGLQHPGGRFSLIGNLTHLGLAYLKLDGCYLHAVDR.EGDKRLFIEAVYRTTHSIDLPLIAEQVET
Pectobacterium HGCALGLQHPGGRFNMIGNLSQWGLAYLKVDGSYIRNIDQ.EDDKQMFIEALYRATNSIALPLIAERVET
Citrobacter HRCALGLQHPGSRFDMIGHLSQWGLAYLKIDSSYIRNIDE.ENDKRLFIEVLCSATRSIDLPLIAERVET
Polaromonas LGCRVGIEYPGQRFAEGDKLADLGLDYIKVHPSYVRDIAH.NPGNQEFLKGLCKVARNFGIVVIALGVES
Rhodoferax LGCRVGIEYPGQRFAEGDKLADLGLDYIKVHPSYVHGIAN.NPGNQEFLKGLCSVARNFGIVVIALGVES
Dechloromonas LGCRLGIEHPGRQFSEIGRLHRIGLDYLKVDGSFIRAIDS.QPGNQAFLKGLCSIAHNIGLTVIAESVQT
Cellvibrio LGCLPGLEHACLEFTQFQQLQDLGLSHLKIDAALVRDIHK.TPANQPFLQSLCRIGHSLGIVMIAEGVNQ
Legionella LGVTVGIDRVGIKFSPLHYLSDLNIDYLKLHGSLVTDIDE.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 SGAEFGVDNPGRNFESLEYIKLFRPNYVKLDYLFTHQLDD..EKQRYILASISRTAQNLGITTIASRVEN
V.angustum NGFRFGIDNYGHSFSSTGYLNRLRPNYVKLDFAYTSQLDD..QIKMDVLQSITRTANNLSVTTIATRVET
α7-E
β8-E
LapD
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.....

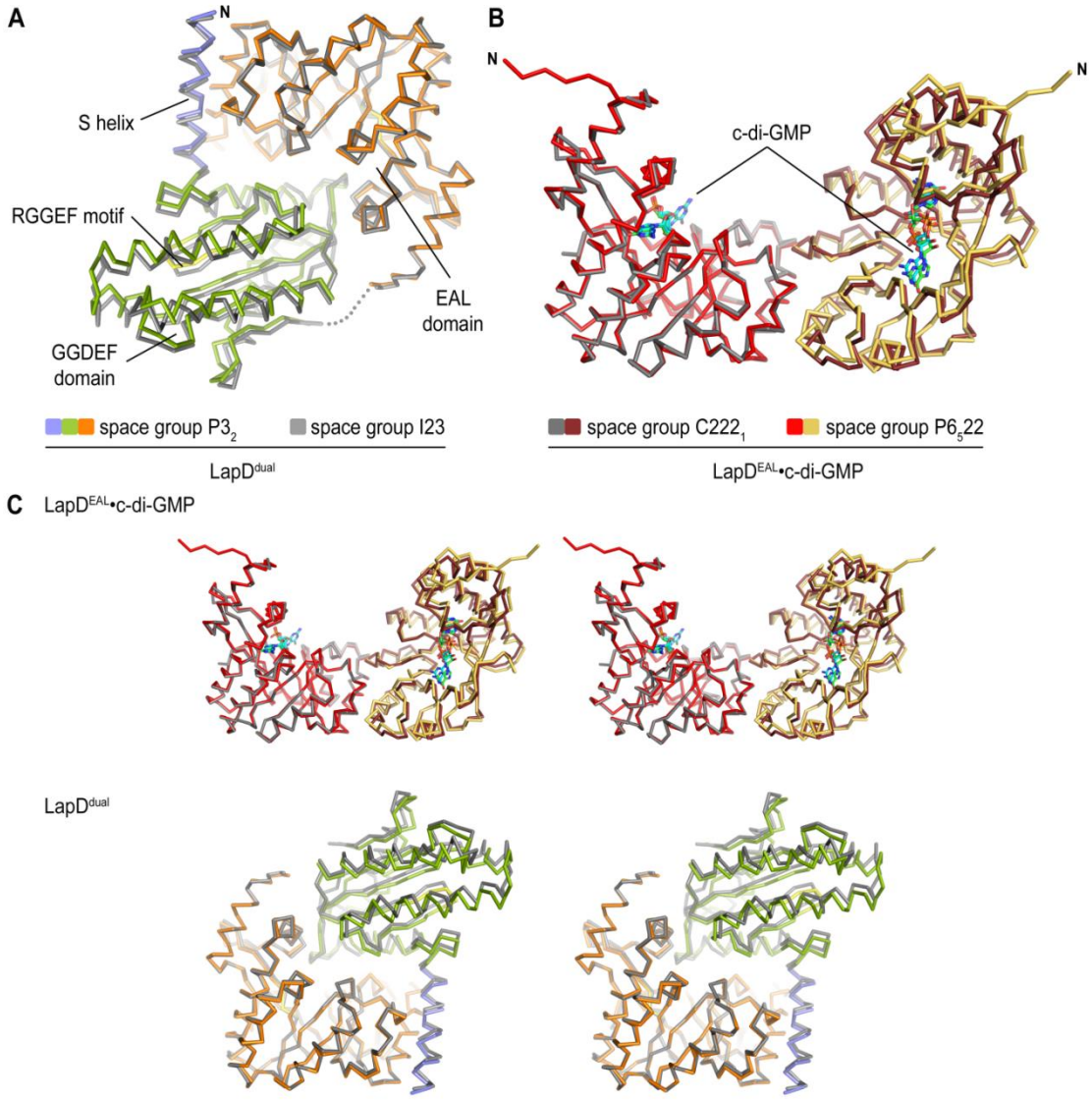


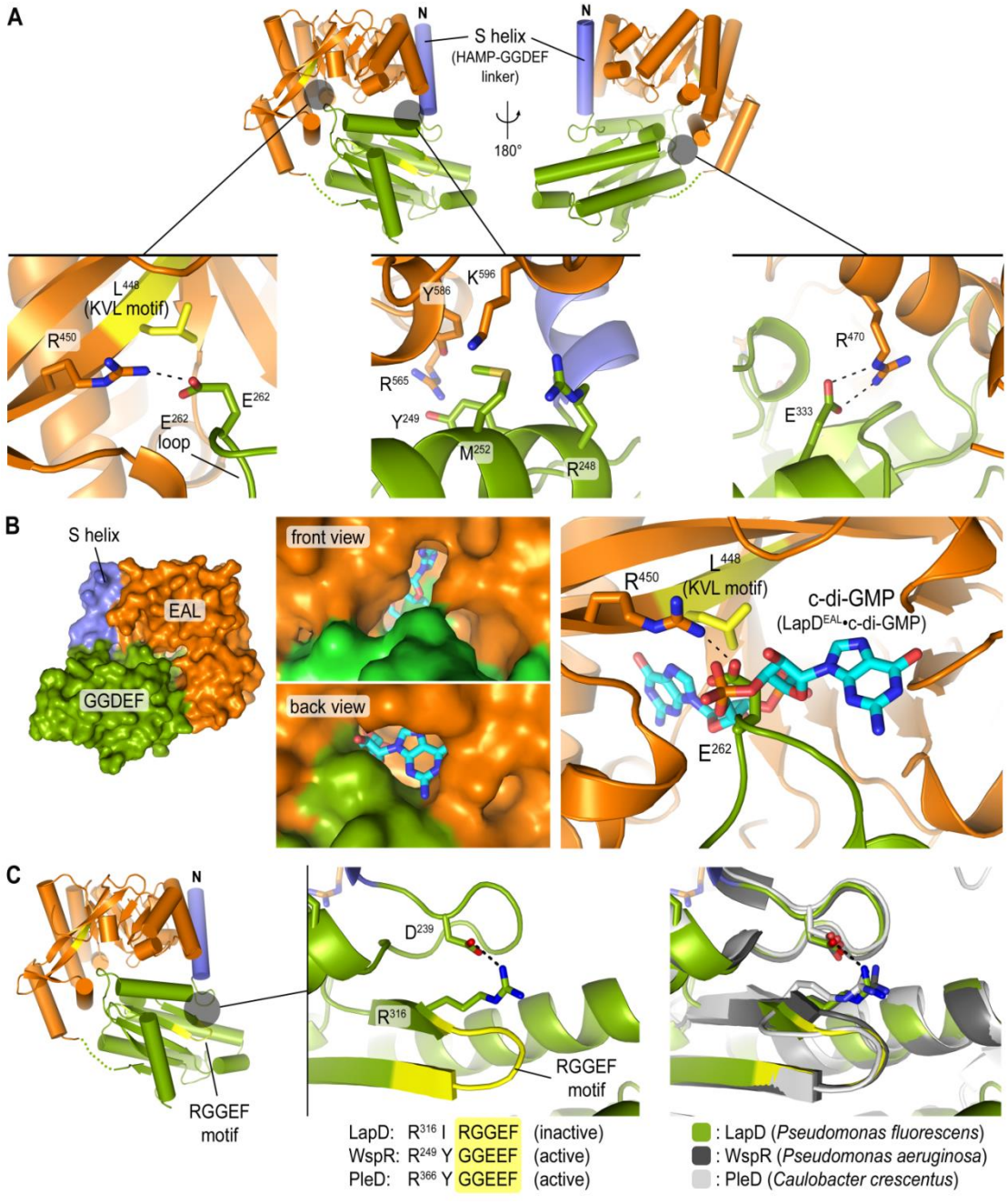

**Figure 2.3: Crystal forms of LapD$^{dual}$ and LapD$^{EAL}$•c-di-GMP.**
**(A)** LapD$^{dual}$: Two independent crystal forms were obtained for LapD$^{dual}$. The resulting structures were superimposed on the EAL domain and shown as protein backbone traces.
**(B)** C-di-GMP-bound LapD$^{EAL}$: Two independent 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.
**(C)** Stereo views: Stereo views of the structural comparisons shown in (A) and (B) are shown. In this view, the EAL domains of LapD$^{dual}$ and LapD$^{EAL}$•c-di-GMP are shown in a similar orientation.

**Figure 2.4: GGDEF–EAL domain interactions and S helix–GGDEF domain linker conformation observed in apo-LapD$^{dual}$.**

**(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 domains are 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). C-di-GMP is shown as a 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 (17),(18),(38),(41). 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
E262
loop
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 fluorescens)
: WspR (Pseudomonas aeruginosa)
: PleD (Caulobacter crescentus)

***Crystal Structure of LapD$^{EAL}$·c-di-GMP***

The crystal structure of LapD$^{EAL}$ bound to c-di-GMP (residues 399–648; LapD$^{EAL}$·c-di-GMP; Fig. 2.5) was solved by molecular replacement (Table 2.1). We obtained crystals in two independent conditions, yielding two different crystal forms (space group *C222$_1$*, two molecules per asymmetric unit; space group *P6$_5$22*, one molecule per asymmetric unit; Fig. 2.3B and 2.3C).

Cyclic-di-GMP binding did not alter the overall conformation of the EAL domain observed in the apo-LapD$^{dual}$ structure (rmsd of 0.6 Å over all atoms) (Fig. 2.5), consistent with the lack of major conformational changes upon dinucleotide binding to the EAL domains of YkuI, TDB1265, and FimX (35–37). Minor changes in the dinucleotide-binding pocket are confined to four c-di-GMP coordinating residues that adopt alternate side chain rotamer conformations (Fig. 2.5A).

The most notable conformational change in LapD$^{EAL}$ upon c-di-GMP binding occurs in the switch loop (Fig. 2.5B). Dinucleotide binding and the absence of the S helix in the isolated EAL domain allow the loop to restructure, resulting in the switching of the conserved phenylalanine residue F$^{566}$ (Fig. 2.5B). In apo-LapD$^{dual}$, the side chain of F$^{566}$ faces inward and is located at the center of the S helix–binding interface (Figure 2.1B). 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. 2.5B and 2.6). Whether this change is due to the flexibility of the loop, adjusting its conformation to accommodate the S helix–bound and dimeric states, or depends on dinucleotide binding awaits further structural analysis. The symmetric LapD$^{EAL}$ domain dimer is reminiscent of the oligomeric state in active EAL domain–containing phosphodiesterases, such as in *P. aeruginosa* SadR/RocR, *Bacillus subtilis* YkuI, *Thiobacillus denitrificans* TDB1265, and the BLUF domain–regulated photoreceptor BlrP1 from *Klebsiella pneumonia* (34, 35, 37, 40). Most importantly, dimerization of

the c-di-GMP-bound EAL domains is incompatible with the conformation observed in the crystals of apo-LapD$^{dual}$ (Fig. 2.6C). The surface occupied by the S helix overlaps significantly with the homodimerization interface, which indicates that dinucleotide-induced conformational changes will include the displacement of the GGDEF domain and the S helix.

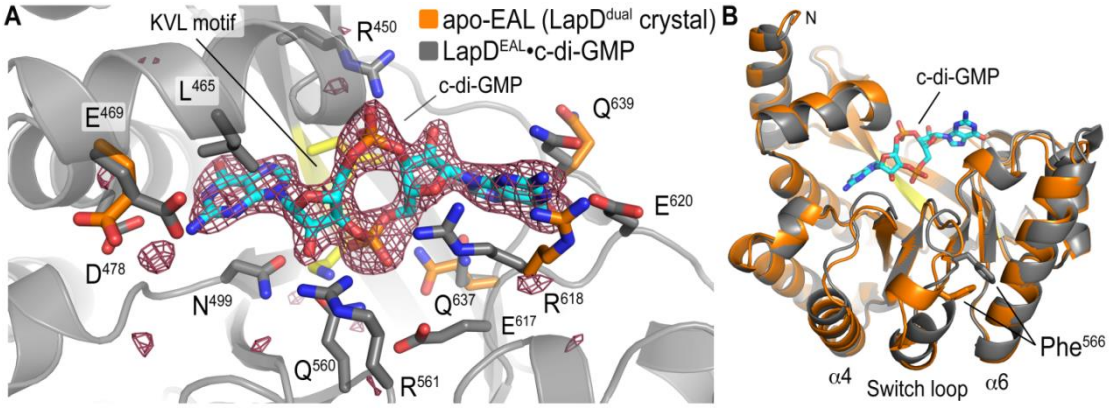



**Figure 2.5: Comparison between the dinucleotide-free and c-di-GMP-bound EAL domain of LapD** (Data courtesy of Petya V. Krasteva)**.**
**(A)** Crystal structure of LapD$^{EAL}$•c-di-GMP: The c-di-GMP-bound structure of LapD$^{EAL}$ (gray) was superimposed onto the dinucleotide-free structure of LapD$^{dual}$ (orange residues). The S helix and GGDEF domain were omitted for clarity. A close-up view of the dinucleotide-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 dinucleotide and is contoured at 3.5σ.
**(B)** Conformational change of the switch loop: C-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 F$^{566}$, a residue involved in both S helix interaction in LapD$^{dual}$ and dimerization of LapD$^{EAL}$, changes position.

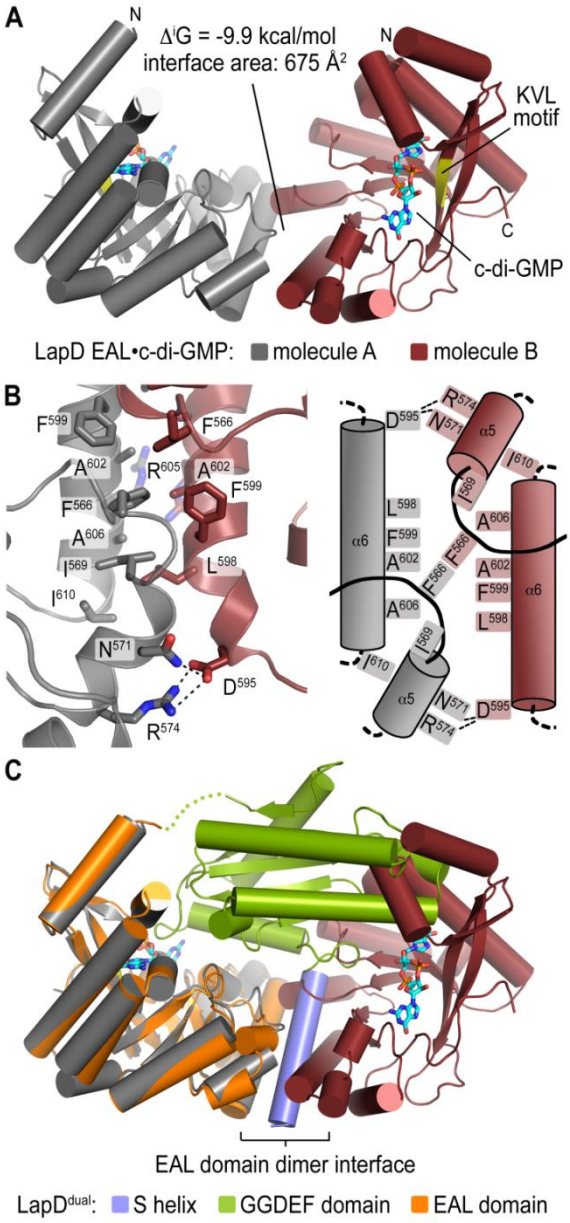



**Figure 2.6: Dimerization of c-di-GMP-bound LapD$^{EAL}$.**
**(A)** EAL domain dimerization: In both crystal forms obtained for LapD$^{EAL}$•c-di-GMP we observe symmetric dimerization between protomers involving helix α6 and the switch loop. Dimerization buries 1,350 Å$^2$ of surface area.
**(B)** Dimer 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 dinucleotide-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. 2.1.

***Analysis of the Regulatory Mechanisms of LapD in Solution***

Based on the crystallographic data, a simple model would suggest that LapD is subject to an autoinhibition mechanism. In contrast to other c-di-GMP receptors with known structures where the dinucleotide-binding site is freely accessible in the apo-state (Fig. 2.8), intramolecular interactions restrict dinucleotide access to the EAL domain in LapD. Cyclic-di-GMP binding would disrupt these interactions, resulting in a change in conformation of the receptor. Alternatively, mutations in the regulatory features predicted to destabilize the interaction should relieve the autoinhibition and alter the shape and activity of the receptor.

To test this model, structure-guided mutations were introduced into LapD to assess the functional relevance of the autoinhibitory conformation and EAL domain dimerization (Fig. 2.7). Site-directed mutations were introduced into the S-helix that was 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. 2.1B). 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. 2.4A). Finally, $A^{602}$ was targeted for mutation as it was identified as a residue at the center of the EAL domain dimerization interface (Fig. 2.6B). The structure of apo-LapD$^{dual}$ showed $A^{602}$ at the periphery of the S helix–EAL domain interaction, suggesting that perturbations at this site may maintain the autoinhibited state (Fig. 2.1B).

Mutations were introduced into LapD$^{dual}$, the EAL domain, and the full-length receptor. It is important to note that LapD is a dimeric receptor via its HAMP and

output domains, and therefore EAL domain dimerization (and dinucleotide binding) represents a conformational change within the receptor, rather than a change in its oligomeric state. The comparative analyses described below reveal the basic properties of the cytoplasmic module of LapD, especially the correlation between c-di-GMP binding and dimerization (Fig. 2.7 and 2.8). However, the specific interaction energies will likely be enhanced in the context of the full-length receptor compared to those of the isolated domains. Cell-based assays elucidate the functional relevance of these properties in intact LapD (Fig. 2.10 and 2.11).

We next analyzed the oligomerization state of $LapD^{dual}$ protein variants in solution, using static multi-angle light scattering (MALS) (Fig. 2.9). This method provides the population-averaged absolute molecular weight and hence quaternary state of proteins eluting from a gel filtration column. The technique measures the intensity of scattered laser light from a particle at multiple angles, which is proportional to the product of the molecular weight and the concentration of the particle, permitting rapid and facile comparison of oligomeric equilibria across a series of mutants (42).

The wild-type $LapD^{dual}$ protein elutes in a single peak from the size exclusion column with a molecular weight of 43.5 kDa, indicating a monomeric state in solution (Fig. 2.9A, left column). Incubation of the protein with c-di-GMP shifted the peak elution volume and increased the molecular weight slightly to 54.5 kDa. While being monomeric in the absence of dinucleotide, 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 towards dimeric species upon c-di-GMP binding (77.5 kDa and 71.4

kDa, respectively; Fig. 2.9A, left column). As predicted on the basis of the structural analysis, LapD$^{dual}$ variants containing a glutamate substitution in place of A$^{602}$ (A$^{602}$E and S$^{229}$D/A$^{602}$E) are monomeric in solution, independent of the presence of dinucleotide and unaffected by the additional mutation S$^{229}$D.

In general, the intermediate molecular weights and non-Gaussian peak shapes observed for wild-type LapD$^{dual}$ and the mutants S$^{229}$D and E$^{262}$A, predicted to be less inhibited, incubated with c-di-GMP prior to gel filtration, may indicate a fast exchange between monomeric and dimeric species relative to the data acquisition time and/or instability of the complex. To further investigate this phenomenon, we conducted concentration-dependent experiments by subjecting LapD$^{dual}$ to light scattering measurements at concentrations between 20 and 320 mM with or without incubation in c-di-GMP. All samples eluted as single peaks from the gel filtration column and showed no signs of unspecific protein aggregation. Protein concentration determination across the peak volume indicated that samples were diluted consistently, 15-fold during the chromatography. All LapD$^{dual}$ variants were monomeric in the absence of c-di-GMP across the entire concentration range (Fig. 2.9B). LapD$^{dual}$ proteins with a mutation at the dimerization interface (A$^{602}$E or S$^{229}$D/A$^{602}$E) were insensitive to c-di-GMP addition and remained monomeric. Wild-type LapD$^{dual}$ showed signs of oligomerization only at the highest concentrations tested. In contrast, the molecular weight of LapD$^{dual}$ variants with single-point mutations S$^{229}$D or E$^{262}$A, predicted to disrupt autoinhibitory features, increased in a concentration-dependent manner in the presence of c-di-GMP, indicative of dimerization of the isolated cytoplasmic domain in solution.

In summary, LapD appears to be autoinhibited for efficient dinucleotide binding by structural features involving the S helix 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 dinucleotide binding, we propose that dimerization and c-di-GMP binding are interdependent events in $LapD^{dual}$ and $LapD^{EAL}$.

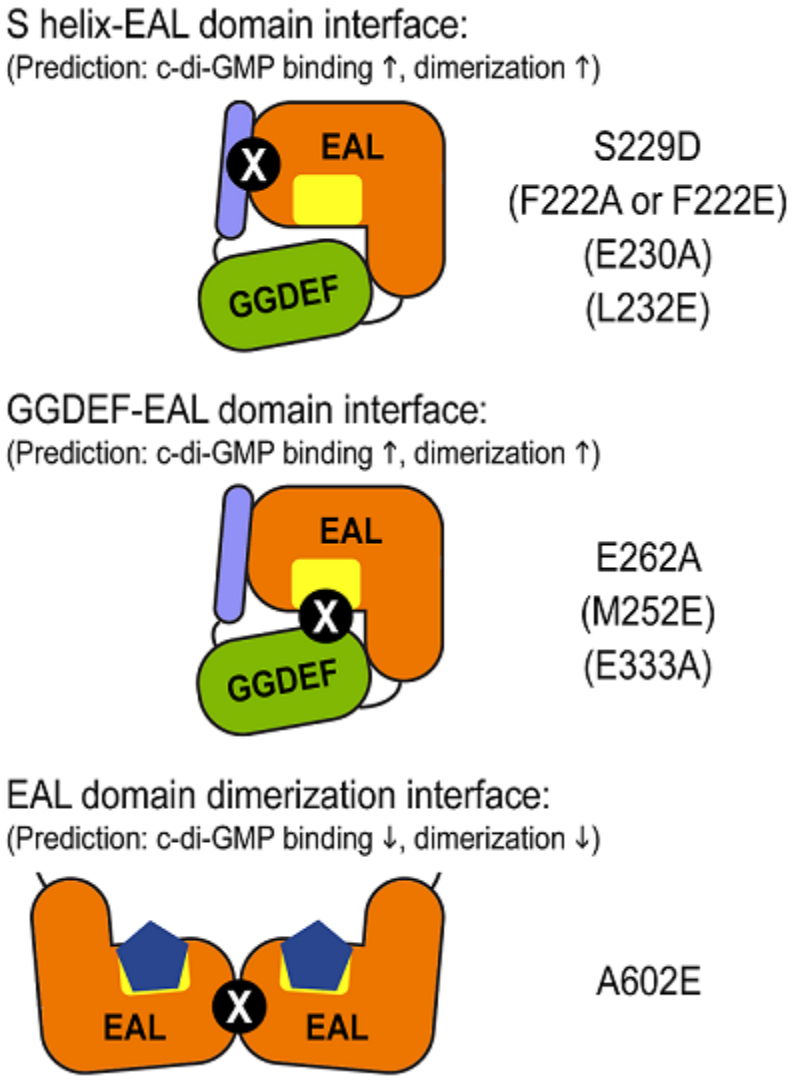



**Figure 2.7: c-di-GMP binding of LapD$^{dual}$ in solution.**
Mutant categories: Structure-guided, site-directed mutants in LapD are illustrated. Mutations in brackets were used in experiments shown in Fig. 2.10 and 2.11. Structure-based predictions regarding the c-di-GMP binding and c-di-GMP-dependent dimerization propensities are indicated.

**Figure 2.8: Structural comparion of LapD with other c-di-GMP receptors.**
**(A)** LapD: The monomoric apo-LapD$^{dual}$ structure is shown as a surface presentation (top). The middle panel shows the c-di-GMP-bound EAL domain of LapD in two orthogonal views. The dinucleotide-binding site is colored in red. The conformation of c-di-GMP is similar to that observed in other EAL domains such as FimX, YkuI, and BlrP1 (bottom panel) (17),(18),(34)–(38).
**(B)** VpsT (PDB IDs 3kln and 3klo): The transcription factor VpsT from *V. cholerae* exists in a monomer–dimer equilibrium. An apo-VpsT monomer is shown as a surface presentation (top panel). The dimeric species is stabilized by c-di-GMP binding to the base of the regulatory receiver domain (middle panel) (14). Two molecules of c-di-GMP form an intercalated dimer, similar to the binding mode observed for the inhibitory site binding in active diguanylate cyclases (17), (38). The dinucleotide binding site is shown in red.
**(C)** PilZ domains (PDB IDs 1yln, 2rde, 3yg, and 3kyf): The PliZ domain–containing protein PlzD/VCA0042 forms homodimers via its YcgR-N* domain. The PilZ domains form separate lobes of the protein. PilZ domain–containing proteins have been shown to bind either one or two mutually intercalated molecules of c-di-GMP (54). The dinucleotide binding site is shown in red.

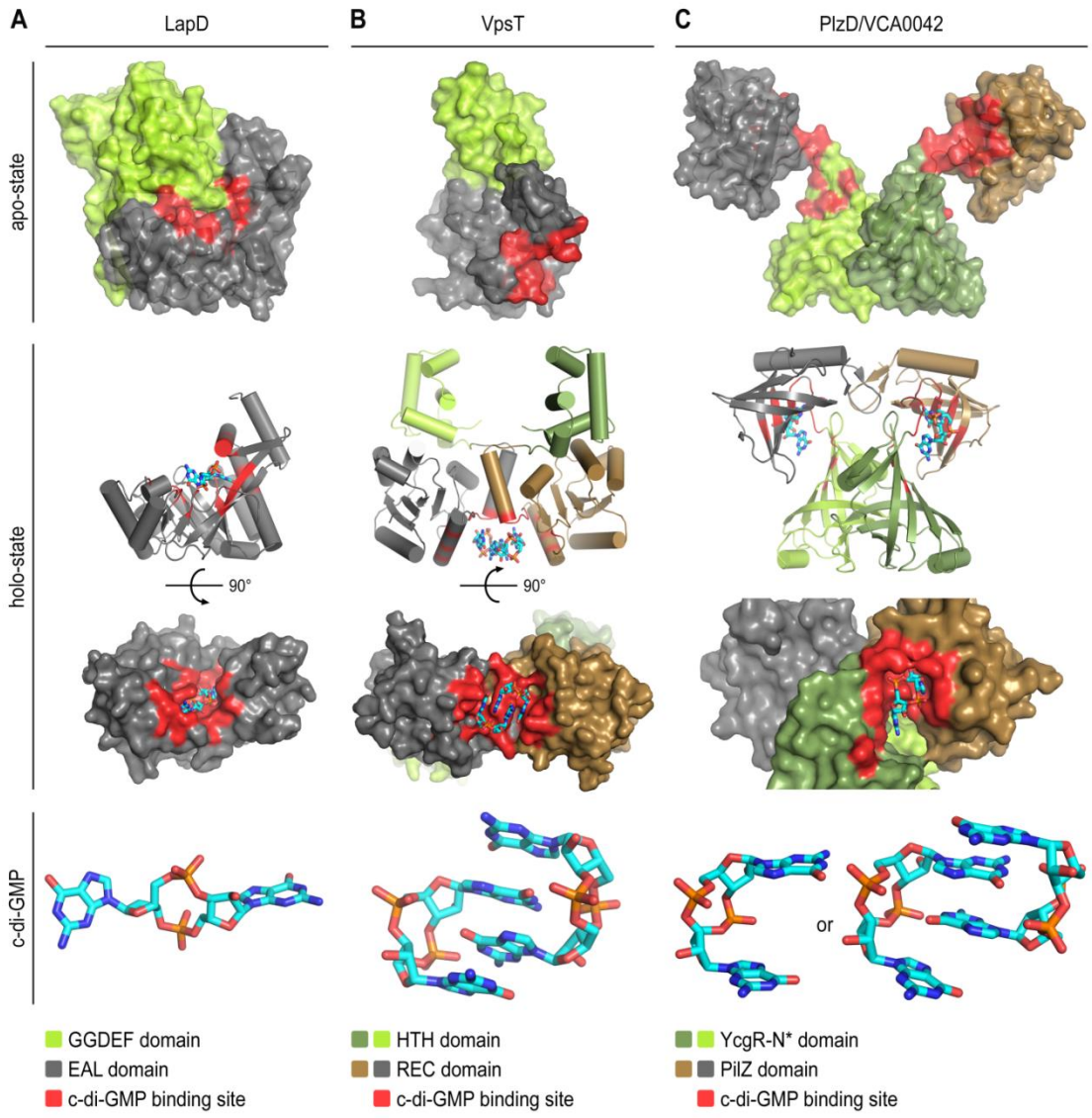
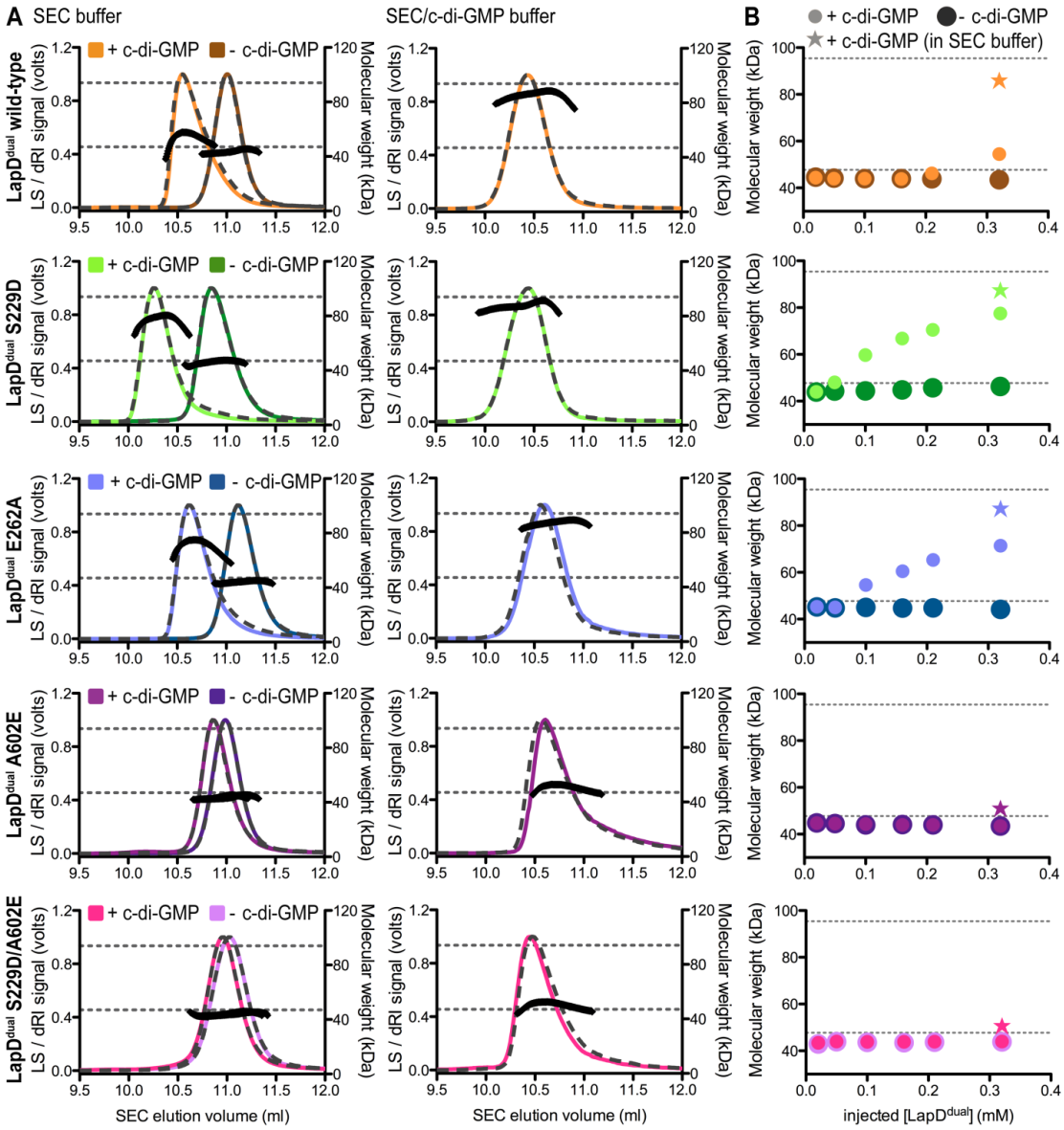

A
LapD
B
VpsT
C
PlzD/VCA0042
apo-state
holo-state
c-di-GMP
90°
90°
or
GGDEF domain
EAL domain
c-di-GMP binding site
HTH domain
REC domain
c-di-GMP binding site
YcgR-N* domain
PilZ domain
c-di-GMP binding site

**Figure 2.9: Quaternary state of LapD$^{dual}$ in solution.**
**(A)** Oligomerization of LapD$^{dual}$ in solution: SEC-coupled MALS analysis of wild-type and mutant LapD$^{dual}$ in the presence and absence of c-di-GMP is shown. The signal from the 90° scattering detector is shown in color, and the signal from the refractive index detector is shown as a 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 gray lines. Injected protein and dinucleotide concentrations were 250 mM and 500 mM, respectively. In the right panel, the mobile phase contained c-di-GMP (50 mM). Earlier elution times may indicate a more elongated conformation of certain mutants in solution (for example, of the mutant S$^{229}$D compared to wild-type or the E$^{262}$A variant in the absence of c-di-GMP), which is probably due to a displacement of the GGDEF domain from the EAL domain.
**(B)** Concentration-dependent dimerization of LapD$^{dual}$: SEC-MALS experiments were carried out with samples of increasing LapD$^{dual}$ concentration. The samples of highest concentration correspond to data shown in (A). The data point shown as a star represents data obtained for samples run in a mobile phase that contained c-di-GMP.


A
SEC buffer
SEC/c-di-GMP buffer
B
+ c-di-GMP
- c-di-GMP
+ c-di-GMP (in SEC buffer)
LapDdual wild-type
LapDdual S229D
LapDdual E262A
LapDdual A602E
LapDdual S229D/A602E
LS / dRI signal (volts)
Molecular weight (kDa)
SEC elution volume (ml)
injected [LapDdual] (mM)

***Effect of Structure-Based Mutations in LapD on Biofilm Formation***

Stable biofilms of *P. fluorescens* require LapD expression and the presence of c-di-GMP (24). 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 ΔlapD mutant strain (Fig. 2.10). 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 ΔLapD is constitutively activated by mutations in the HAMP domain (24) (Fig. 2.10 and 2.11). The mutation that we predict to disrupt the S helix–EAL interface in the autoinhibited conformation, $S^{229}D$, caused an ''activated'' phenotype. Similar results were obtained with the mutant $F^{222}E$, whereas a less disruptive alanine substitution was tolerated at this position.

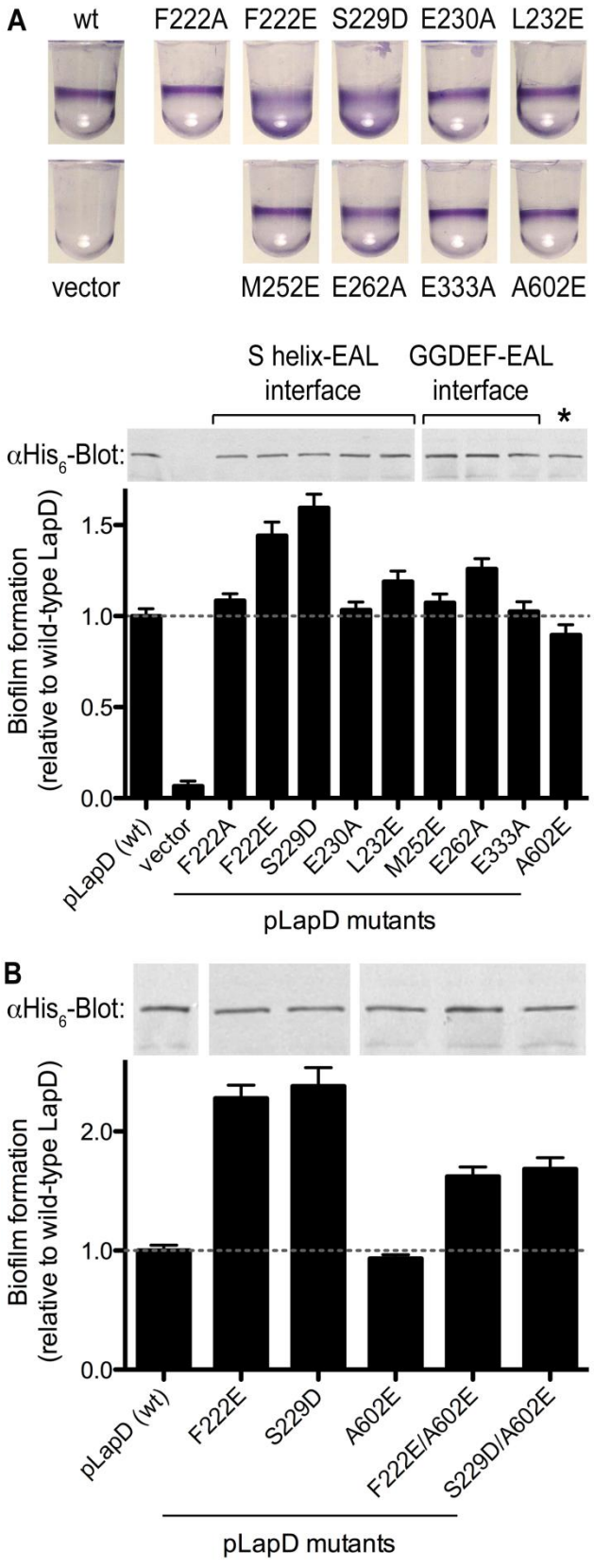
In the apo-$LapD^{dual}$ structure, the $E^{262}$ residue is positioned such that it would occlude binding of c-di-GMP to the EAL domain (Fig. 2.4B). Consistent with this, the $E^{262}A$ mutation results in an increase in biofilm formation relative to the wild-type allele (Fig. 2.10A). Yet, the $E^{262}A$ 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* (Fig. 2.9). This suggests that the $E^{262}A$ mutant is still subject to autoinhibition *in vivo*, albeit with higher sensitivity for c-di-GMP than the wild-type protein. Structurally, this may be explained by removal of the side chain that directly occupies the c-di-GMP-binding site without disturbing the S helix–EAL domain interaction. Other mutations showed intermediate ($L^{232}E$ and $M^{252}E$) or no significant changes ($F^{222}A$, $E^{230}A$, and $E^{333}A$) in phenotype, roughly

corresponding to their surface exposure in the autoinhibited state structure (Fig. 2.1B, 2.4A, and 2.10A). 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. 2.9), led to a small but significant decrease in biofilm formation relative to the wild-type allele (Fig. 2.10). 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 dinucleotide binding pocket (24), argues that dimerization increases the stability of the dinucleotide 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 $F^{222}E$. The reduction in biofilm formation in the double mutants was significant, corroborating that EAL domain dimerization plays a role in LapD function *in vivo* (Fig. 2.10B). The single mutants were also tested for their response to phosphate starvation, a physiological input for LapD-mediated signaling that leads to a reduction of cellular c-di-GMP concentration (24, 28). At low c-di-GMP concentration, wild-type LapD activity is downregulated, which results in the release of the adhesin LapA from the cell surface and thus a reduction in biofilm formation (Fig. 2.11A, top) (24). Mutations in the S helix–EAL domain interface ($F^{222}E$ and $S^{229}D$) failed to respond to phosphate starvation efficiently, showing little to no reduction in biofilm formation (Fig. 2.11A). 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.

2.11A) (24). In contrast, mutation of the residue in the GGDEF domain that occupies the c-di-GMP binding site ($E^{262}A$) showed an intermediate response to phosphate starvation, suggesting that mutant receptor function is still controlled by c-di-GMP, albeit not as effectively as in wild-type LapD (Fig. 2.11A). Similar to the trends observed in the static biofilm assay (Fig. 2.10A), other mutations in LapD showed more subtle effects in the phosphate starvation experiments (Fig. 2.11B).

Collectively, these results suggest that the S helix–EAL domain interface stabilizes the off state. The interaction is the dominant autoinhibitory feature responsible for positioning the GGDEF domain to occlude the c-di-GMP-binding pocket and therefore ensure 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, although it appears to be a secondary component of the activation mechanism.

**Figure 2.10: Phenotypic analyses of lapD mutants** (Data courtesy of Peter D. Newell)**.**
**(A)** Biofilm phenotypes: Biofilm formation of ΔlapD cells 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 6 SD of eight 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)** Biofilm phenotypes of double mutants: The analysis was carried out as described in (A). Data are means 6 SD of eight replicates.


A
wt
F222A
F222E
S229D
E230A
L232E
vector
M252E
E262A
E333A
A602E
S helix-EAL interface
GGDEF-EAL interface
*
αHis6-Blot:
Biofilm formation (relative to wild-type LapD)
1.5
1.0
0.5
0.0
pLapD (wt)
vector
F222A
F222E
S229D
E230A
L232E
M252E
E262A
E333A
A602E
pLapD mutants
B
αHis6-Blot:
Biofilm formation (relative to wild-type LapD)
2.0
1.0
0.0
pLapD (wt)
F222E
S229D
A602E
F222E/A602E
S229D/A602E
pLapD mutants

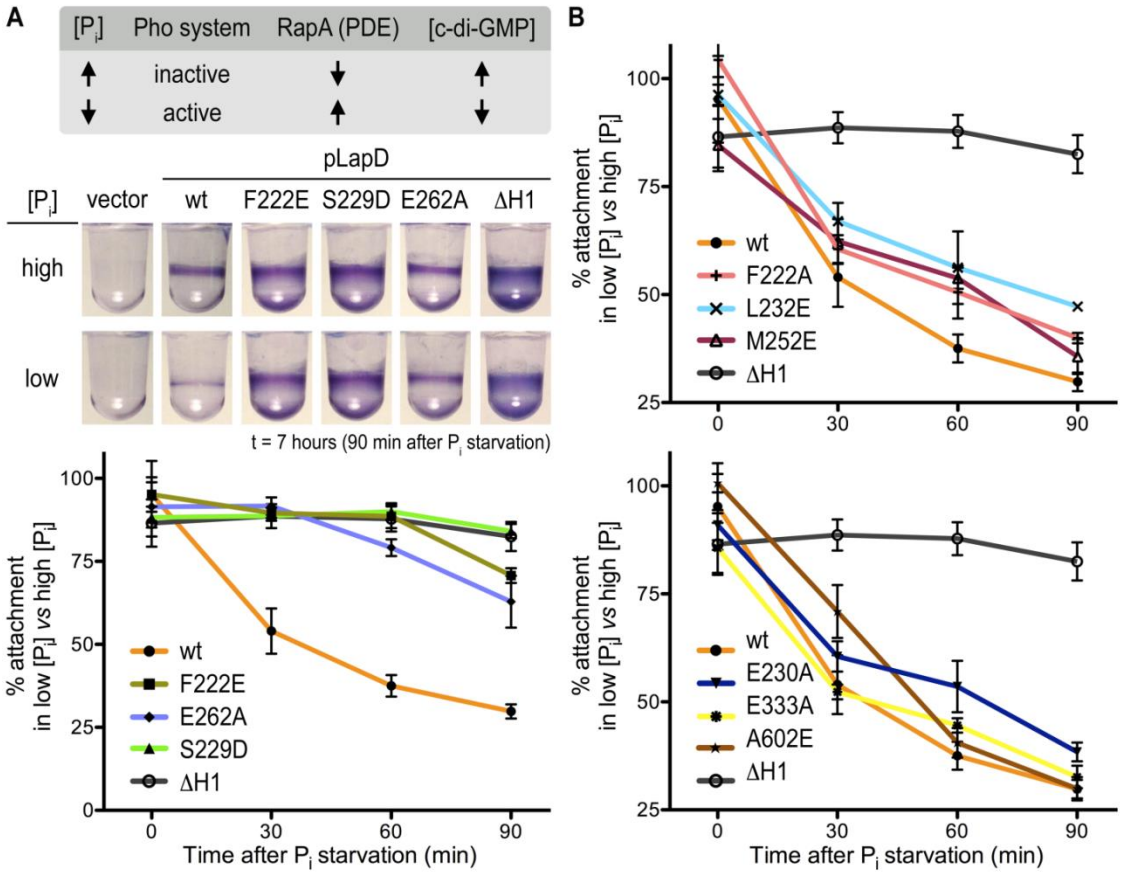



**Figure 2.11: Phosphate-regulated c-di-GMP signaling via LapD** (Data courtesy of Peter D. Newell)**.**

**(A)** Phosphate-regulated c-di-GMP signaling: Phosphate ($P_i$) starvation leads to the expression of the active phosphodiesterase RapA and a reduction in cellular c-di-GMP concentration (24). 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 medium, and compared to biofilm formation in phosphate-rich medium. The mutant DH1 contains an activating deletion in the HAMP domain and has been described previously (24). Data are means ± SD of eight replicates.

**(B)** Mutants showing intermediate responses: The analysis was carried out as described in (A). Data are means ± SD of eight replicates.

***Crystal Structure of LapD's Periplasmic Output 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. 2.1A). Crystals grown with selenomethionine-derivatized protein diffracted X-rays to a maximum resolution of 1.8 Å (Table 2.1). The structure was solved by single-wavelength anomalous dispersion phasing (SAD). The final model consists of two molecules per asymmetric unit spanning residues 23–150 (Fig. 2.12A and 2.13A).

The periplasmic output domain of LapD forms an extensively interwoven, domain-swapped dimer sharing 3,429 $Å^2$ interfacial surface area between the protomers (1/3 of LapD's output domain molecular surface) (Fig. 2.12 and 2.13B). 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. 2.13A).

A DALI (distance-matrix alignment) search comparing LapD's output domain to proteins in the RSCB Protein Data Bank (PDB) 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) (43–45). 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 (46, 47). 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 (44, 47, 48).

A sequence alignment of 18 sequences was constructed, including LapD homologs from other Pseudomonas strains and extending to more distantly related sequences from other bacterial genera (Fig. 2.2; Table 2.2). Mapping sequence conservation onto the accessible molecular surface revealed a few potentially important motifs (Fig. 2.12C and 2.14). 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. 2.12C).

Given its strict conservation and peculiar conformation, we targeted $W^{125}$ in a site-directed mutagenesis study, 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 ($His_6$)–tagged LapG, a periplasmic cysteine protease that binds to LapD's output domain in a c-di-GMP-dependent manner (29), we could efficiently pull down the untagged wild-type output domain (Fig. 2.12D). 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 ΔlapD genetic background (Fig. 2.12E). 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.

**Figure 2.12: Structure–function analysis of the 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 a 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.
**(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 as a surface presentation while the other is shown as a ribbon presentation. Conserved motifs and individual residues are highlighted.
**(D)** $LapD^{output}$–LapG complex formation: Purified $His_6$-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 $W^{125}$ has been replaced with a glutamate. The Coomassie-stained gel shows eluates of NiNTA-bound proteins.
**(E)** Biofilm phenotypes and LapD stability: Biofilm formation of ΔlapD cells 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.

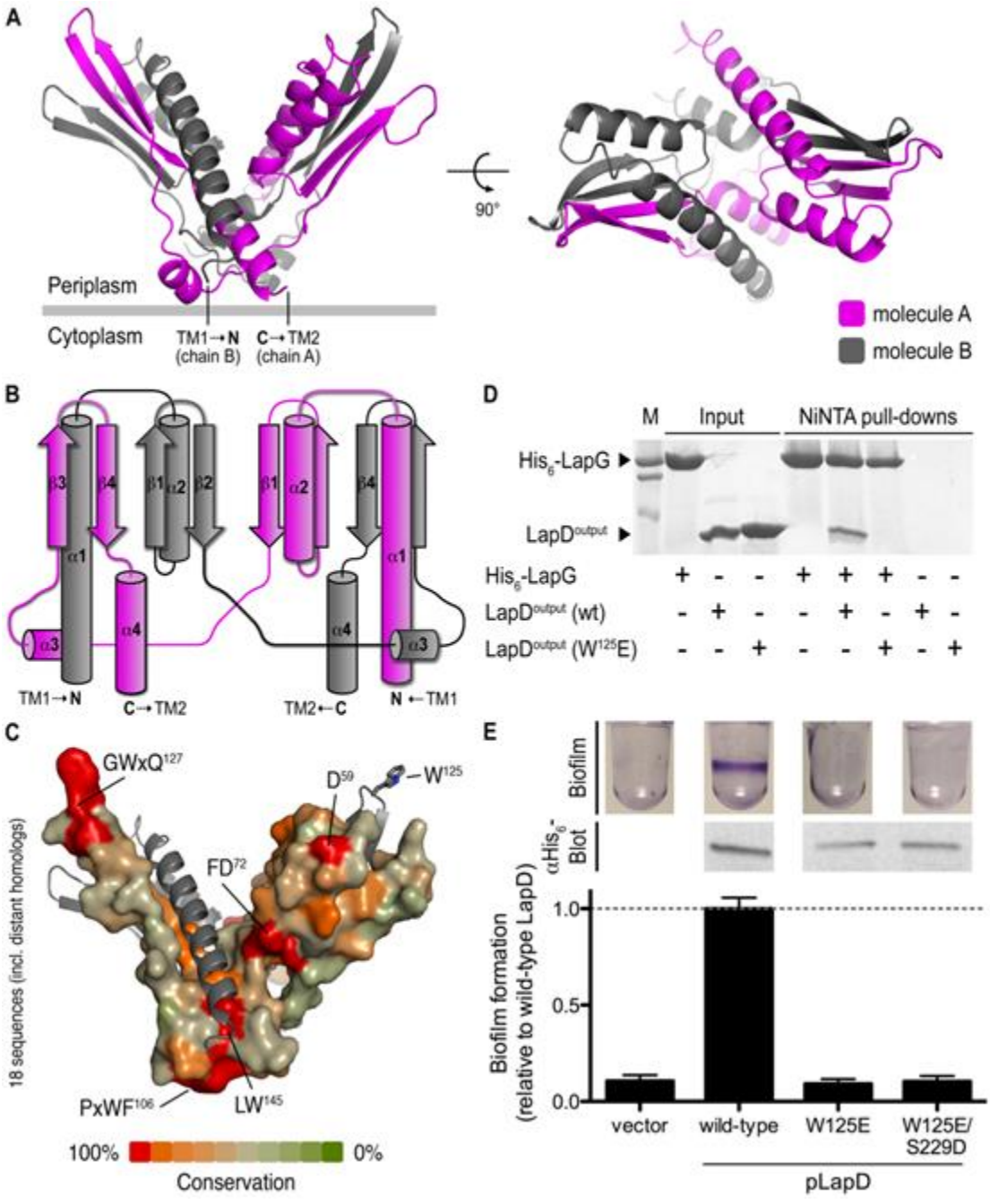
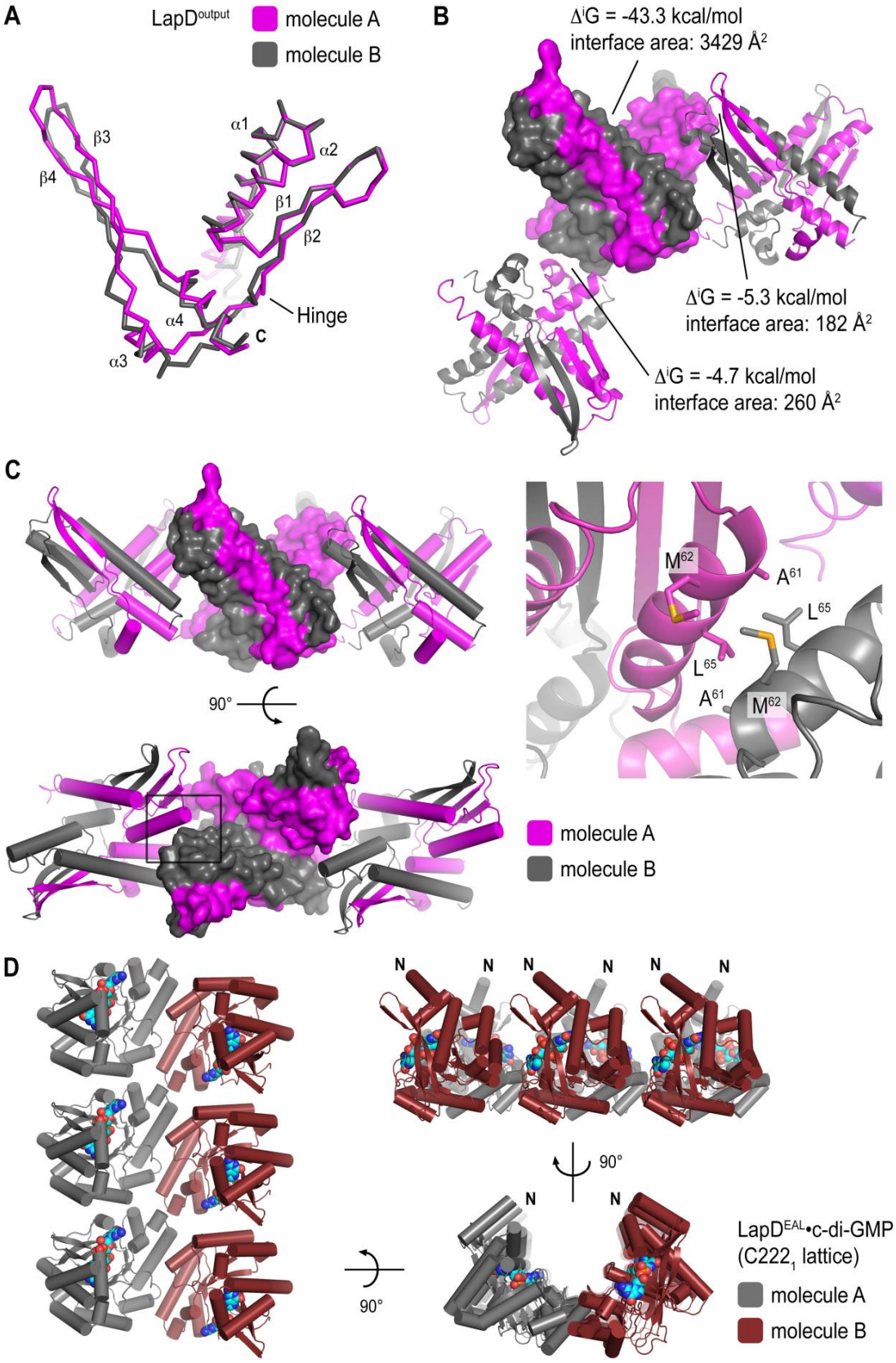

A
Periplasm
Cytoplasm
TM1→N (chain B)
C→TM2 (chain A)
90°
molecule A
molecule B
B
β3
β4
β1
α2
β2
α1
α3
α4
TM1→N
C→TM2
TM2←C
N←TM1
C
GWxQ127
D59
W125
FD72
PxWF106
LW145
18 sequences (incl. distant homologs)
100%
0%
Conservation
D
M
Input
NiNTA pull-downs
His6-LapG
LapDoutput
His6-LapG + - - + + + - -
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

**Figure 2.13: 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
(C222₁ lattice)
molecule A
molecule B

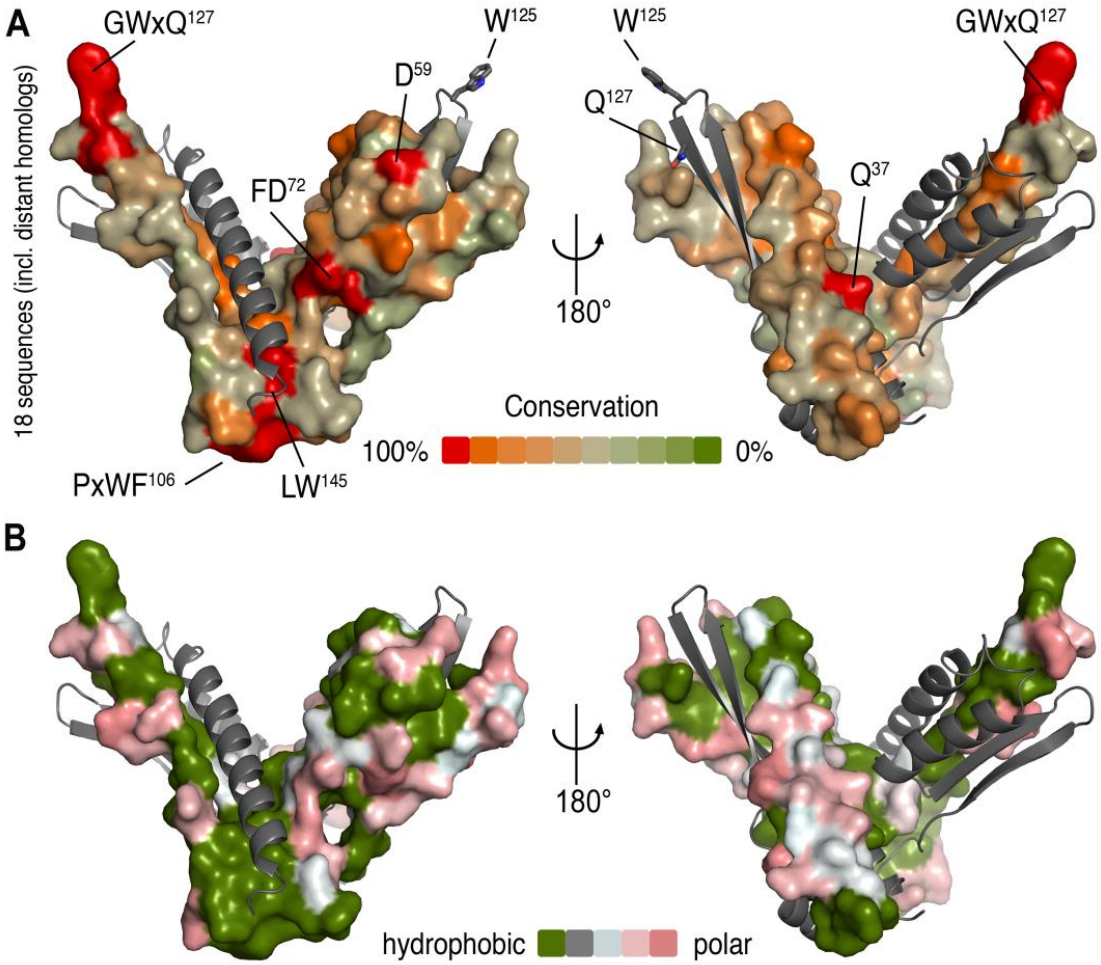



**Figure 2.14: Surface conservation and hydrophobicity of LapD$^{output}$.**
**(A)** Surface conservation: Based on an alignment of 18 sequences of LapD homologs (Fig. 2.2), the sequence conservation was mapped onto the solvent-accessible surface of the output domain. One protomer is shown as a surface presentation while the other is shown as a 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.

## Table 2.2: 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 $W^{125}E$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $W^{125}E,S^{229}D$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $F^{222}A$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $F^{222}E$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $S^{229}D$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $E^{230}A$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $L^{232}E$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $M^{252}E$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $E^{262}A$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $E^{333}A$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $A^{602}E$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $F^{222}E,A^{602}E$ | pLapD point mutation in the indicated codon or codons | This study |
| pLapD $S^{229}D,A^{602}E$ | pLapD point mutation in the indicated codon or codons | This study |

***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 (49, 50). 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. 2.15). 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, concomitant with a sampling of a c-di-GMP-binding-competent conformation of LapD, 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. 2.10 and 2.11) (24). These data suggest that intrinsic autoinhibitory interactions are indeed necessary to prevent the system from adopting a constitutively active conformation.

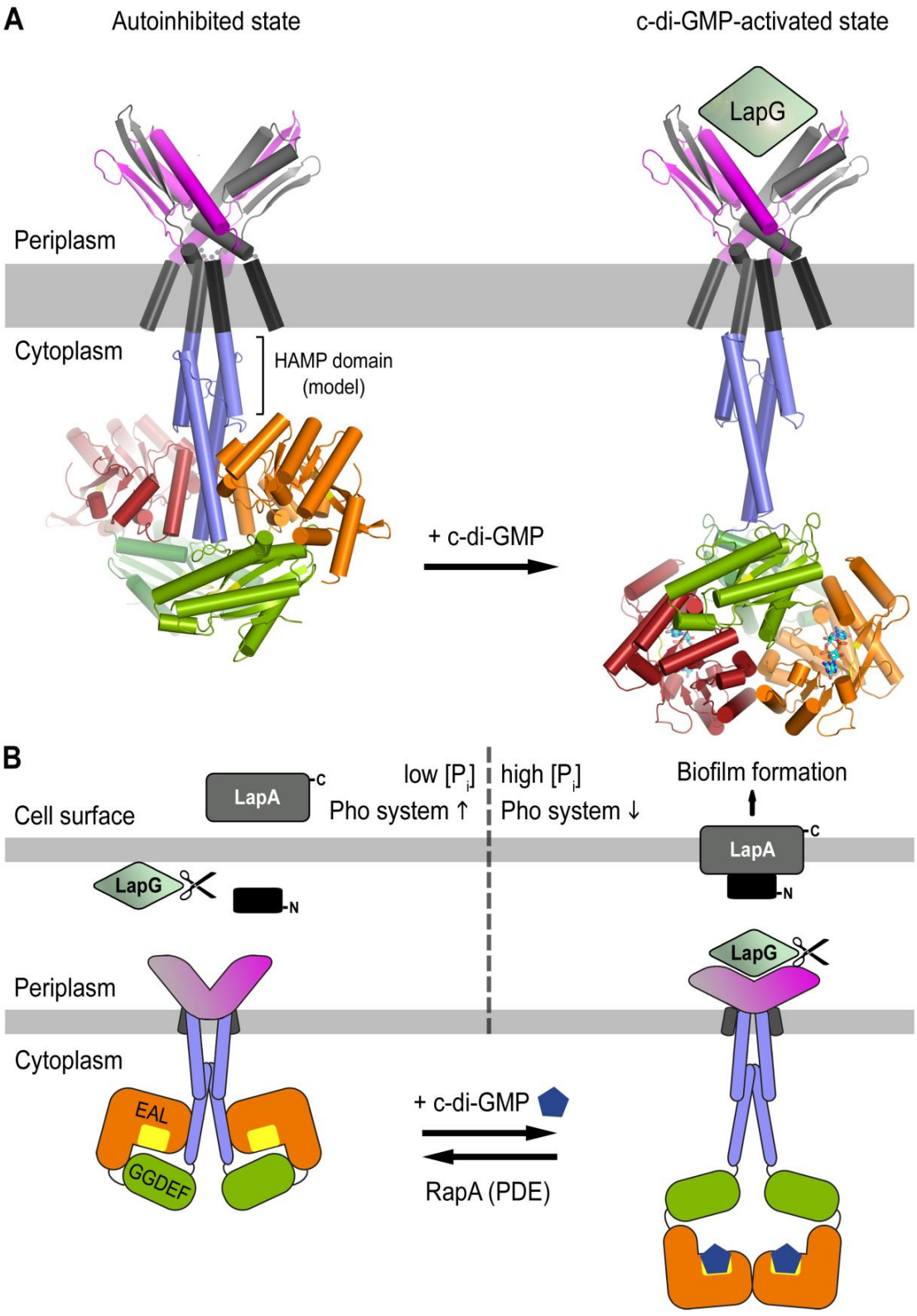
Based on the primary sequence and secondary structure predictions, the HAMP domain is directly linked to the GGDEF–EAL domain module via the S helix. HAMP domains occur in a large number of predominantly transmembrane sensor proteins that transmit signals from the environment across the cell membrane to elicit an intracellular response (outside-in signaling) (21). Rotation of the helices in HAMP dimers has been described as the main mechanism for signal transmission (49). It is conceivable that the EAL domain–S helix interaction stabilizes the off state, and that the release of the EAL domain from the S helix will allow the receptor to relax. The disengagement may trigger a rotation in the HAMP domain in a similar fashion to in other HAMP domains (49, 50), yielding a conformational change in the output domain and allowing the periplasmic domain of LapD to sequester LapG.

What is the relevance of 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 speculate that a domain-swapped fold would respond more efficiently and precisely in coupling conformational changes in the cytosolic domains across the membrane than canonical dimeric periplasmic domains. One may consider the periplasmic domain of LapD as a single domain given the extensive sharing of structural elements and a negligible monomer–dimer transition. 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 the 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, assuming that both tryptophan residues of the dimeric, periplasmic fold interact with LapG (monomers or dimers).

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 (P. D. N., unpublished data). It is likely that the intracellular and transmembrane domains facilitate the formation of a stable, high-affinity state. In addition, removal of the domain from its native context may alter its conformation. The observation that the isolated output domain can bind LapG is consistent with a model in which the dinucleotide-free, intracellular domains hold the receptor in an autoinhibited conformation that relaxes into a LapG-binding state upon activation. Consequently, deletion 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. 2.13C and 2.13D). Further experiments will be required to determine the oligomeric state of full-length LapD in the absence and presence of c-di-GMP.

**Figure 2.15: 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 (49-50).
**(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 (24–26, 29).


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

***Conservation of Signaling Systems Involving LapD Homologs***

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. 2.2; Table 2.2) (24, 27). While LapD and LapG from *P. aeruginosa* (PA1433 and PA1434, respectively) show a high degree of sequence conservation and functionally rescue deletions in these genes in *P. fluorescens*, no biofilm phenotype has been associated with this signaling system in their native strain (23), 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. 2.2; Table 2.2). LapD from *V. 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 *V. cholerae*, associated with increased biofilm formation (51).

The bioinformatic analysis also detected the presence of associated ABC transporters in genomes encoding LapD homologs, as in the case of *P. fluorescens*. Putative substrates of the cysteine protease LapG may fall into one of two categories. The large adhesin LapA as a LapG substrate, involved in biofilm formation and stability in *P. fluorescens* (29). 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. Regions with homology to the LapG-cleavage site of LapA have been identified 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 (52). Other proteins containing the tandem domain module with a higher degree of conservation at the putative enzyme active sites exist in association with a HAMP domain in some bacterial genomes (e.g., *V. cholerae*). 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 the outside-in signaling mechanism, thus leading to changes in cellular c-di-GMP levels.

## 2.4. Conclusions

Here, we elucidated the molecular mechanism underlying the function and regulation of *P. fluorescens* LapD, a transmembrane receptor essential for biofilm formation in this strain. Similar receptors are conserved in many bacteria where they control a LapG-type, periplasmic protease. LapD is autoinhibited with regard to c-di-GMP binding by interactions of the EAL domain with the S helix and the GGDEF domain. Receptor activation requires the concurrent release of the EAL domain from these interactions and the binding of c-di-GMP, which triggers a conformational change in the output domain from an incompetent to a competent state with regard to LapG binding (29). Mutations in the regulatory features that weaken the autoinhibitory interactions render LapD constitutively active even under phosphate starvation (low c-di-GMP levels; Fig. 2.11). This is in contrast to other c-di-GMP receptors with known structure, such as PilZ domain–containing proteins (53, 54), VpsT (14), and the GGDEF–EAL domain–containing protein FimX (36). In PlzD, dinucleotide binding introduces a conformational change that changes the relative orientation of its two domains (53). In FimX, the EAL domains form the distal tips of an elongated, dimeric protein (36). Cyclic-di-GMP binding to the isolated EAL domain or the full-length protein is indistinguishable, and no major conformational change has been observed for FimX upon dinucleotide binding (36, 55). Given the occurrence of the HAMP–GGDEF–EAL domain module in many other proteins from different free-living and pathogenic bacterial species, the results discussed here will have broad implications for receptors predicted to mediate either inside-out or outside-in signaling involving the bacterial second messenger c-di-GMP.

## 2.5. 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 (24). Proteins used in crystallization and *in vitro* studies were expressed and purified as follows. Construct boundaries were chosen based on secondary structure predictions and sequence alignments. 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 *P. 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 (14). Briefly, native proteins were expressed in T7 Express cells (NEB), grown at 37ºC 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 18ºC and protein expression was induced with 1 mM 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 5 mM β-mercaptoethanol) and, where applicable, affinity tags were removed by incubation with the yeast protease Ulp-1 or PreScission Protease at 4ºC overnight. Cleaved proteins were collected in the flow-through during a second step of NiNTA affinity chromatography (HisTrap; GE Healthcare). As a 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 -80ºC.

***Crystallization, X-ray data collection, and structure solution***

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 20ºC 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 *I23* 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 1 mM 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 4ºC 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 20ºC 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 4ºC 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 (62) and XDS (63). 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 (64). For LapD$^{dual}$, initial phases were extended by using the software DM (65), and the first model was built into the electron density map automatically by using the software Buccaneer (66). 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 (64, 67). Data collection and refinement statistics are summarized in Table 2.1. Illustrations were made in Pymol (DeLano Scientific).

***Size Exclusion Chromatography–Coupled Static MALS***

For MALS measurements, purified proteins (20–320 μM, injected concentration) were subjected to size exclusion chromatography (SEC) using a WTC-

030S5 (for $LapD^{dual}$) or WTC-015S5 (for $LapD^{EAL}$) column (Wyatt Technology) 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, for 30 min at room temperature prior to SEC. The SEC system was coupled to an 18-angle static light scattering detector and a refractive index detector (DAWN HELEOS-II and Optilab T-rEX, respectively, Wyatt Technology). Data were collected at 25°C every second at a flow rate of 1.0 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.

***Protein Pull-Down Assay***

$His_6$-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 h at 4°C 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, and 300 mM Imidazole) and visualized using standard denaturing gel electrophoresis (SDS-PAGE). For quantification, gels were stained with SYPRO Ruby gel stain (Molecular Probes) following the manufacturer's directions, and imaged on a VersaDoc MP system (Bio-Rad).

***Strains and Growth Conditions***

Routine culturing of *P. fluorescens* Pf0-1 and *E. coli* was done in lysogeny broth at 30°C and 37°C, respectively. When appropriate, antibiotics were added to the medium at the following concentrations: *E. coli*, 10 μg/ml gentamicin; *P. fluorescens*, 20 μg/ml gentamicin. Plasmids were introduced into *P. fluorescens* by electroporation as described previously (58). K10T medium for biofilm assays was prepared as described previously. K10T-π is 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 in Table 2.2.

***Quantitative Biofilm Formation and Surface Attachment Assays***

To quantify biofilm formation, strains were grown statically for 6 h in K10T-1 medium as described previously (24). Biofilm biomass was stained with 0.1% crystal violet for 15 min, the stain was dissolved, and the biofilm quantified by spectrophotometry, measuring the optical density at 550 nm. We analyzed the effects of inorganic phosphate starvation on attachment by comparing biofilm levels in high-phosphate (K10T-1) and low-phosphate (K10T-π) media over time, as done previously (24).

***Assessment of LapD Protein Levels by Western Blot***

LapD proteins expressed in *P. fluorescens* Pf0-1 were visualized by Western blot as described previously (24), 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 3×10s in 500 µl of buffer (20 mM Tris [pH 8] and 10 mM $MgCl_2$), and pelleting debris at 15,000*g* for 12 min. Samples were normalized to protein concentration using the BCA kit (Pierce).

***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 (18). 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 standards (Biolog Life Science Institute).

***Accession Numbers***

Atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (PDB) under the ID codes 3pjt, 3pju, 3pjv, 3pjw, and 3pjx.

## References


1. Hall-Stoodley L, Costerton J. W, Stoodley P. Bacterial biofilms: from the natural environment to infectious diseases. 2004. Nat Rev Microbiol. 2:95–108.

2. O'Toole G, Kaplan H. B, Kolter R. Biofilm formation as microbial development. 2000. Annu Rev Microbiol. 54:49–79.

3. Parsek M. R, Singh P. K. Bacterial biofilms: an emerging link to disease pathogenesis. 2003. Annu Rev Microbiol. 57:677–701.

4. Hengge R. Principles of c-di-GMP signalling in bacteria. 2009. Nat Rev Microbiol. 7:263–273.

5. Weinhouse H, Sapir S, Amikam D, Shilo Y, Volman G, et al. c-di-GMP-binding protein, a new factor regulating cellulose synthesis in *Acetobacter xylinum*. 1997. FEBS Lett. 416:207–211.

6. Schirmer T, Jenal U. Structural and mechanistic determinants of c-di-GMP signalling. 2009. Nat Rev Microbiol. 7:724–735.

7. Tal R, Wong H, Calhoon R, Gelfand D, Fear A, et al. Three cdg operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. 1998. J Bacteriol.180:4416–4425.

8. Ryan R. P, Fouhy Y, Lucey J. F, Crossman L. C, Spiro S, et al. Cell-cell signaling in*Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. 2006. Proc Natl Acad Sci U S A.103:6712–6717.

9. Simm R, Morr M, Kader A, Nimtz M, Römling U. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. Mol Microbiol. 2004. 53:1123–1134.

10. D'Argenio D, Miller S. Cyclic di-GMP as a bacterial second messenger. 2004. Microbiology.150:2497–2502.

11. Sudarsan N, Lee E. R, Weinberg Z, Moy R. H, Kim J. N, et al. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. 2008. Science. 321:411–413.

12. Amikam D, Galperin M. Y. PilZ domain is part of the bacterial c-di-GMP binding protein. 2006. Bioinformatics. 22:3–6.

13. Ryjenkov DA, Simm R, Romling U, Gomelsky M. The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. 2006. J Biol Chem. 281:30310–30314.

14. Krasteva P. V, Fong J. C, Shikuma N. J, Beyhan S, Navarro M. V, et al. *Vibrio cholerae*VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. 2010. Science. 327:866–868.

15. Hickman J. W, Harwood C. S. Identification of FleQ from *Pseudomonas aeruginosa* as a c-di-GMP-responsive transcription factor. 2008. Mol Microbiol. 69:376–389.

16. Leduc J. L, Roberts G. P. Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of *Xanthomonas axonopodis pv. citri*. 2009. J Bacteriol.191:7121–7122.

17. Chan C, Paul R, Samoray D, Amiot N. C, Giese B, et al. Structural basis of activity and allosteric control of diguanylate cyclase. 2004. Proc Natl Acad Sci U S A.101:17084–17089.

18. De N, Pirruccello M, Krasteva P. V, Bae N, Raghavan R. V, et al. Phosphorylation-independent regulation of the diguanylate cyclase WspR. 2008. PLoS Biol. 6:e67.

19. Lee V. T, Matewish J. M, Kessler J. L, Hyodo M, Hayakawa Y, et al. A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. 2007. Mol Microbiol. 65:1474–1484.

20. Beyhan S, Odell L. S, Yildiz F. H. Identification and characterization of cyclic diguanylate signaling systems controlling rugosity in *Vibrio cholerae*. 2008. J Bacteriol.190:7392–7405.

21. Galperin M. Y, Nikolskaya A. N, Koonin E. V. Novel domains of the prokaryotic two-component signal transduction systems. 2001. FEMS Microbiol Lett. 203:11–21.

22. Christen M, Christen B, Folcher M, Schauerte A, Jenal U. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. 2005. J Biol Chem. 280:30829–30837.

23. Kulasakara H, Lee V, Brencic A, Liberati N, Urbach J, et al. Analysis of *Pseudomonas aeruginosa* diguanylate cyclases and phosphodiesterases

reveals a role for bis-(3′-5′)-cyclic-GMP in virulence. 2006. Proc Natl Acad Sci U S A. 103:2839–2844.

24. Newell P. D, Monds R. D, O'Toole G. A. LapD is a bis-(3′,5′)-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0-1. 2009. Proc Natl Acad Sci U S A.106:3461–3466.

25. Hinsa S. M, Espinosa-Urgel M, Ramos J. L, O'Toole G. A. Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. 2003. Mol Microbiol. 49:905–918.

26. Hinsa S. M, O'Toole G. A. Biofilm formation by *Pseudomonas fluorescens* WCS365: a role for LapD. 2006. Microbiology.152:1375–1383.

27. Gjermansen M, Nilsson M, Yang L, Tolker-Nielsen T. Characterization of starvation-induced dispersion in *Pseudomonas putida* biofilms: genetic elements and molecular mechanisms. 2009. Mol Microbiol. 75:815–826.

28. Monds R. D, Newell P. D, Gross R. H, O'Toole G. A. Phosphate-dependent modulation of c-di-GMP levels regulates *Pseudomonas fluorescens* Pf0-1 biofilm formation by controlling secretion of the adhesin LapA. 2007. Mol Microbiol. 63:656–679.

29. Newell P. D, Boyd C. D, Sondermann H, O'Toole G. A. A C-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage. 2011. PLoS Biol. 9:e587.

30. Falke J. J, Hazelbauer G. L. Transmembrane signaling in bacterial chemoreceptors. 2001. Trends Biochem Sci. 26:257–265.

31. Anantharaman V, Balaji S, Aravind L. The signaling helix: a common functional theme in diverse signaling proteins. 2006. Biol Direct.1:25.

32. Stewart V, Chen L. L. The S helix mediates signal transmission as a HAMP domain coiled-coil extension in the NarX nitrate sensor from *Escherichia coli* K-12. 2010. J Bacteriol.192:734–745.

33. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. 2007. J Mol Biol. 372:774–797.

34. Barends T. R, Hartmann E, Griese J. J, Beitlich T, Kirienko N. V, et al. Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. 2009. Nature. 459:1015–1018.

35. Minasov G, Padavattan S, Shuvalova L, Brunzelle J. S, Miller D. J, et al. Crystal structures of YkuI and its complex with second messenger c-di-GMP suggests catalytic mechanism of phosphodiester bond cleavage by EAL domains. 2009. J Biol Chem. 284:13174–13184.

36. Navarro M. V, De N, Bae N, Wang Q, Sondermann H. Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX. 2009. Structure. 17:1104–1116.

37. Tchigvintsev A, Xu X, Singer A, Chang C, Brown G, et al. Structural insight into the mechanism of cyclic di-GMP hydrolysis by EAL domain phosphodiesterases. 2010. J Mol Biol. 402:524–538.

38. Wassmann P, Chan C, Paul R, Beck A, Heerklotz H, et al. Structure of BeF3- -modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. 2007. Structure. 15:915–927.

39. Rao F, Yang Y, Qi Y, Liang Z. X. Catalytic mechanism of cyclic di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from *Pseudomonas aeruginosa*. 2008. J Bacteriol. 190:3622–3631.

40. Rao F, Qi Y, Chong H. S, Kotaka M, Li B. The functional role of a conserved loop in EAL domain-based cyclic di-GMP-specific phosphodiesterase. 2009. J Bacteriol. 191:4722–4731.

41. De N, Navarro M. V, Raghavan R. V, Sondermann H. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. 2009. J Mol Biol. 393:619–633.

42. De N, Navarro M. V, Wang Q, Krasteva P. V, Sondermann H. Biophysical assays for protein interactions in the Wsp sensory system and biofilm formation. 2010. Methods Enzymol. 471:161–184.

43. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. 2008. Bioinformatics. 24:2780–2781.

44. Reinelt S, Hofmann E, Gerharz T, Bott M, Madden D. R. The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. 2003. J Biol Chem. 278:39189–39196.

45. Moglich A, Ayers R. A, Moffat K. Structure and signaling mechanism of Per-ARNT-Sim domains. 2009. Structure. 17:1282–1294.

46. Chang C, Tesar C, Gu M, Babnigg G, Joachimiak A, et al. Extracytoplasmic PAS-like domains are common in signal transduction proteins. 2010. J Bacteriol. 192:1156–1159.

47. Cheung J, Bingman C. A, Reyngold M, Hendrickson W. A, Waldburger C. D. Crystal structure of a functional dimer of the PhoQ sensor domain. 2008. J Biol Chem. 283:13762–13770.

48. Sevvana M, Vijayan V, Zweckstetter M, Reinelt S, Madden D. R, et al. A ligand-induced switch in the periplasmic domain of sensor histidine kinase CitA. 2008. J Mol Biol. 377:512–523.

49. Hulko M, Berndt F, Gruber M, Linder J. U, Truffault V, et al. The HAMP domain structure implies helix rotation in transmembrane signaling. 2006. Cell. 126:929–940.

50. Airola M. V, Watts K. J, Bilwes A. M, Crane B. R. Structure of concatenated HAMP domains provides a mechanism for signal transduction. 2010. Structure. 18:436–448.

51. Beyhan S, Bilecen K, Salama S. R, Casper-Lindley C, Yildiz F. H. Regulation of rugosity and biofilm formation in *Vibrio cholerae*: comparison of VpsT and VpsR regulons and epistasis analysis of vpsT, vpsR, and hapR. 2007. J Bacteriol. 189:388–402.

52. Suzuki K, Babitzke P, Kushner S. R, Romeo T. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. 2006. Genes Dev. 20:2605–2617.

53. Benach J, Swaminathan S. S, Tamayo R, Handelman S. K, Folta-Stogniew E, et al. The structural basis of cyclic diguanylate signal transduction by PilZ domains. 2007. EMBO J. 26:5153–5166.

54. Ko J, Ryu K. S, Kim H, Shin J. S, Lee J. O, et al. Structure of PP4397 reveals the molecular basis for different c-di-GMP binding modes by Pilz domain proteins. 2010. J Mol Biol. 398:97–110.

55. Guzzo C. R, Salinas R. K, Andrade M. O, Farah C. S. PILZ protein structure and interactions with PILB and the FIMX EAL domain: implications for control of type IV pilus biogenesis. 2009. J Mol Biol. 393:848–866.

56. Laue T. M, Shah B. D, Ridgeway T. M, Pelletier S. L. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding S. E, Rowe A. J, Horton J. C, editors. Analytical ultracentrifugation in biochemistry and polymer science. 1992. Cambridge: Royal Society of Chemistry. pp. 90–125.

57. Stafford W. F, Sherwood P. J. Analysis of heterologous interacting systems by sedimentation velocity: curve fitting algorithms for estimation of sedimentation coefficients, equilibrium and kinetic constants. 2004. Biophys Chem. 108:231–243.

58. Choi K. H, Kumar A, Schweizer H. P. A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: application for DNA fragment transfer between chromosomes and plasmid transformation. 2006. J. Microbiol. Methods. 64:391–397.

59. Monds R. D, Newell P. D, Schwartzman J. A, O'Toole G. A. Conservation of the Pho regulon in *Pseudomonas fluorescens* Pf0-1. 2006. Appl Environ Microbiol. 72:1910–1924.

60. Larkin M. A, Blackshields G, Brown N. P, Chenna R, McGettigan PA. Clustal W and Clustal X version 2.0. 2007. Bioinformatics. 23:2947–2948.

61. Gouet P, Courcelle E, Stuart D. I, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. 1999. Bioinformatics.15:305–308.

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

63. Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr 66: 125-132.

64. Adams P, Grosse-Kunstleve R, Hung L, Ioerger T, McCoy A, et al. (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 58: 1948-1954.

65. 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.

66. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr 62: 1002-1011.

67. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography 60: 2126-2132.

# CHAPTER 3

## Structural characterization of a conserved, calcium-dependent periplasmic protease from *Legionella pneumophila**

### 3.1. Abstract

The bacterial dinucleotide second messenger c-di-GMP has emerged as a central molecule in regulating bacterial behavior, including motility and biofilm formation. Proteins for the synthesis and degradation of c-di-GMP and effectors for its signal transmission are widely used in the bacterial domain. Previous work established that the GGDEF-EAL domain-containing receptor LapD senses c-di-GMP inside the cytosol and relays this signal to the outside by the differential recruitment of the periplasmic protease LapG. Here we identify the core components of an orthologous system in *Legionella pneumophila*. Despite only moderate sequence conservation at the protein level, key features concerning the regulation of LapG are retained. The output domain of the LapD-like receptor from *L. pneumophila*, CdgS9, binds the LapG ortholog involving a strictly conserved surface tryptophan residue. While the endogenous substrate for *L. pneumophila* LapG is unknown, the enzyme processed the corresponding *P. fluorescens* substrate, indicating a common catalytic mechanism and substrate recognition. Crystal structures of *L. pneumophila* LapG provide the first atomic models of bacterial proteases of the DUF920 family and reveal a conserved calcium-binding site important for LapG function.

## 3.1. Introduction

Bacteria sense and respond to their environment by a multitude of physiological programs, allowing them to adapt to changing and often hostile conditions. Biofilm formation is one such mechanism that is used widely by many pathogenic and environmental bacteria (16). Cyclic di-GMP, a molecule unique to bacteria, has emerged as an important intracellular second messenger that regulates the formation of biofilms at multiple levels (17). The majority of the bacterial genomes sequenced to date encode enzymes for the production and turnover of c-di-GMP, diguanylate cyclases with GGDEF domains, and phosphodiesterases with EAL or HD-GYP domains, respectively (12). Receptors for c-di-GMP are a less-well-defined group that includes receiver domains in transcription factors, PilZ domain-containing proteins, riboswitches, and proteins with catalytically inactive GGDEF or EAL domains constituting a distinct class (36). These proteins exploit their degenerate active sites or regulatory c-di-GMP binding sites to sense the cellular concentration of the dinucleotide and to solicit a specific response.

One such receptor, the transmembrane protein LapD from *Pseudomonas fluorescens*, contains both catalytically inactive GGDEF and EAL domains, with the latter serving as the exclusive c-di-GMP binding module (31). Previous work established LapD as a prototypical receptor for mediating inside-out signaling from the cytosol to the periplasm. LapD responds to a rise in c-di-GMP levels triggered by the availability of $P_i$, by undergoing a conformational change from a nucleotide-free, autoinhibited state to a c-di-GMP-activated state (19, 28–30) (Fig. 3.1A). Only the activated state is capable of sequestering a periplasmic protease, LapG, to LapD's

output domain which, in turn, stabilizes LapG's substrate, LapA (30). LapA is a large adhesin protein that is embedded in the outer membrane (18). Proteolytic processing by LapG releases LapA from the membrane, leading to biofilm dispersion when $P_i$ is limiting (30). In conjunction with revealing the signaling and output system for biofilm formation in *P. fluorescens*, we determined the crystal structures of the functional domains of LapD spanning almost the entire receptor (29). These studies revealed the main regulatory features controlling LapD function. Based on a bioinformatic analysis, we also identified orthologous signaling systems in many other bacteria. Two of those have been characterized independently in *Pseudomonas putida* and *Shewanella oneidensis* (14, 37).

Another orthologous system that we predicted, based on bioinformatics, is that of *Legionella pneumophila*, the causative agent of Legionnaires' disease (29). Unlike *Pseudomonas*, *L. pneumophila* is a facultative intracellular bacterium that is also able to grow in biofilms. Its genome encodes 21 predicted proteins with GGDEF and/or EAL domains and a single PilZ protein, and a subset of these have been shown to impact intracellular growth, motility, or biofilm formation (5, 25, 34). However, the underlying signaling mechanisms and networks and their regulation are largely unknown for the majority of these proteins.

In this study, we focused on the periplasmic protease LapG of *L. pneumophila*, which belongs to the domain of unknown function 920 (DUF920) (or COG3672) family, and hence, little is known regarding its molecular mechanism. Sequence and fold recognition methods classified these proteins as bacterial transglutaminase-like cysteine proteases (BTLCPs) and predicted a cysteine-histidine-aspartate (C-H-D)

catalytic triad and core structural motif at the active site (13). O'Toole and colleagues corroborated the notion that LapG functions as a cysteine protease (30). In an effort to further our mechanistic understanding of LapG and related proteases, we determined crystal structures of the LapG ortholog from *L. pneumophila*. In a parallel study, the O'Toole group noted a dependence of *P. fluorescens* LapG activity on calcium ions (4), and the structures allowed us to identify a strictly conserved calcium-binding site in LapG and BTLCPs. In addition, we demonstrate that the *L. pneumophila* LapD-LapG system utilizes an output mechanism similar to that which we previously described in *P. fluorescens* (29, 30).

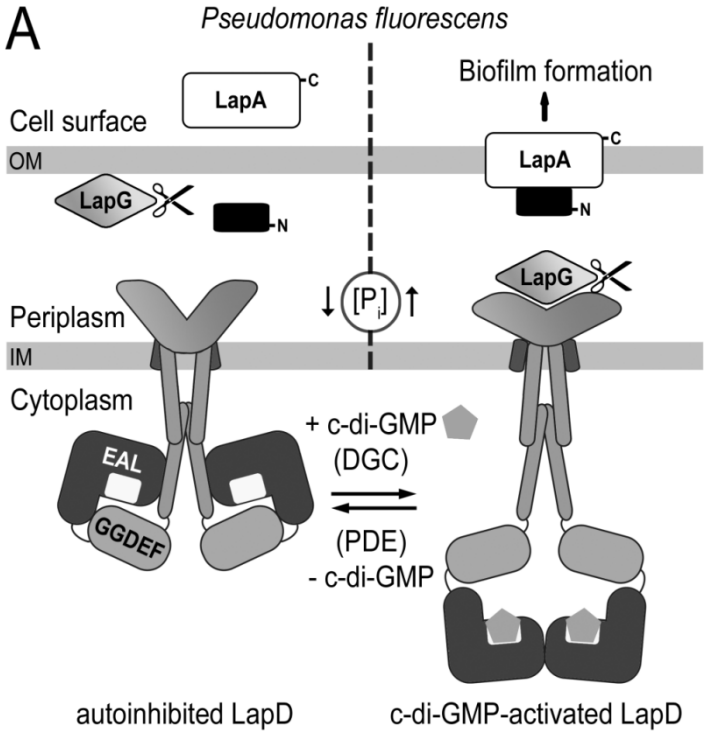



**Figure 3.1: The LapADG signaling system.**
**(A)** Model of LapD-mediated regulation of biofilm formation in *P. fluorescens*. The c-di-GMP receptor LapD localizes to the inner membrane (IM), where it senses cytoplasmic c-di-GMP levels. It controls the stability of the large adhesin LapA in the outer membrane (OM) by sequestration of the periplasmic protease LapG. The underlying signaling pathway is controlled by the availability of $P_i$.

## 3.2. Materials and Methods

### *Protein expression and purification*

DNA fragments encoding *L. pneumophila* LapG lacking the signal peptide (*lpg0828*; residues 52 to 244) and the periplasmic output domain of the LapD ortholog CdgS9 (*lpg0829*; residues 22 to 152) were amplified from genomic DNA by standard PCR and cloned into a bacterial expression vector based on pET28a (Novagen) that adds an N-terminal, cleavable $His_6$-SUMO tag (Table 3.1).

Native and selenomethionine-derivatized proteins were overexpressed in *Escherichia coli* BL21 T7 Express or T7 Express Crystal cells (New England BioLabs), respectively. For the expression of native proteins, cultures were grown at 37°C in terrific broth medium supplemented with 50 µg/ml kanamycin. At an optical density at 600 nm ($OD_{600}$) of ~1, the temperature was reduced to 18°C and protein expression was induced by adding 1 mM IPTG. Selenomethionine-derivatized proteins were expressed in cells grown at 37°C in M9 minimal medium supplemented with 50 µg/ml kanamycin, vitamins (1 µg/ml thiamine and 1 µg/ml biotin), a carbon source (0.4% glucose), trace elements, and amino acids (each of the 20 amino acids at 40 µg/ml, with selenomethionine substituting for methionine). Protein expression was induced at an $OD_{600}$ of 0.4 to 0.5. In both cases, protein expression proceeded for 16 h at 18°C, after which cells were harvested by centrifugation, resuspended in Ni-nitrilotriacetic acid (NTA) buffer A (25 mM Tris-HCl [pH 8.5], 500 mM NaCl and 20 mM imidazole), and flash frozen in liquid nitrogen. Cell suspensions were thawed and lysed by sonication. Cell debris was removed by centrifugation, and the clear lysates were incubated with Ni-NTA resin (Qiagen) that was pre-equilibrated with Ni-NTA

buffer A. The resin was washed with 20 column volumes of buffer A, followed by protein elution with 5 column volumes of Ni-NTA buffer B (25 mM Tris-HCl [pH 8.5], 500 mM NaCl and 300 mM imidazole). The eluted proteins were buffer exchanged into a low-salt buffer (25 mM Tris-HCl [pH 8.5] and 150 mM NaCl) on a fast desalting column (GE Healthcare). Proteins were subjected to size exclusion chromatography on a Superdex 200 column (GE Healthcare) pre- equilibrated with gel filtration buffer (25mMTris-HCl [pH 8.5] and 150 mM NaCl). Where indicated, the $His_6$-SUMO moiety was cleaved off by using the yeast protease Ulp-1 following desalting. Ulp-1, uncleaved protein, and the cleaved fusion tags were removed by Ni-NTA affinity chromatography prior to the final gel filtration. Purified proteins were concentrated on Amicon filters with an appropriate size cut-off to concentrations of 25 mg/ml, flash frozen in liquid nitrogen, and stored at -80°C.

The expression and purification of the corresponding proteins from *P. fluorescens* were described previously (29). The construction, expression, and purification of *P. fluorescens* $LapA^{Nterm}$ were described elsewhere (30). Site-directed mutagenesis was carried out by using the QuikChange kit (Agilent Technologies) and following the manufacturer's instructions.

***Crystallization, data collection, and structure solution***

Crystals were obtained by hanging-drop vapor diffusion mixing equal volumes of protein (10 to 30 mg/ml) and reservoir solution, followed by incubation at 4°C. For the native, apo-state crystal form, the reservoir solution contained 0.14 M ammonium

tartrate dibasic and 20% polyethylene glycol 3350. The selenomethionine-derivatized protein crystallized in the same condition supplemented with 0.1 M Bis-tris (pH 7.0).

Crystals for calcium-bound and EGTA-treated LapG were obtained in the same condition as the apo-protein, supplemented with 2 mM $CaCl_2$ and EGTA, respectively. For cryoprotection, crystals were soaked in reservoir solution supplemented with 25% xylitol. Cryopreserved crystals were flash frozen and stored in liquid nitrogen. For optimal diffraction, crystals of LapG-$Ca^{2+}$ were grown at 20°C and transferred to 4°C for 1 h prior to cryoprotection and freezing. Data on frozen crystals were collected at 100 K using synchrotron radiation at the Cornell High Energy Synchrotron Source (CHESS; Cornell University, Ithaca, NY).

Data reduction was carried out with the software package HKL2000 (33). Experimental phases for the initial structure determination were obtained from single-wavelength anomalous diffraction (SAD) experiments with crystals grown from selenomethionine-derivatized proteins by using the software package PHENIX (2). Refinement in PHENIX and COOT (11) yielded the final models. Data collection and refinement statistics are summarized in Table 3.2. Illustrations were made in Pymol (Schrödinger). Alignments were generated using ClustalW2 (24) and formatted with ESPript (15). Sequence logos were generated using WebLogo (8, 35).

***Protein pull-down assay***

$His_6$-SUMO-tagged *L. pneumophila* LapG or $His_6$-tagged *P. fluorescens* LapG was incubated with Ni-NTA resin (Qiagen) for 1 h at 4°C in binding buffer (25 mM Tris-HCl [pH 8.5], 75 mM NaCl, 25 mM KCl and 40 mM imidazole). Following the

| Table 3.1. Strains and plasmids used in this study. | | |
|---|---|---|
| **Strain or plasmid** | **Genotype or Description** | **Reference** |
| ***Escherichia coli*** | | |
| Top-10 | F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G | Invitrogen |
| T7 Express cells | *fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10* | NEB |
| T7 Express Crystal cells | *fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 metB1 Δ(mcrC-mrr)114::IS10* | NEB |
| **Plasmids** | | |
| pET21a | *E. coli* expression vector, $Amp^R$ | Novagen |
| pET28a | *E. coli* expression vector, $Kana^R$ | Novagen |
| pET28-$His_6$-SUMO | pET28 expressing $His_6$-SUMO, $Kana^R$ | This study |
| pET21-Pfl-LapG | pET21 expressing *P. fluorescens* LapG (Pfl0130), residues 50-251, $Amp^R$ | (1) |
| pSUMO-Lpg-LapG | pET28 expressing *L. pneumophila* LapG (Lpg0828), residues 52-244, fused to $His_6$-SUMO at its N-terminus, $Kana^R$ | This study |
| pSUMO-Pfl-$LapD^{output}$ | pET28 expressing *P. fluorescens* LapG (Pfl0131), residues 52-244, fused to $His_6$-SUMO at its N-terminus, $Kana^R$ | (1) |
| pSUMO-Lpg-$CdgS9^{output}$ | pET28 expressing *L. pneumophila* CdgS9 (Lpg0829), residues 22-152, fused to $His_6$-SUMO at its N-terminus, $Kana^R$ | This study |
| $pLapA^{Nterm}$ | pMQ72 expressing N-term-LapA, $Gm^R$ | (2) |

three washing steps with 5 column volumes of binding buffer each, LapG-bound resin (corresponding to ~50 µg of protein) was incubated with an excess of the untagged output domain (250 µg or 20 µM) of either *L. pneumophila* or *P. fluorescens* LapD for 30 min at 4°C. After the resin was washed three times with 5 column volumes of binding buffer, proteins were eluted from the resin with elution buffer (25 mM Tris-HCl [pH 8.5], 500 mM NaCl and 300 mM imidazole). Eluates were analyzed using standard denaturing SDS-PAGE and Coomassie staining. All incubations were carried out under gentle agitation in spin columns.

***LapA cleavage assay***

Purified *P. fluorescens* LapG variants (wild type or $D^{136}A$; 0.4 µM) and *L. pneumophila* LapG variants (wild type, $D^{136}A$, $E^{138}A$, or $D^{139}A$; 40 to 150 µM) were incubated at the indicated concentrations with purified $LapA^{Nterm}$ (2 µM) in reaction buffer (25 mM Tris-HCl [pH 8.5], 150 mM NaCl and 20 mM $MgCl_2$) overnight at room temperature in the presence or absence of 10 mM EGTA. The reaction products were separated by SDS-PAGE and analyzed by Western blotting using a monoclonal antibody raised against pentahistidine (Qiagen) which was detected by a horseradish peroxidase-coupled anti-mouse antibody. Blots were developed by using enhanced chemiluminescence (GE Healthcare) and exposed to film.

***Size Exclusion Chromatography-Multi Angle Light Scattering***

Size exclusion chromatography-coupled multiangle light scattering (SEC-MALS) measurements were carried out by injecting purified proteins (100 µM) onto a

WTC-030S5 gel filtration column (Wyatt Technology) pre-equilibrated with gel filtration buffer (25 mM Tris-HCl [pH 8.5] and 150 mM NaCl). The SEC system was coupled to an 18-angle, static light scattering detector and a refractive index detector (DAWN HELEOS-II and Optilab T-rEX, respectively; Wyatt Technology). Data were collected at 25°C every second at a flow rate of 1 ml/min and analyzed with the software ASTRA, yielding the molecular mass 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.

***Isothermal Titration Calorimetry***

Isothermal titration calorimetry (ITC) was used to determine the apparent dissociation constant ($K_d$) and the stoichiometry of interactions using a VP calorimeter (Microcal). Calorimetric titrations of calcium (2 mM in the syringe; 10 µl injections) and wild-type or mutant *L. pneumophila* LapG (200 µM in the cuvette) were carried out at 20°C in gel filtration buffer (25 mM Tris-HCl [pH 8.5] and 150 mM NaCl) with a delay of 300s between injections. The data obtained were analyzed by integrating heat effects normalized to the amount of injected protein and curve fitting based on a single-site binding model by using the Origin software package (Microcal).

***Protein structure accession numbers***

Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) and assigned accession numbers 4FGQ, 4FGP, and 4FGO.

## 3.3. Results

### *The Lap operon in L. pneumophila*

We previously predicted the existence of proteins with sequence similarity to *P. fluorescens* LapD and LapG in several bacterial species, including *L. pneumophila* (29). *In silico* genomic analysis indicates that both genes map to an operon containing at least five genes (Fig. 3.2). It encodes a predicted type I secretion outer membrane protein (*lpg0827*/TolC), a LapG-like protease (*lpg0828*), a LapD ortholog with a degenerate GGDEF-EAL domain module (*lpg0829*/CdgS9), and two putative proteins (*lpg0830*, predicted thioesterase/lipase activity; *lpg0831*, predicted flavin-containing monooxygenase). While the functional relevance of the latter two gene products within this cluster remains to be established, a type I secretion system is required in *P. fluorescens* for the translocation of the LapG substrate to the outer membrane (18), and *L. pneumophila* TolC may fulfill a similar function. Considering the limited mechanistic characterization of DUF920-containing proteins and the importance of LapG as a part of a c-di-GMP-dependent signaling system (29, 30), we set out to determine the molecular mechanism and structure of a LapG ortholog.

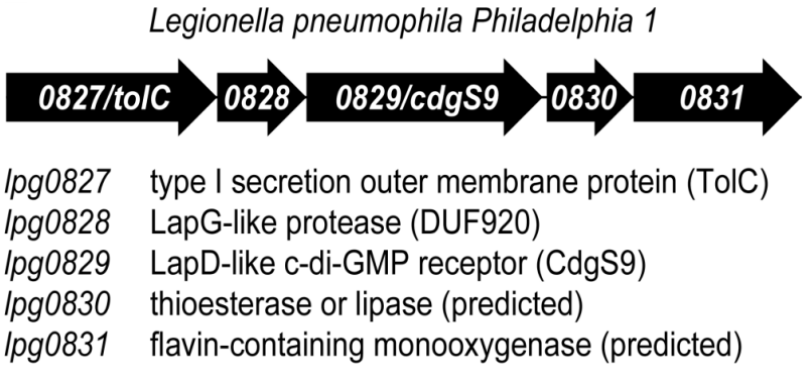



**Figure 3.2: The Lap operon in *L. pneumophila*.**
Genetic map of the LapD ortholog containing operon in *L. pneumophila*.

***Conservation of the LapD-LapG interaction in L. pneumophila***

To establish that the LapG and LapD orthologs indeed form a complex as part of a regulatory system, we purified the periplasmic output domain of *L. pneumophila* CdgS9 (the LapD ortholog) and LapG. Both proteins were expressed most stably with cleavable, N-terminal hexahistidine-SUMO ($His_6$-SUMO) tags. For comparison, we used the respective protein constructs from *P. fluorescens*, with the exception that LapG contained a C-terminal $His_6$ tag instead of the $His_6$-SUMO tag (29). We previously demonstrated that the interaction between *P. fluorescens* LapG and LapD relies on a strictly conserved tryptophan residue that is present at the tip of a beta-hairpin motif in LapD's output domain (29) (Fig. 3.3A). Mutation of this residue to glutamate abolished LapG binding and signaling through LapD, serving as an invaluable specificity control.

$His_6$-tagged (or $His_6$-SUMO-tagged) LapG orthologs were bound to Ni-NTA Sepharose, washed, and incubated with the purified, untagged output domain of LapD or CdgS9, respectively. Proteins were eluted and analyzed by SDS-PAGE. Despite overall low sequence conservation of LapD orthologs (~23% across the entire receptor), *P. fluorescens* LapG adsorbed not only the cognate LapD output domain but also equally efficiently the corresponding domain of the *L. pneumophila* ortholog (Fig. 3.3B). A similar complex formation by *L. pneumophila* LapG and the corresponding LapD output domain was observed. Binding of *L. pneumophila* LapG to the output domain of *P. fluorescens* LapD was detectable but weaker. All interactions were sensitive to a non-conservative (W-to-E) mutation at the critical tryptophan residue at

the center of the output domain beta-hairpin motif, indicating that the mode of binding is specific and conserved across distantly related bacterial species (Fig. 3.3B).

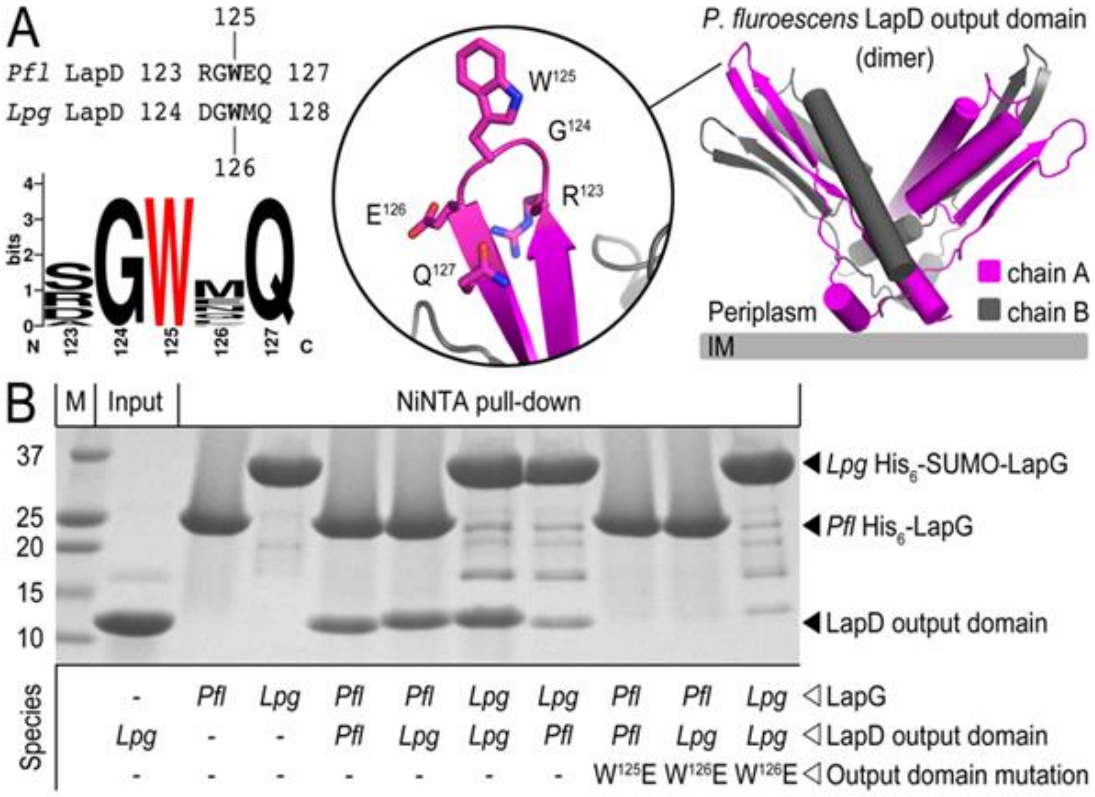



**Figure 3.3: Conserved interaction of LapD and LapG**.
**(A)** Sequence conservation of a loop in the periplasmic output domain of LapD (PDB code 3PJV) that is critical for LapG interaction in *P. fluorescens*.
**(B)** Interaction between LapD's output domain with LapG: Purified, $His_6$-tagged LapG was bound to Ni-NTA and incubated in the absence or presence of the untagged LapD output domain. Proteins from *P. fluorescens* (*Pfl*) or *L. pneumophila* (*Lpg*) were used. A specific output domain mutation ($W^{125}E$ in *P. fluorescens* LapD; $W^{126}E$ in *L. pneumophila* CdgS9) was included. Eluted complexes were analyzed by SDS-PAGE, followed by Coomassie staining.

***The crystal structure of L. pneumophila LapG***

The LapG ortholog from *L. pneumophila* (residues 52 to 244; lacks the signal peptide) was expressed in *E. coli* as a soluble protein and purified by standard liquid chromatography. Upon crystallization (space group $P2_12_12$; two molecules per asymmetric unit), the high-resolution structure was determined by SAD phasing with selenomethionine-substituted protein crystals (Fig. 3.5A; Table 3.2).

The structure reveals a bilobal fold of LapG (Fig. 3.5A). The N-terminal lobe is formed by five α-helices folding into a globular structure. In contrast, the C-terminal lobe consists of a five stranded anti-parallel β-sheet. The three central strands are longer than the two flanking strands. The extreme C terminus folds into a helix that is connected to the bulk of the protein by a flexible linker and is buttressed by the N-terminal lobe via largely hydrophobic interactions. The strictly conserved active site, the C-H-D catalytic triad ($C^{137}$, $H^{172}$, $D^{189}$; Fig. 3.5B), is located at the interface between the two halves of the protein, with the histidine and aspartate residues being contributed by the C-terminal lobe and the catalytic cysteine residue by the central helix α5 of the N-terminal lobe (Fig. 3.5A). The hydroxyl group of a serine residue ($S^{190}$) points toward the active site and engages in a hydrogen bond with $D^{189}$ of the catalytic triad.

The catalytic triad is equally conserved within the LapG subgroup and all BTLCPs. The residue corresponding to position 190 in *L. pneumophila* LapG can be either a serine or an asparagine residue in the LapG subset, as well as the wider BTLCP family (Fig. 3.5B). In order to more globally visualize the conservation of LapG-type proteases, we mapped the conservation scores of 24 LapG orthologs onto

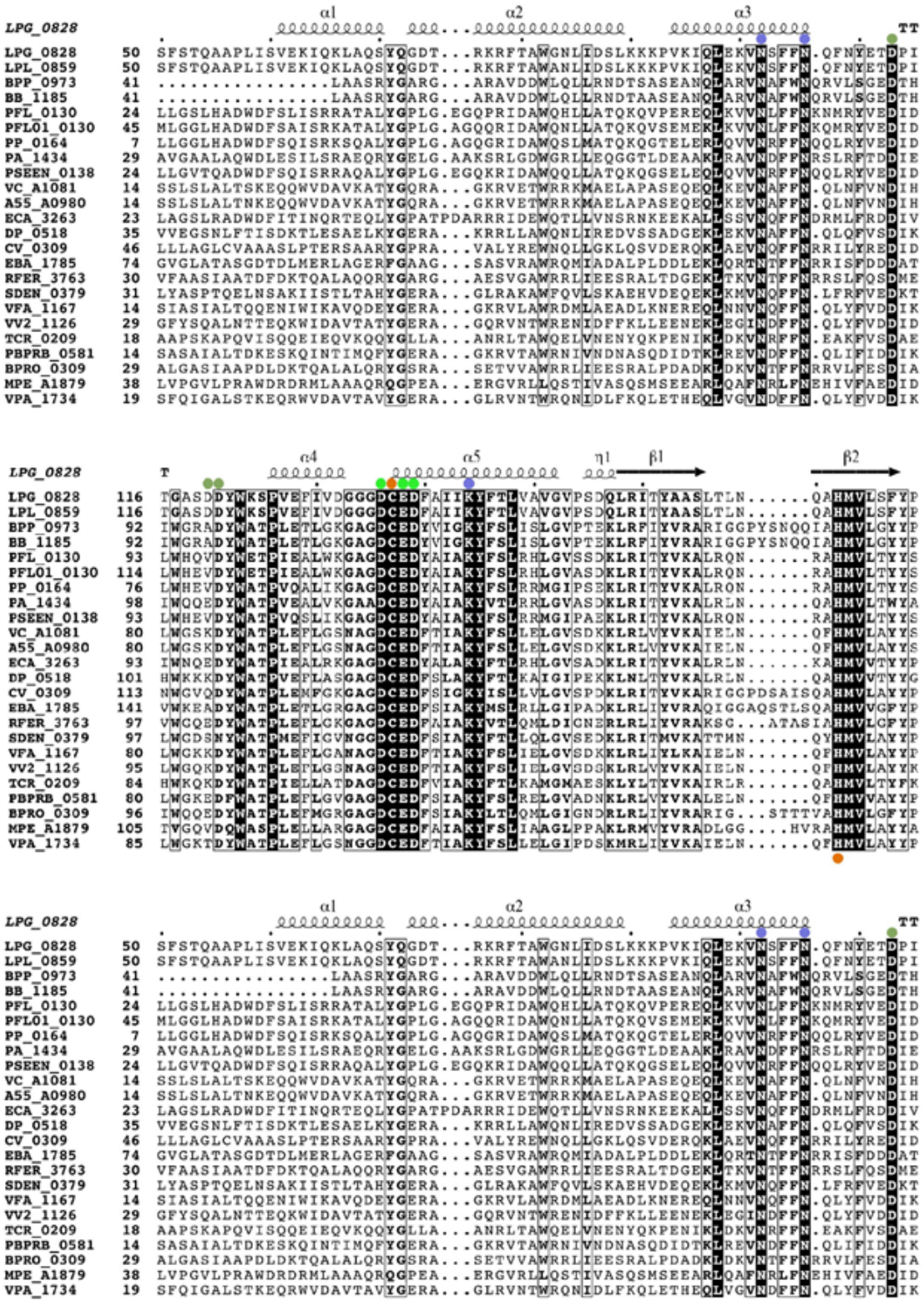
the accessible surface of the protease fold. Orthologs from distantly related species, including *Pseudomonas*, *Legionella*, and *Vibrio* species, were used to create the alignment for this analysis (Fig. 3.4). Interestingly, not only is the catalytic triad strictly conserved, but we also noted a fairly conserved surface patch extending from the active site (Fig. 3.5C). While the functional relevance remains to be established, the hydrophobic nature of this region may suggest a role as an interaction interface, for example, for substrate binding, considering its close proximity to the active site.

A structural role can be attributed to several hydrophilic residues. Consistent with the bioinformatic and modeling study of BTLCPs (13), in addition to the invariant catalytic triad, there are several conserved, polar, or charged residues that form a hydrogen bond network stabilizing some of the core secondary structure elements adjacent to the active site ($N^{91}$, $N^{95}$, $K^{130}$, and $N^{173}$ in BTLCP; $N^{102}$, $N^{106}$, $K^{144}$, and the aforementioned $S^{190}$ in *L. pneumophila* LapG) (Fig. 3.6). Residues with similar function but more specific to the LapG subfamily of BTLCPs are $R^{201}$ and $D^{203}$ (Fig. 3.5B), located in the C-terminal lobe (Fig. 3.5A). Positioned by $D^{203}$, $R^{201}$ coordinates $D^{189}$ of the catalytic triad. Together, these residues are part of the hydrogen bond network that coordinates $D^{189}$ at the active site.

A comparison against the entire PDB using the DALI server (20) was used to identify structurally related proteins. With the structure of *L. pneumophila* LapG as the search model (Fig. 3.7A), we identified eukaryotic protein transglutaminases as some of the closest structural neighbors (Fig. 3.7B; PDB codes 1g0d, 1kv3, 1ggt, and 1nud) (3, 27, 32, 40). In addition, the search identified an arylamine *N*-acetyltransferase (Fig. 3.6C; PDB code 2bsz) (21) and putative bacterial cysteine proteases (Fig. 3.7D and E;

PDB codes 3isr and 3kd4) with deposited structures but no associated publication as proteins that contain a similar fold. All but one protein (the putative bacterial protease with PDB code 3kd4) display a catalytic triad that is conserved in sequence (C-H-D) and position relative to that of LapG, with the cysteine residue being located at the tip of the central helix (Fig. 3.7). Notably, the activity of several transglutaminases depends on the presence of calcium ions (1, 41), and calcium-binding sites have been identified in a subset of crystal structures and by modeling approaches (e.g., PDB code 1nud; Fig. 3.7B) (3, 6, 9). These studies identified three distinct calcium-binding sites in the human transglutaminase 3 enzyme with site 1 being located adjacent to the active site. For transglutaminase, this site has been proposed to stabilize the enzyme, yet a similar role in LapG could not be established. However, *P. fluorescens* LapG is sensitive to EGTA treatment and requires calcium ions for the specific cleavage of its substrate, LapA (4), suggesting a crucial function during catalysis.

**Figure 3.4: Sequence alignment of LapG orthologs.**
A sequence alignment of LapG-like proteins from various species was generated. Key residues discussed in the manuscript are marked and colored according to the main figures. The following sequences were used to generate the alignment: *L. pneumophila* Philadelphila 1 (LPG_0828), *L. pneumophila* Lens (LPL_0859), *Bordetella parapertussis* (BPP_0973), *B. bronchiseptica* (BB_1185), *P. fluorescens* Pf0-1 (PFL_0130, PFL01_0130), *P. putida* KT2440 (PP_0164), *P. aeruginosa* PAO1 (PA_1434), *P. entomophila* (PSEEN_0138), *Vibrio cholerae* O1 biovar El Tor N16961 (VC_A1081), *V. cholerae* 1587 (A55_A0980), *Pectobacterium atrosepticum* (ECA_3263), *Desulfotalea psychrophila* (DP_0518), *Chromobacterium violaceum* (CV_0309), *Aromatoleum aromaticum* EbN1 (EBA_1785), *Rhodoferax ferrireducens* (RFER_3763), *Shewanella denitrificans* (SDEN_0379), *V. fischeri* ES114 (VF_A1167), *V. vulnificus* CMCP6 (VV2_1126), *Thiomicrospira crunogena* (Tcr_0209), *Photobacterium profundum* (PBPRB_0581), *Polaromonas* sp. JS666 (BP RO_0309), *Methylibium petroleiphilum* (Mpe_A1879), *V. parahaemolyticus* (VPA_1734).


α1
α2
α3
α4
α5
η1
β1
β2
LPG_0828
LPG_0828 50 SFSTQAAPLISVEKIQKLAQSYQGDT...RKRFTAWGNLIDSLKKKPVKIQLEKVNSFFN.QFNYETDPI
LPL_0859 50 SFSTQAAPLISVEKIQKLAQSYQGDT...RKRFTAWANLIDSLKKKPVKIQLEKVNSFFN.QFNYETDPI
BPP_0973 41 .....LAASRYGARG...ARAVDDWLQLLRNDTSASEANQLARVNAFWNQRVLSGEDTH
BB_1185 41 .....LAASRYGARG...ARAVDDWLQLLRNDTAASEANQLARVNAFWNQRVLSGEDTH
PFL_0130 24 LLGSLHADWDFSLISRRATALYGPLG.EGQPRIDAWQHLLATQKQVPEREQLKVVNLFFNKNMRYVEDID
PFL01_0130 45 MLGGLHADWDFSAISRKATALYGPLG.AGQQRIDAWQNLLATQKQVSEMEKLKVVNLFFNKQMRYVEDID
PP_0164 7 LLGGLHADWDFSQISRKSQALYGPLG.AGQGRIDAWQSLMATQKQGTELERLQVVNRFFNQQLRYVEDID
PA_1434 29 AVGAALAQWDLESILSRAEQRYGELG.AAKSRLGDWGRLLEQGGTLDEAAKLRAVNDFFNRSLRFTDDIE
PSEEN_0138 24 LLGVTQADWDFSQISRRAQALYGPLG.EGQKRIDAWQQLLATQKQGSELEQLQVVNRFFNQQLRYVEDID
VC_A1081 14 SSLSLALTSKEQQWVDAVKATYGQRA...GKRVETWRRKMAELAPASEQEQLKEVNAFFN.QLNFVNDIH
A55_A0980 14 SSLSLALTNKEQQWVDAVKATYGQRA...GKRVETWRRKMAELAPASEQEQLKEVNAFFN.QLNFVNDIH
ECA_3263 23 LAGSLRADWDFITINQRTEQLYGPATPDARRRIDEWQTLLVNSRNKEEKALLSSVNQFFNDRMLFRDDIV
DP_0518 35 VVEGSNLFTISDKTLESAELKYGERA...KRRLLAWQNLIREDVSSADGEKLEKVNAFFN.QLQFVSDIK
CV_0309 46 LLLAGLCVAAASLPTERSAARYGPRA...VALYREWNQLLGKLQSVDERQKLAEVNQFFNRRILYREDID
EBA_1785 74 GVGLATASGDTDLMERLAGERFGAAG...SASVRAWRQMIADALPLDDLEKLQRTNTFFNRRISFDDDAT
RFER_3763 30 VFAASIAATDFDKTQALAQQRYGARG...AESVGAWRRLIEESRALTDGEKLTKVNTFFNRRSLFQSDME
SDEN_0379 31 LYASPTQELNSAKIISTLTAHYGERA...GLRAKAWFQVLSKAEHVDEQEKLKMVNQFFN.LFRFVEDKT
VFA_1167 14 SIASIALTQQENIWIKAVQDEYGERA...GKRVLAWRDMLAEADLKNEREQLNNVNQFFN.QLYFVDDIK
VV2_1126 29 GFYSQALNTTEQKWIDAVTATYGERA...GQRVNTWRENIDFFKLLEENEKLEGINDFFN.QLYFVDDID
TCR_0209 18 AAPSKAPQVISQQEIEQVKQQYGLLA...ANRLTAWQELVNENYQKPENIKLDKVNRFFN.EAKFVSDAE
PBPRB_0581 14 SASAIALTDKESKQINTIMQFYGERA...GKRVTAWRNIVNDNASQDIDTKLREVNDFFN.QLIFIDDIK
BPRO_0309 29 ALGASIAAPDLDKTQALALQRYGSRA...SETVVAWRRLIEESRALPDADKLDKVNTFFNRRVLFESDIA
MPE_A1879 38 LVPGVLPRAWDRDRMLAAAQRQGPEA...ERGVRLLQSTIVASQSMSEEARLQAFNRLFNEHIVFAEDID
VPA_1734 19 SFQIGALSTKEQRWVDAVTAVYGERA...GLRVNTWRQNIDLFKQLETHEQLVGVNDFFN.QLYFVDDIK
LPG_0828 116 TGASDDYWKSPVEFIVDGGGDCEDFAIIKYFTLVAVGVPSDQLRITYAASLTLN......QAHMVLSFYP
LPL_0859 116 TGASDDYWKSPVEFIVDGGGDCEDFAIIKYFTLVAVGVPSDQLRITYAASLTLN......QAHMVLSFYP
BPP_0973 92 IWGRADYWATPLETLGKGAGDCEDYVIGKYFSLISLGVPTEKLRFIYVRARIGGPYSNQQIAHMVLGYYP
BB_1185 92 IWGRADYWATPLETLGKGAGDCEDYVIGKYFSLISLGVPTEKLRFIYVRARIGGPYSNQQIAHMVLGYYP
PFL_0130 93 LWHQVDYWETPIEALWKGAGDCEDYAIAKYFSLRHLGVSSDKLRITYVKALRQN......RAHMVLTYYS
PFL01_0130 114 LWHEVDYWETPIEALWKGAGDCEDYAIAKYFSLRHLGVASDKLRITYVKALRQN......RAHMVLTYYS
PP_0164 76 LWHEVDYWATPVQALIKGAGDCEDYAIAKYFSLRRMGIPSEKLRITYVKALRQN......RAHMVLTYYS
PA_1434 98 IWQQEDYWATPVEALVKGAADCEDYAIAKYVTLRRLGVASDKLRITYVKALRLN......QAHMVLTWYA
PSEEN_0138 93 LWHEVDYWATPVQSLIKGAGDCEDYAIAKYFSLRRMGIPAEKLRITYVKALRQN......RAHMVLTYYS
VC_A1081 80 LWGSKDYWATPLEFLGSNAGDCEDFTIAKYFSLLELGVSDKKLRLVYVKAIELN......QFHMVLAYYS
A55_A0980 80 LWGSKDYWATPLEFLGSNAGDCEDFTIAKYFSLLELGVSDKKLRLVYVKAIELN......QFHMVLAYYS
ECA_3263 93 IWNQEDYWATPIEALRKGAGDCEDYALAKYFTLRHLGVSADKLRITYVKALRLN......KAHMVVTYYP
DP_0518 101 HWKKKDYWATPVEFLASGAGDCEDFSLAKYFTLKAIGIPEKKLNLTYVKALRLN......QAHMVVTYYG
CV_0309 113 NWGVQDYWATPLEMFGKGAGDCEDFSIGKYISLLVLGVSPDKLRITYVKARIGGPDSAISQAHMVLAYYP
EBA_1785 141 VWKEADYWATPLETLGRGAGDCEDFSIAKYMSLRLLGIPADKLRLIYVRAQIGGAQSTLSQAHMVGFYP
RFER_3763 97 VWCQEDYWATPLEFLGKGAGDCEDFAIAKYVTLQMLDIGNERLRLIYVRAKSG...ATASIAHMVLGFYP
SDEN_0379 97 LWGDSNYWATPMEFIGVNGGDCEDFSIAKYFTLLQLGVSEDKLRITMVKATTMN......QYHMVLAYYP
VFA_1167 80 LWGKKDYWATPLEFLGANAGDCEDFTIAKYFSLIELGVSDKKLRLIYLKAIELN......QFHMVLAYYP
VV2_1126 95 LWGQKDYWATPLEFLGSNAGDCEDFTIAKYFSLIELGVSDSKLRLVYVKAIELN......QFHMVLAYYK
TCR_0209 84 HWKQKDYWATPIELLATDAGDCEDFVIAKYFTLKAMGMAESKLYLTYVKAIRLR......QAHMVLTYFK
PBPRB_0581 80 LWGKEDFWATPLEFLGVGAGDCEDFSIAKYFSLRELGVADNKLRLVYVKALELN......QFHMVVAYYP
BPRO_0309 96 IWQQEDYWATPLEFMGRGAGDCEDFSIAKYLTLQMLGIGNDRLRLIYVRARIG...STTTVAHMVLGFYP
MPE_A1879 105 TVGQVDQWASPLELLARGAGDCEDFAIAKYFSLIAAGLPPAKLRMVIYVRADLGG....HVRAHMVLAYYA
VPA_1734 85 LWGKTDYWATPLEFLGSNGGDCEDFTIAKYFSLLELGIPDSKMRLIYVKAIELN......QFHMVLAYYP
LPG_0828
LPG_0828 50 SFSTQAAPLISVEKIQKLAQSYQGDT...RKRFTAWGNLIDSLKKKPVKIQLEKVNSFFN.QFNYETDPI
LPL_0859 50 SFSTQAAPLISVEKIQKLAQSYQGDT...RKRFTAWANLIDSLKKKPVKIQLEKVNSFFN.QFNYETDPI
BPP_0973 41 .....LAASRYGARG...ARAVDDWLQLLRNDTSASEANQLARVNAFWNQRVLSGEDTH
BB_1185 41 .....LAASRYGARG...ARAVDDWLQLLRNDTAASEANQLARVNAFWNQRVLSGEDTH
PFL_0130 24 LLGSLHADWDFSLISRRATALYGPLG.EGQPRIDAWQHLLATQKQVPEREQLKVVNLFFNKNMRYVEDID
PFL01_0130 45 MLGGLHADWDFSAISRKATALYGPLG.AGQQRIDAWQNLLATQKQVSEMEKLKVVNLFFNKQMRYVEDID
PP_0164 7 LLGGLHADWDFSQISRKSQALYGPLG.AGQGRIDAWQSLMATQKQGTELERLQVVNRFFNQQLRYVEDID
PA_1434 29 AVGAALAQWDLESILSRAEQRYGELG.AAKSRLGDWGRLLEQGGTLDEAAKLRAVNDFFNRSLRFTDDIE
PSEEN_0138 24 LLGVTQADWDFSQISRRAQALYGPLG.EGQKRIDAWQQLLATQKQGSELEQLQVVNRFFNQQLRYVEDID
VC_A1081 14 SSLSLALTSKEQQWVDAVKATYGQRA...GKRVETWRRKMAELAPASEQEQLKEVNAFFN.QLNFVNDIH
A55_A0980 14 SSLSLALTNKEQQWVDAVKATYGQRA...GKRVETWRRKMAELAPASEQEQLKEVNAFFN.QLNFVNDIH
ECA_3263 23 LAGSLRADWDFITINQRTEQLYGPATPDARRRIDEWQTLLVNSRNKEEKALLSSVNQFFNDRMLFRDDIV
DP_0518 35 VVEGSNLFTISDKTLESAELKYGERA...KRRLLAWQNLIREDVSSADGEKLEKVNAFFN.QLQFVSDIK
CV_0309 46 LLLAGLCVAAASLPTERSAARYGPRA...VALYREWNQLLGKLQSVDERQKLAEVNQFFNRRILYREDID
EBA_1785 74 GVGLATASGDTDLMERLAGERFGAAG...SASVRAWRQMIADALPLDDLEKLQRTNTFFNRRISFDDDAT
RFER_3763 30 VFAASIAATDFDKTQALAQQRYGARG...AESVGAWRRLIEESRALTDGEKLTKVNTFFNRRSLFQSDME
SDEN_0379 31 LYASPTQELNSAKIISTLTAHYGERA...GLRAKAWFQVLSKAEHVDEQEKLKMVNQFFN.LFRFVEDKT
VFA_1167 14 SIASIALTQQENIWIKAVQDEYGERA...GKRVLAWRDMLAEADLKNEREQLNNVNQFFN.QLYFVDDIK
VV2_1126 29 GFYSQALNTTEQKWIDAVTATYGERA...GQRVNTWRENIDFFKLLEENEKLEGINDFFN.QLYFVDDID
TCR_0209 18 AAPSKAPQVISQQEIEQVKQQYGLLA...ANRLTAWQELVNENYQKPENIKLDKVNRFFN.EAKFVSDAE
PBPRB_0581 14 SASAIALTDKESKQINTIMQFYGERA...GKRVTAWRNIVNDNASQDIDTKLREVNDFFN.QLIFIDDIK
BPRO_0309 29 ALGASIAAPDLDKTQALALQRYGSRA...SETVVAWRRLIEESRALPDADKLDKVNTFFNRRVLFESDIA
MPE_A1879 38 LVPGVLPRAWDRDRMLAAAQRQGPEA...ERGVRLLQSTIVASQSMSEEARLQAFNRLFNEHIVFAEDID
VPA_1734 19 SFQIGALSTKEQRWVDAVTAVYGERA...GLRVNTWRQNIDLFKQLETHEQLVGVNDFFN.QLYFVDDIK

| Table 3.2: X-ray data collection and refinement statistics. | | | | |
|---|---|---|---|---|
| | LapG, apo (selenomethionine) | LapG, apo (native) | LapG•$Ca^{2+}$ (selenomethionine) | LapG•EGTA (native) |
| **Data Collection** | | | | |
| X-ray source | CHESS, A1 | CHESS, A1 | CHESS, A1 | CHESS, A1 |
| Wavelength (Å) | 0.9771 | 0.9771 | 0.9771 | 0.9771 |
| Space group | $P2_12_12$ | $P2_12_12$ | $P4_32_12$ | C2 |
| Unit cell | | | | |
| a, b, c (Å) | 100.9, 104.5, 43.8 | 100.9, 105.1, 43.3 | 60.0, 60.0, 110.4 | 123.8, 43.7, 100.7 |
| α, β, γ (°) | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 122.1, 90 |
| Resolution (Å)[a] | 50-2.0 (2.04-1.97) | 50-1.7 (1.71-1.65) | 50-1.9 (1.97-1.90) | 50-1.7 (1.79-1.73) |
| No. of reflections | | | | |
| Total | 406,849 (26,517) | 414,388 (26,665) | 40,642 (20,120) | 195,054 (9,471) |
| Unique | 33,274 (3,048) | 55,010 (4,938) | 16,128 (1,315) | 45,446 (3,266) |
| Completeness (%) | 98.9 (92.4) | 96.8 (88.5) | 97.7 (81.3) | 95.3 (69.6) |
| Redundancy | 12.2 (8.7) | 7.5 (5.4) | 25.2 (15.3) | 4.3 (2.9) |
| I/σ(I) | 31.3 (4.8) | 22.8 (1.9) | 49.3 (2.5) | 21.3 (2.0) |
| $R_{meas}$ (%) | 8.5 (33.5) | 7.7 (72.7) | 5.9 (47.5) | 6.1 (37.0) |
| **Refinement** | | | | |
| $R_{work}$ / $R_{free}$ (%) | 18.9 / 22.0 | | 25.9 / 26.7 | 17.7 / 20.7 |
| rms deviations | | | | |
| Bond length (Å) | 0.006 | | 0.008 | 0.006 |
| Bond angles (°) | 1.095 | | 1.111 | 1.067 |
| No. of atoms | | | | |
| Protein | 2910 | | 1440 | 2984 |
| Water | 505 | | 35 | 453 |
| Ave. B-factors ($Å^2$) | | | | |
| Protein | 22.3 | | 62.7 | 25.0 |
| Water | 35.5 | | 49.7 | 32.7 |
| Ramachandram (%) | | | | |
| Favored | 98.6 | | 95.5 | 96.9 |
| Allowed | 1.4 | | 4.5 | 2.5 |
| Disallowed | 0 | | 0 | 0.6 |
| (a) Values in brackets are for the highest resolution bin. | | | | |

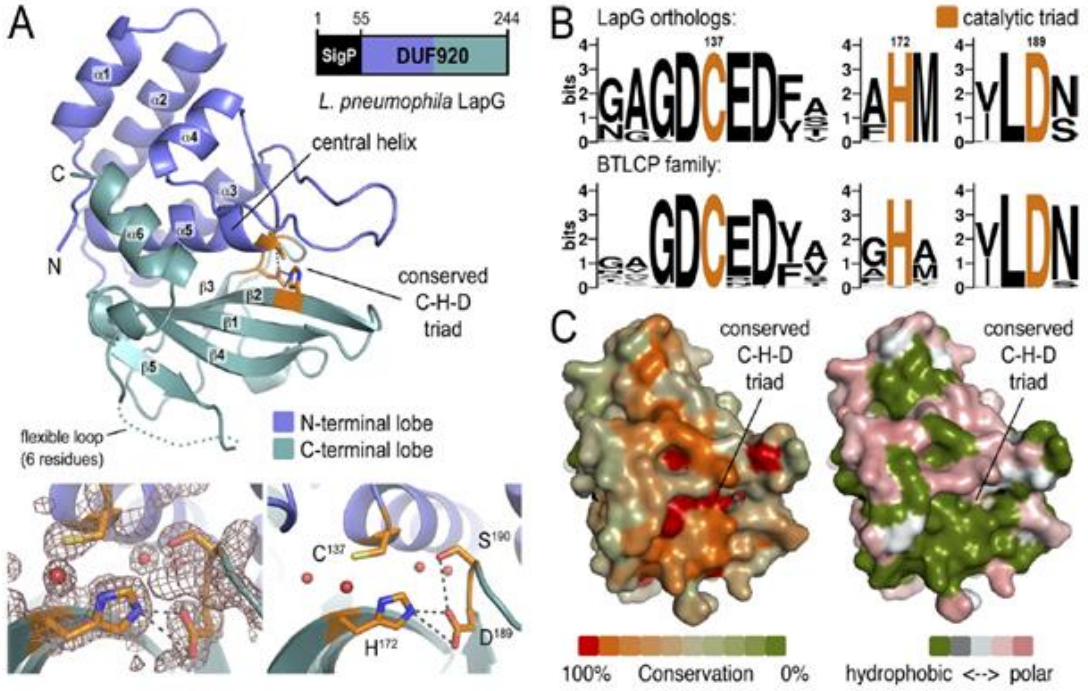



**Figure 3.5: Crystal structure of *L. pneumophila* LapG.**
**(A)** Cartoon presentation of the LapG fold: The N- and C-terminal lobes are shown in slate and cyan, respectively. The conserved catalytic triad (cysteine-histidine-aspartate; C-H-D) is shown as sticks with the carbon atoms in orange. The electron density map at the active site (bottom left inset) has amplitudes of $2|F_o|$ to $|F_c|$, with $F_o$ and $F_c$ being the observed and calculated structure factors. The electron density contour is at 1.8 σ. Water molecules are shown as red spheres. A hydrogen bond network involving residues of the catalytic triad is shown (bottom right inset).
**(B)** Conservation of the catalytic triad based on the sequence alignment shown in Fig. 3.3: Separate sequence logos are shown for the LapG subfamily and the wider family of bacterial transglutaminase-like cysteine proteases (BTLCPs), which was generated based on an alignment that covers a representative, nonredundant set of sequences.
**(C)** Surface conservation and hydrophobicity: Based on the alignment of 24 sequences of LapG orthologs, the sequence conservation was mapped to the accessible surface (left). In the right panel, the surface is colored according to the hydrophobicity of accessible residues. Hydrophobic residues are shown in green, and polar and charged residues are gray and pink, respectively.

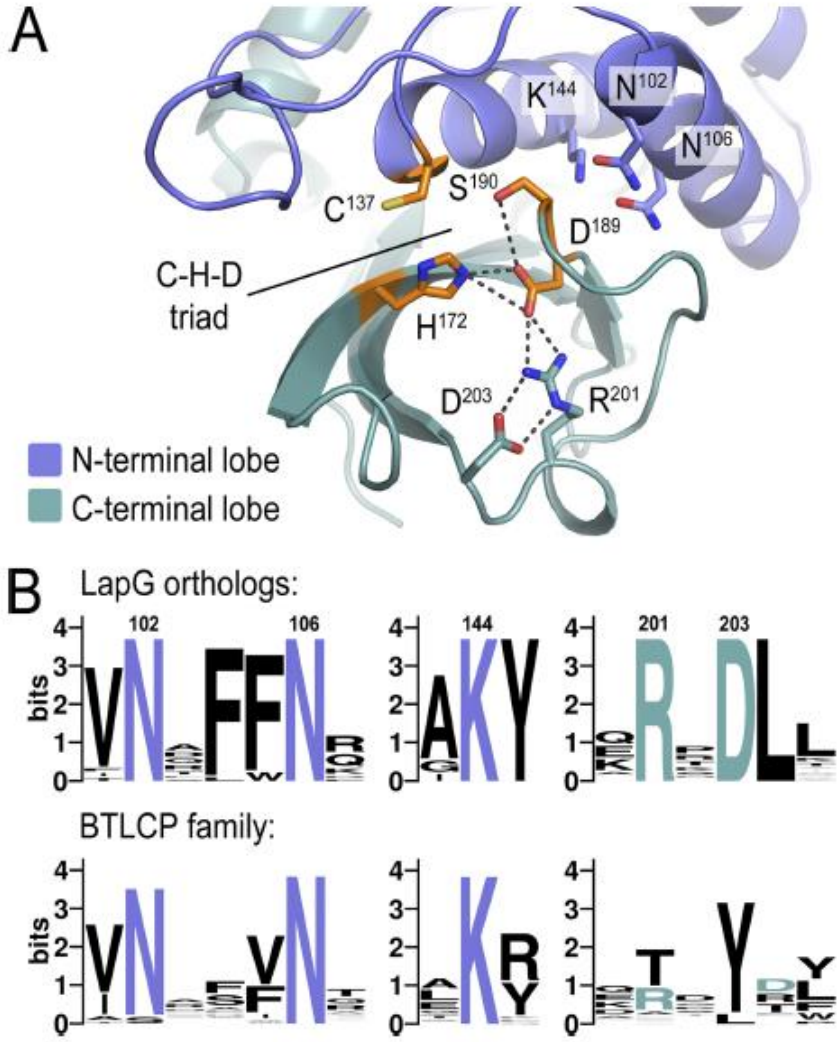



**Figure 3.6: Conservation of a hydrogen bond network around the catalytic triad.**
**(A)** Residues from the N- and C-terminal lobes involved in a hydrogen bond network around the active site are shown as sticks.
**(B)** Sequence conservation: Sequence logos for LapG orthologs and the entire BTLCP family are shown.

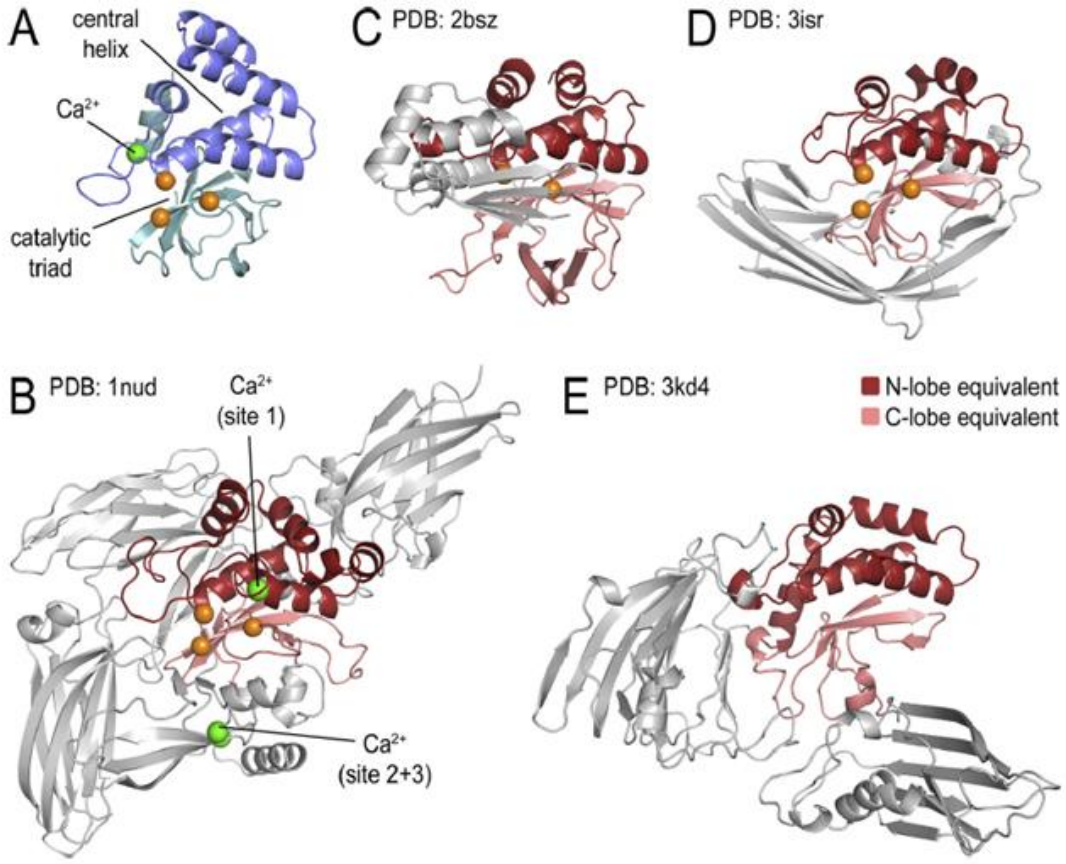



**Figure 3.7: Structural similarities of LapG to transglutaminases, arylamine *N*-acetyltransferases, and putative bacterial proteases.**
A DALI protein structure database search highlighted features structurally conserved between LapG **(A)** and distinct classes of proteins. Regions corresponding to the N and C lobes of LapG, are in red and pink respectively, in panels B to E. Conserved catalytic (C-H-D) triads are shown as orange spheres. Calcium ions are shown as green spheres.
**(B)** Human transglutaminase 3: Z score, 7.4; rmsd, 3.6; aligning fragment spans 135 of 673 residues. Several transglutaminases were identified. The particular structure was chosen to highlight the positions of calcium-binding sites in these enzymes.
**(C)** *Mesorhizobium loti* arylamine *N*-acetyltransferase 1: Z score, 7.8; rmsd, 3.6; aligning fragment spans 127 of 267 residues.
**(D)** Putative cysteine protease from *Cytophaga hutchinsonii*: Z score, 9.4; rmsd, 3.1; aligning fragment spans 121 of 285 residues.
**(E)** Putative protease from *Parabacteroides distasonis* ATCC 8503: Z score, 7.8; rmsd, 3.2; aligning fragment spans 121 of 505 residues.

***Crystal structures of calcium-bound and EGTA-treated LapG***

Incited by the observation that *P. fluorescens* LapG's activity depends on calcium (4) yet no bound calcium ion was apparent in the crystal structure, we initially used pattern prediction algorithms with the apo-LapG structure as the input. The MUG calcium-binding site prediction server (38) identified a patch of four conserved, negatively charged residues adjacent to the catalytic triad that had the potential to accommodate a calcium ion (Fig. 3.8A). While not identical in exact position, the predicted site is as close to the active site as is calcium-binding site 1, which was found to be crucial for transglutaminase function (Fig. 3.7B). In LapG, three of the residues, $D^{136}$, $E^{138}$, and $D^{139}$, flank $C^{137}$ of the catalytic triad, whereas the fourth residue ($D^{120}$) is less conserved and farther away with regard to the primary sequence (Fig. 3.8A). However, other aspartate or glutamate residues are located in close proximity to $D^{120}$ and could function redundantly (Fig. 3.8A). While both the N- and C-terminal lobes contribute residues to the catalytic triad, the calcium-binding site is located entirely in the N-terminal lobe. This motif is not unique to LapG-like proteins but appears to be conserved in the entire BTLCP family on the basis of sequence alignments (Fig. 3.8A), indicating a general molecular mechanism of these proteases.

To confirm the prediction, we crystallized *L. pneumophila* LapG in the presence of calcium ions, yielding a new crystal form. We solved the structure by SAD phasing (space group $P4_32_12$; one molecule per asymmetric unit; Table 3.2). While the maximum resolution was 1.9 Å, crystals diffracted X-rays anisotropically with a resolution of ~2.8 Å in the worst orientation. This observation is consistent with

poor packing interactions along one crystal axis and with the two N-terminal helices being poorly resolved (with high B factors for residues 58 to 88; data not shown).

The overall fold of LapG in the new crystal form was preserved (Fig. 3.9; [rmsd] with or without calcium of 0.6 Å). We observed clear density around the residues predicted to form the calcium-binding site (Fig. 3.8B, inset). A model with a calcium ion at that site refined well. Furthermore, the coordination of the ion by the protein side chains is in good agreement with other calcium binding motifs, revealing seven ionic or polar interactions, including one water-mediated contact and a carbonyl backbone contact with $Y^{122}$ (Fig. 3.8B). To accomplish this, only the side chain of $D^{120}$ had to adopt an alternative rotamer conformation, flipping toward the binding site, while the other three residues are essentially in the same conformation as in the apo-state structure. We noticed only a small shift of the $D^{120}$-presenting loop relative to the body of LapG. However, we cannot distinguish whether this change is calcium induced or due to the altered packing interactions in this crystal form.

We also crystallized LapG treated with EGTA (Table 3.2). This protein preparation crystallized in the original condition. While we obtained crystals within the same space group as LapG without EGTA treatment, it also crystallized in a second space group under identical conditions and similar crystal packing. We solved the structure by molecular replacement using the first LapG crystal structure as the search model. The structure of the EGTA-treated monomer is almost identical to the initial calcium-free state (rmsd of 0.12 Å; Fig. 3.9), and very similar to the calcium-bound state despite differences in crystal contacts.

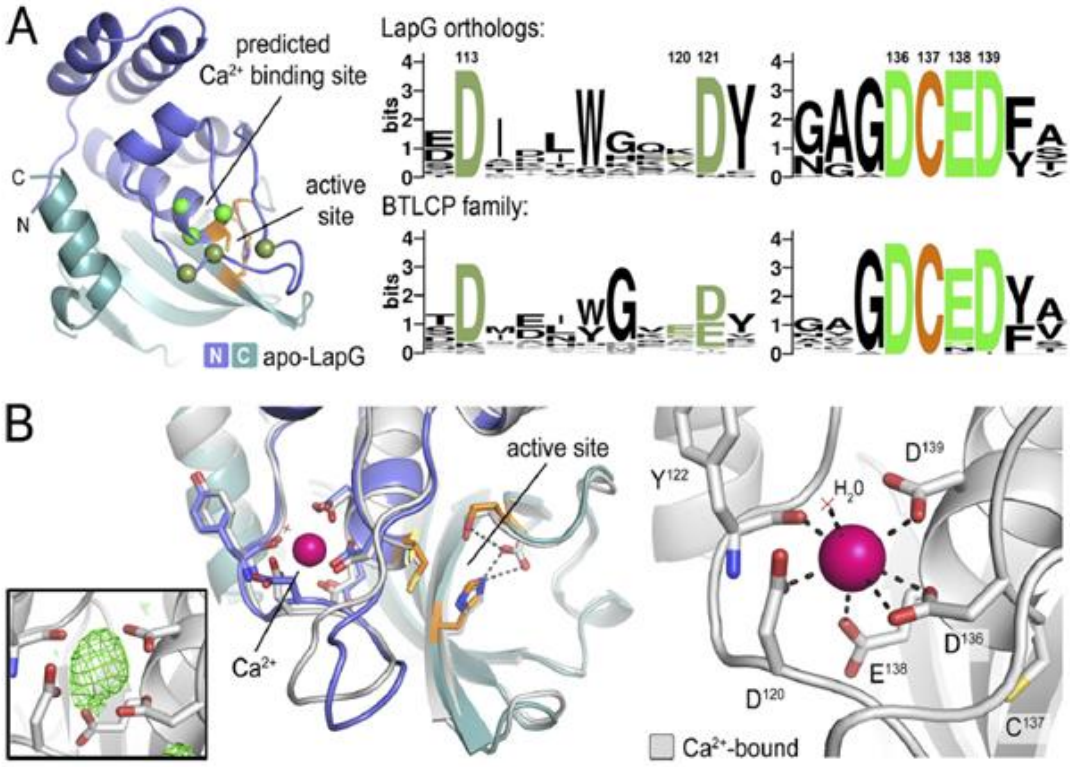



**Figure 3.8: Structural identification of a conserved calcium-binding site in LapG.**
**(A)** Position of a putative calcium-binding site in LapG: Sequence conservation indicates that at least 3 of the 4 residues predicted to form an ion-binding site are conserved in LapG and BTLCPs.
**(B)** Crystal structure of calcium-bound LapG: The crystal structure of LapG (gray) was determined in the presence of calcium ions (pink sphere). The left inset shows uninterpreted electron density that was observed only in the presence of calcium. The density has amplitudes of $|F_o|$ to $|F_c|$ and is contoured at 3.6 σ prior to the inclusion of calcium in the refinement. The structure of non-calcium-bound LapG was superimposed. On the right, a close-up view of the calcium-binding site and calcium coordination is shown.

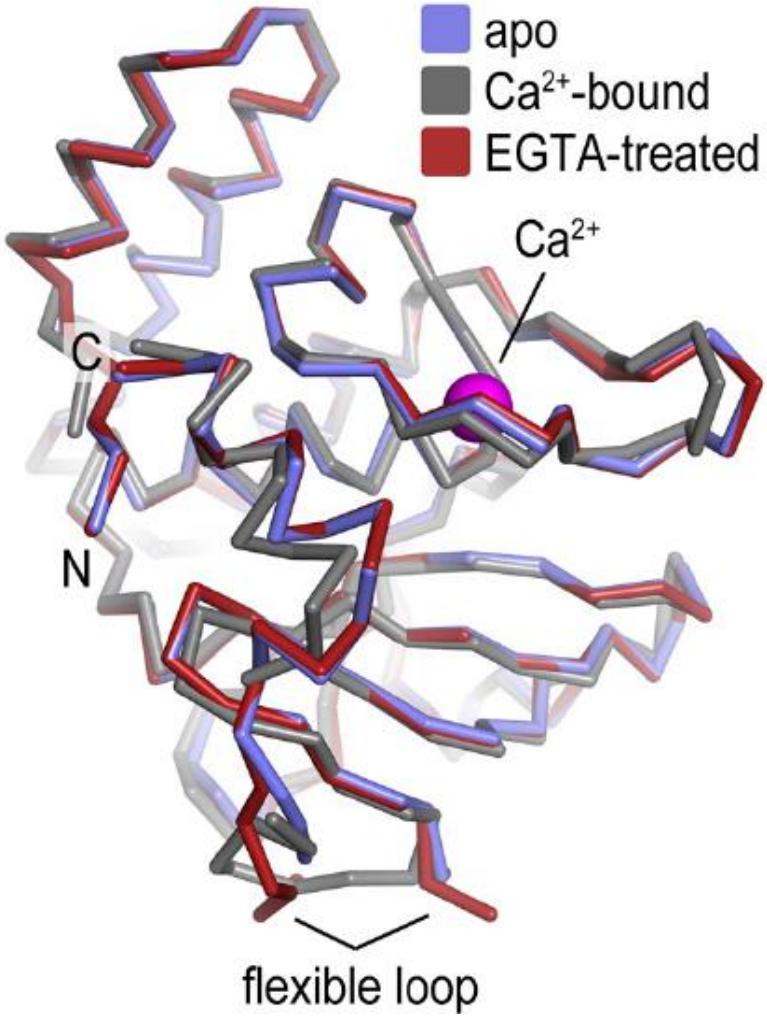



**Figure 3.9: Superposition of *L. pneumophila* LapG crystal structures.**
The protomers of the three crystal structures (protein purified from *E. coli*, "apo"; calcium-bound protein; EGTA-treated protein) were superimposed on all residues. Structural models are shown as $C_\alpha$ traces.

***Calcium binding by L. pneumophila LapG***

To further validate and quantify calcium binding, we turned to isothermal titration calorimetry. The titrations of calcium into a solution of LapG protein are shown in Fig. 3.10. Purified protein bound calcium sub-stoichiometrically with 0.6 calcium molecule per molecule of LapG and with an affinity of 6.2 μM (Fig. 3.9A; Table 3.3). EGTA treatment and repurification of LapG by gel filtration in EGTA-free buffer as the mobile phase yielded protein that bound calcium with the same affinity and negative change in enthalpy as the starting material but with a stoichiometry of close to 1 (Fig. 3.10A; Table 3.3). Together, these data indicate that LapG purifies partially bound to calcium, which can be extracted by EGTA treatment, yielding calcium-free LapG. Based on the titration data, we expect only one calcium-binding site per LapG monomer. Given that no bound calcium was observed in the initial crystal structure, we infer that the crystallization conditions are sensitive to the calcium-bound state of LapG. Finally, mutations in the key residues for calcium coordination ($D^{136}A$, $E^{138}A$, or $D^{139}A$) that were revealed by the crystal structure significantly reduce calcium binding by factors of ~50 (Fig. 3.10B), well within the range one would expect on the basis of similar mutations in other calcium-binding proteins (39). In summary, the structural studies elucidate the overall fold, catalytic core machinery, and a calcium-binding site as conserved features of LapG-like proteases and members of the wider BTLCP superfamily.

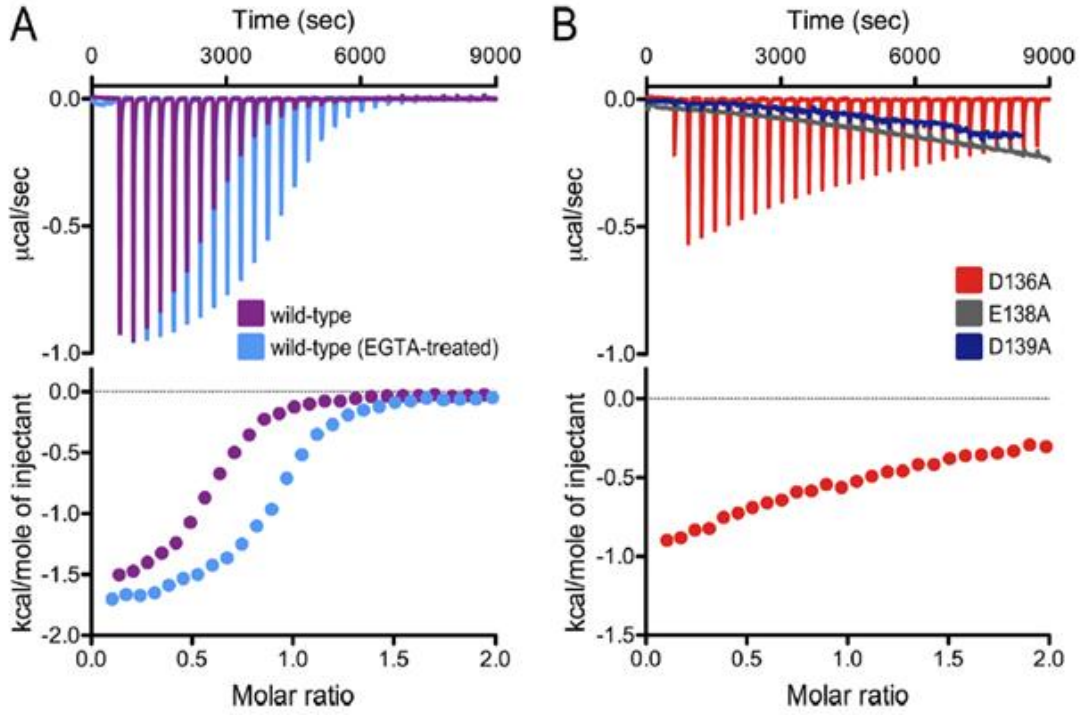



**Table 3.3. Calcium binding measured by isothermal titration calorimetry.**

| ***L. pneumophila* LapG** | $n^b$ | $K_d$ (µM) | $\Delta H$ (cal/mole) | $\Delta S$ (cal/mole/deg) |
|---|---|---|---|---|
| wild-type | 0.59 ± 0.003 | 6.17 ± 0.22 | -1615 ± 10.0 | 18.3 |
| EGTA-treated wild-type | 0.91 ± 0.005 | 4.98 ± 0.32 | -1720 ± 13.0 | 18.4 |
| $D^{136}A$ | 0.95 ± 0.142 | 349.7 ± 59.3 | -2700 ± 569.9 | 6.6 |
| $E^{138}A$ | N.D. | N.D. | N.D. | N.D. |
| $D^{139}A$ | N.D. | N.D. | N.D. | N.D. |

N.D. not detected.

**Figure 3.10: Isothermal titration calorimetry data for calcium binding to LapG.**
**(A)** Binding of calcium to wild-type *L. pneumophila* LapG: Calorimetric titration for calcium (2 mM) titrated into LapG (0.2 mM) is shown (top, raw data; bottom, binding isotherms). LapG purified from *E. coli* was used with or without EGTA treatment during its purification.
**(B)** Binding of calcium to mutant LapG forms: Single-point mutations were introduced into the calcium-binding site of LapG and analyzed as described for panel A. Data analyses are summarized in Table 3.3.

***Oligomeric state of LapG in solution***

It can be assumed that the transmembrane receptor LapD, which interacts with LapG, forms a constitutive homodimer mediated at least by its HAMP domains, a two-helix module that forms a four-helix bundle within a dimeric assembly as the functional unit. This prompted us to analyze the oligomeric state of LapG in the crystals and in solution.

Two different crystal forms of calcium-free LapG were obtained, both with two molecules in their asymmetric units (rmsd of protomer A versus protomer B, 0.08 Å). A potential dimer interface can be identified on the basis of crystal packing interactions. Although the protomer interactions are identical in the two forms (rmsd considering dimeric assemblies, 0.17 Å), the interface is marginal, spanning only 670 $Å^2$. In both crystal forms, protomers interact via a hydrophobic interface formed by the β-sheets of the C-terminal lobe (Fig. 3.11A). Another part of the interface in the crystallographic dimer is presented by the loop protruding from the N-terminal lobe that contains a residue involved in calcium binding ($D^{120}$), in addition to the conserved residue, $D^{113}$. However, the overall packing and calculated binding enthalpies suggest that the dimerization interface is rather weak overall (23).

We next evaluated LapG's oligomerization propensity in solution by using purified proteins. We used SEC-MALS, a method that yields the absolute molecular mass of a protein as it is eluted from a gel filtration column (10). *L. pneumophila* LapG purified from *E. coli* is monomeric in solution with a molecular mass of ~25 kDa (theoretical molecular mass based on sequence, 21.6 kDa; Fig. 3.11B). There was no change in its monomeric state when the protein was incubated with calcium prior to

SEC-MALS analysis or even when calcium was added to the mobile phase, so that calcium was present throughout the experiment.

While we observed only monomeric LapG in solution, the crystallographic dimer interface is moderately conserved across various bacterial strains (Fig. 3.4 and 3.5C), which could suggest that a subset of LapG-like orthologs may utilize this surface. Also, considering that the receptor LapD forms constitutive dimers (29), such a mode of LapG dimerization may still be relevant when LapG is bound to the output domains of intact LapD.

***Conservation of the basic catalytic mechanism of LapG***

Unlike in the case of *P. fluorescens*, the LapG-containing operon in *L. pneumophila* lacks a putative LapA-like substrate protein. Although the identity of the substrate for the *L. pneumophila* ortholog of LapG remains unknown, we evaluated LapG's catalytic activity by using *P. fluorescens* LapA as a model. LapA is a secreted, large adhesin protein. Previous experiments established that shorter constructs including the LapG cleavage site, a conserved TAAG motif, are processed efficiently by *P. fluorescens* LapG *in vitro* (30). The model substrate we use here comprises a 235-residue-long, N-terminal fragment of LapA with a $His_6$ tag at its C-terminus ($LapA^{Nterm}$) (Fig. 3.12). Proteolysis of $LapA^{Nterm}$ by purified *P. fluorescens* LapG produces a fragment of 15 kDa that can be detected by Western blotting with primary antibodies that recognize the $His_6$ tag. $D^{134}A$ mutation in *P. fluorescens* LapG, which corresponds to the calcium-binding mutation $D^{136}A$ in *L. pneumophila* LapG,

abolishes catalytic activity completely, corroborating a calcium-dependent mechanism for LapG function.

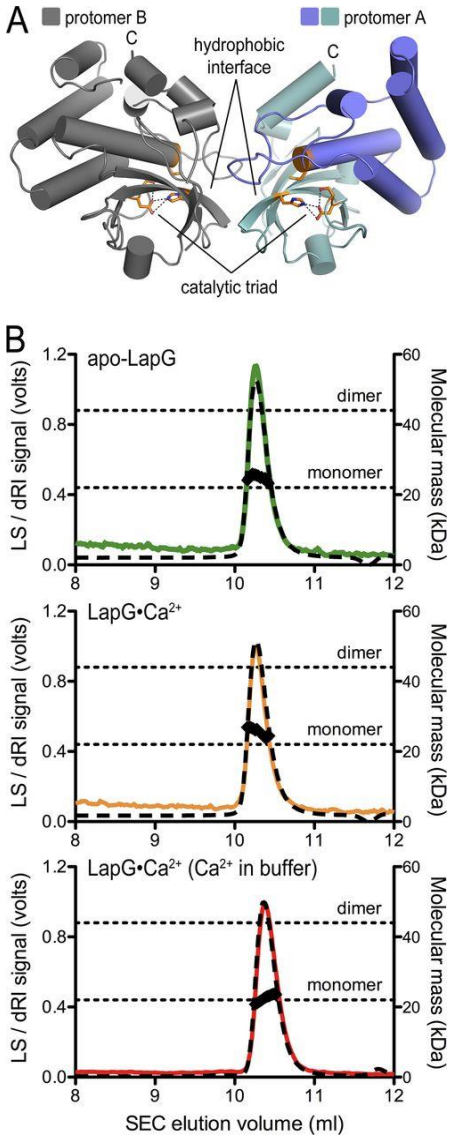



**Figure 3.11: Oligomerization of LapG.**
**(A)** Crystallographic dimer: Analysis of crystal packing interactions reveals a putative LapG dimer that forms via partially hydrophobic interactions between protomers.
**(B)** Oligomeric state in solution: SEC-MALS was used to determine the absolute molecular mass of LapG in solution. The effect of calcium on LapG's oligomeric state was assessed by adding calcium to the purified protein and by additionally adding calcium to the chromatography mobile phase. The averaged molecular masses were 25.1±0.3 kDa for apo-LapG, 25.7±0.3 kDa for LapG+$Ca^{2+}$, and 22.5±0.2 kDa for LapG with calcium in the mobile phase.

Similarly, *L. pneumophila* LapG successfully processed $LapA^{Nterm}$, albeit less efficiently than *P. fluorescens* LapG. In these assays, we used 100-, 250-, or 375-fold more LapG than in the samples incubated with the *P. fluorescens* protease. While we were able to detect a cleavage product with a molecular mass similar to that produced by *P. fluorescens* LapG, indicating that the same cleavage site was recognized, reaction mixtures containing *L. pneumophila* LapG still contained uncleaved substrate. In addition, we observed a second, minor cleavage product with a slightly higher molecular mass at the two highest protease concentrations, which may indicate an aspecific activity under these conditions and with this model substrate. However, proteolysis efficiency was LapG concentration dependent. Furthermore, the calcium-binding single-point mutations ($D^{136}A$, $E^{138}A$, and $D^{139}A$ in *L. pneumophila* LapG) produced markedly reduced catalytic activity. Consistently, both protease orthologs failed to cleave $LapA^{Nterm}$ when incubated with EGTA, indicating that calcium is essential for LapG's function as a periplasmic protease.

Altogether, we validated our prediction of the existing LapD-LapG system in *L. pneumophila*. More importantly, we showed that it utilizes an output and proteolytic mechanism similar to that of its *P. fluorescens* counterparts.

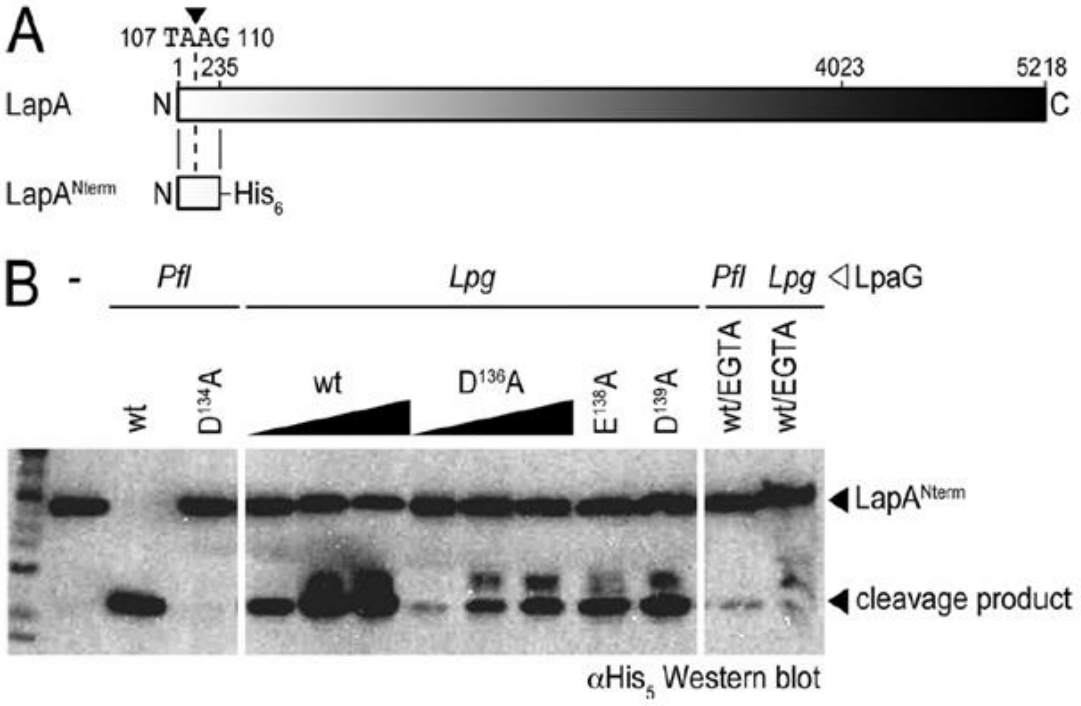



**Figure 3.12: Proteolytic activity of *L. pneumophila* LapG.**
**(A)** A model substrate based on *P. fluorescens* LapA: Previous studies of *P. fluorescens* identified the LapG-cleavage site of LapA within the first 235 residues of the adhesin. A minimal, $His_6$-tagged construct containing this region ($LapA^{Nterm}$) was proteolyzed specifically and efficiently by *P. fluorescens* LapG.
**(B)** *L. pneumophila* LapG activity: Site-specific cleavage of purified $LapA^{Nterm}$ was monitored by using SDS-PAGE, followed by Western blotting with pentahistidine-specific antibodies. *L. pneumophila* LapG with the $E^{138}A$ or $D^{139}A$ mutation was used only at the highest concentration. *P. fluorescens* LapG and a corresponding calcium-binding mutant were included as positive and negative controls, respectively.
Sensitivity to EGTA treatment was also assessed. wt, wild type.

### 3.4. Discussion

The superfamily of DUF920 domain-containing proteins or BTLCPs consists of more than 400 members and is conserved across the bacterial domain (13), yet functional and structural information has been scarce. We recently identified LapG as a functional example of this family, assigning it a biological role as a calcium-dependent, periplasmic protease that is involved in biofilm dispersal of *P. fluorescens* (4, 30). Since we were unable to obtain crystals of *P. fluorescens* LapG thus far, we instead determined the high-resolution crystal structure of the LapG ortholog from *L. pneumophila* in its apo and calcium-bound states. Considering the sequence conservation in functionally important motifs, the structures yield prototypic models of LapG-like proteins and the wider family of BTLCPs. In addition, the structures confirmed their structural relationship to eukaryotic transglutaminases. In analogy to transglutaminases, LapG also shows calcium dependence, yet the exact molecular mechanism remains elusive in both cases despite the available structural data.

Calcium has been known to play myriad roles in modulating protein stability and function. In order for a protein to interact with calcium, mainly through electrostatic forces, it usually has negative charges on its surface. In the case of LapG, the N-terminal lobe displays residues $D^{136}$, $E^{138}$, and $D^{139}$, creating a highly negative surface potential with which $Ca^{2+}$ can coordinate. It is possible that calcium binding to LapG aids in catalysis by providing increased access to the substrate or by altering enzyme structure or dynamics. While three of the four calcium-coordinating residues are strictly conserved, the fourth residue ($D^{120}$) appears more variable. On the other hand, this residue is located in a flexible loop that often harbors other, more conserved

aspartate or glutamate residues in close proximity (Fig. 3.7), which may have a redundant calcium-coordinating function. It is possible that crystallographic packing biases the use of one position over another. The use of alternative aspartate or glutamate residues would likely impact the conformation of the loop, which could potentially alter the active or substrate-binding site. A similar mechanism has been discussed for mammalian tissue transglutaminases (26).

Another effect of calcium binding on enzyme catalysis could simply be a change in electrostatic potential close to the active site. Calcium binding is also known to influence the specific subcellular localization of proteins. For example, in the case of calpain, calcium is required for the association of the protein with cellular membranes (7). In the case here, calcium binding may preferentially place LapG in close proximity to its substrate, LapA, which is anchored in the outer membrane.

The identity of the physiological target of *L. pneumophila* LapG is unknown. While the structural similarity of LapG to transglutaminases may indicate different catalytic functions, no transamidase activity was detected in an *in vitro* assay (D.C. and H.S., unpublished data). However, *L. pneumophila* LapG can cleave $LapA^{Nterm}$ and thus is likely to function as a periplasmic cysteine protease. The reduced activity of *L. pneumophila* LapG for $LapA^{Nterm}$ compared to that of the *P. fluorescens* enzyme could have several causes. Although we previously determined a consensus site for proteolytic cleavage of LapA by LapG, we also noticed that a minimal peptide containing that motif evades processing by the enzyme, suggesting that substrate recognition requires other parts of LapA (Peter D. Newell and G.A.O., unpublished data). Furthermore, *L. pneumophila* has a more complex life-style since unlike *P.*

*fluorescens*, it can grow autonomously or within host cells, increasing the potential target pool vastly. The notion that a putative target may be encoded outside the LapG-containing operon is supported by the presence of an incomplete type I secretion system at this *L. pneumophila* locus. Furthermore, the secretion system may be complemented by subunits encoded outside the *lapDG*-containing operon. Intracellular growth and exposure to different environments (e.g., a different pH or temperature) may also have an impact on the optimal point at which *L. pneumophila* LapG is active. Altogether, these observations hamper the identification of potential *L. pneumophila* substrates by using bioinformatics.

Yet, an interesting speculation is the possibility that LapG or related proteases could degrade host proteins or have a role not only in their own biofilm formation but also in those formed by competing bacteria. Substrate topology and protease access may determine the feasibility of such a model. At least in *P. fluorescens*, LapG localization to the periplasm is crucial for its regulatory effect on biofilm formation, as externally added protease does not appear to alter biofilm phenotypes (C.D.B. and G.A.O., unpublished data).

Recently, a LapADG-like system was identified in *S. oneidensis* (37). In this system, a large outer membrane protein, BpfA, is crucial for surface attachment and subsequent biofilm formation. The gene that encodes BpfA is located in a seven-gene operon which also encodes a type I secretion system, a GGDEF-EAL domain-containing protein similar to LapD, and a hypothetical conserved BTLCP-like protein homologous to LapG. An interesting finding pertains to a positive effect of calcium on *S. oneidensis* biofilm formation, which is BpfA dependent. Considering the presence

of putative calcium-binding sites in LapA and BpfA, these large adhesin proteins could function as sinks for calcium ions that, in turn, are required for proteolytic processing via LapG-like proteases. Such a mechanism is supported by the observation that bacteria have the ability to concentrate calcium in their periplasm to levels that are comparable to or above the measured affinity of LapG for calcium (22). Although the exact mechanism by which these proteins interact and function has yet to be established, it is quite possible that they are regulated in a fashion very similar to that of the Lap system in *P. fluorescens* (4, 29, 30).

Although the corresponding regulators and mechanism remain elusive for the orthologous system in *L. pneumophila*, a growth phenotype in macrophages has been observed upon the overexpression of CdgS9, the LapD ortholog in *L. pneumophila*, indicating that the system is operational in this opportunistic pathogen (25). In particular, cells showed a moderate or severe growth reduction in rich broth or in macrophages, respectively. Our structural and mechanistic studies described here may facilitate future research to uncover the physiological role of this system in *L. pneumophila* and other bacteria. The protease and its interaction with the receptor LapD/CdgS9 also provide an attractive angle for the development of specific inhibitors.

## References


1. Achyuthan KE, Greenberg CS. 1987. Identification of a guanosine triphosphate-binding site on guinea pig liver transglutaminase. Role of GTP and calcium ions in modulating activity. J. Biol. Chem. 262:1901–1906.

2. Adams PD, et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66:213–221.

3. Ahvazi B, Boeshans KM, Idler W, Baxa U, Steinert PM. 2003. Roles of calcium ions in the activation and activity of the transglutaminase 3 enzyme.J. Biol. Chem. 278:23834 –23841.

4. Boyd CD, Chatterjee D, Sondermann H, O'Toole GA. 2012. LapG, required for modulating biofilm formation by *Pseudomonas fluorescens* Pf0-1, is a calcium-dependent protease. J. Bacteriol. 194:4406–4414.

5. Carlson HK, Vance RE, Marletta MA. 2010. H-NOX regulation of c-di-GMP metabolism and biofilm formation in *Legionella pneumophila*. Mol. Microbiol. 77:930 –942.

6. Casadio R, et al. 1999. The structural basis for the regulation of tissue transglutaminase by calcium ions. Eur. J. Biochem. 262:672– 679.
7. Croall DE, Ersfeld K. 2007. The calpains: modular designs and functional diversity. Genome Biol. 8:218.

8. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188 –1190.

9. Datta S, Antonyak MA, Cerione RA. 2006. Importance of Ca(2+)-dependent transamidation activity in the protection afforded by tissue transglutaminase against doxorubicin-induced apoptosis. Biochemistry 45:13163–13174.

10. De N, Navarro MV, Wang Q, Krasteva PV, Sondermann H. 2010. Biophysical assays for protein interactions in the Wsp sensory system and biofilm formation. Methods Enzymol. 471:161–184.

11. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60:2126 –2132.

12. Galperin MY, Nikolskaya AN, Koonin EV. 2001. Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol. Lett. 203:11–21.

13. Ginalski K, Kinch L, Rychlewski L, Grishin NV. 2004. BTLCP proteins: a novel family of bacterial transglutaminase-like cysteine proteinases. Trends Biochem. Sci. 29:392–395.

14. Gjermansen M, Nilsson M, Yang L, Tolker-Nielsen T. 2010. Characterization of starvation-induced dispersion in *Pseudomonas putida* biofilms: genetic elements and molecular mechanisms. Mol. Microbiol. 75:815–826.

15. Gouet P, Courcelle E, Stuart DI, Metoz F. 1999. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15:305–308.

16. Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat. Rev. Microbiol.2:95–108.

17. Hengge R. 2009. Principles of c-di-GMP signalling in bacteria. Nat. Rev.Microbiol. 7:263–273.

18. Hinsa SM, Espinosa-Urgel M, Ramos JL, O'Toole GA. 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. 49:905–918.

*19.* Hinsa SM, O'Toole GA. 2006. Biofilm formation by *Pseudomonas fluorescens* WCS365: a role for LapD. Microbiology 152:1375–1383.

20. Holm L, Rosenstrom P. 2010. Dali server: conservation mapping in 3D. Nucleic Acids Res. 38:W545–W549.

21. Holton SJ, et al. 2005. Structure of *Mesorhizobium loti* arylamine Nacetyltransferase 1. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 61(Pt 1):14 –16.

22. Jones HE, Holland IB, Campbell AK. 2002. Direct measurement of free Ca(2+) shows different regulation of Ca(2+) between the periplasm and the cytosol of *Escherichia coli*. Cell Calcium 32:183–192.

23. Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372:774 –797.

24. Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

25. Levi A, Folcher M, Jenal U, Shuman HA. 2011. Cyclic diguanylate signaling proteins control intracellular growth of *Legionella pneumophila*. mBio 2:e00316–00310.

26. Li B, Cerione RA, Antonyak M. 2011. Tissue transglutaminase and its role in human cancer progression. Adv. Enzymol. Relat. Areas Mol. Biol. 78:247–293.

27. Liu S, Cerione RA, Clardy J. 2002. Structural basis for the guanine nucleotide-binding activity of tissue transglutaminase and its regulation of transamidation activity. Proc. Natl. Acad. Sci. U. S. A. 99:2743–2747.

28. Monds RD, Newell PD, Gross RH, O'Toole GA. 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. 63:656–679.

29. Navarro MV, et al. 2011. Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. PLoS Biol. 9:e1000588. doi:10.1371/journal.pbio.1000588.

30. Newell PD, Boyd CD, Sondermann H, O'Toole GA. 2011. A c-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage. PLoS Biol. 9:e1000587.

31. Newell PD, Monds RD, O'Toole GA. 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. 106:3461–3466.

32. Noguchi K, et al. 2001. Crystal structure of red sea bream transglutaminase. J. Biol. Chem. 276:12055–12059.

33. Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. Macromol. Crystallogr. A 276:307–326.

34. Plate L, Marletta MA**.** 2012. Nitric oxide modulates bacterial biofilm formation through a multicomponent cyclic-di-GMP signaling network. Mol. Cell 46:449–460.

35. Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18:6097– 6100.

36. Sondermann H, Shikuma NJ, Yildiz FH. 2012. You've come a long way: c-di-GMP signaling. Curr. Opin. Microbiol. 15:140 –146.

37. Theunissen S, et al. 2010. The 285 kDa Bap/RTX hybrid cell surface protein (SO4317) of *Shewanella oneidensis* MR-1 is a key mediator of biofilm formation. Res. Microbiol. 161:144 –152.

38. Wang X, Kirberger M, Qiu F, Chen G, Yang JJ. 2009. Towards predicting $Ca^{2+}$-binding sites with different coordination numbers in proteins with atomic resolution. Proteins 75:787–798.

39. Yang JJ, Gawthrop A, Ye Y. 2003. Obtaining site-specific calcium-binding affinities of calmodulin. Protein Pept. Lett. 10:331–345.

40. Yee VC, et al. 1994. Three-dimensional structure of a transglutaminase: human blood coagulation factor XIII. Proc. Natl. Acad. Sci. U. S. A. 91: 7296–7300.

41. 41. Zhang J, Lesort M, Guttmann RP, Johnson GV. 1998. Modulation of the in situ activity of tissue transglutaminase by calcium and GTP. J. Biol. Chem. 273:2288 –2295.

# CHAPTER 4

## A model for the allosteric regulation of periplasmic proteolysis by c-di-GMP

### 4.1. Abstract

Biofilm formation, the switch between a free-swimming to a sessile life-style, is a prevalent adaptational strategy of many bacteria and other microbes as a response to changing environments. In *P. fluorescens*, one of the early pivotal steps, stable adhesion of bacterial cells to a surface, is controlled by the Lap operon. At its center, the transmembrane protein LapD undergoes a conformational change upon binding to the second messenger c-di-GMP in the cytosol, which is propagated through a HAMP domain to the periplasmic domain. This switch renders LapD competent to bind to a periplasmic protease, LapG, which in turn results in the stabilization of LapG's substrate, the large adhesion protein LapA, at the cell surface. The net result of the inside-out signaling event is an increase in biofilm formation. While previous work has revealed regulatory steps, the basic structure of several functional units of the signaling system, and the orthologous LapD-LapG pairs in several other bacteria, the mechanism of switching and LapG recruitment was poorly understood. Here, we report the structure of a complex consisting of the periplasmic output domain of a LapD ortholog bound to LapG. A comparison with the corresponding output domain in its apo-state yields a comprehensive insight into a conserved mechanism for the allosteric regulation of periplasmic proteolysis.

## 4.2. Introduction

Biofilms are complex agglomerations of sessile, microbial cells encased in a self-secreted extracellular matrix composed primarily of exopolysaccharides, proteins and nucleic acids (1). They are prevalent in natural as well as industrial and hospital settings, and can form on a wide range of biotic and abiotic surfaces (1). Biofilm dwelling pathogenic bacteria have been associated with numerous persistent, nosocomial infections in humans such as infection of the ear or urinary tract or in patients suffering from cystic fibriosis (2). Since bacteria in biofilms can withstand antibiotic treatment, it has rendered many clinically relevant antibiotics ineffective in the treatment of biofilm related bacterial infections (3). This is a matter of concern especially in the context of increasing number of multi-drug resistant strains and a rather slow rate of discovery of new antimicrobial agents (4). Hence it becomes crucial to understand the molecular basis of biofilm formation, maintenance and dispersal in order to identify novel targets that could potentially be used for disrupting these bacterial aggregates.

The decision to transition between a planktonic and a biofilm lifestyle is orchestrated by the dinucleotide based bacterial second messenger c-di-GMP, which modulates many different aspects of bacterial physiology (5). Higher cytosolic levels of c-di-GMP are often associated with stable surface attachment and biofilm formation, concomitant with a down-regulation in the expression of genes associated with motility and acute virulence. C-di-GMP is synthesized from two molecules of GTP by proteins with a GGDEF domain, and hydrolyzed by proteins with either an EAL or a HD-GYP domain (6, 7). Over the last decade or so, these c-di-GMP turnover

enzymes have been well characterized both structurally and functionally (6, 8-10). However, the receptors for c-di-GMP, effector molecules that translate the second messenger signal into physiological responses, are a much more diverse group and are slowly but steadily beginning to emerge. C-di-GMP functions as an allosteric modulator of these effector proteins, producing a change in their conformation upon binding and thus transmitting the signal to the target components mainly via macromolecular interactions. Some examples are proteins with PilZ domains, transcription factors like VpsT in *Vibrio cholerae* and FleQ in *Pseudomonas aeruginosa*, riboswitch aptamers of *V. Cholerae*, the cAMP-binding domain in Clp of *Xanthomonas campestris* (11-15). Another group of c-di-GMP receptors are proteins with degenerate GGDEF and/or EAL domains that lost their ability to produce or degrade c-di-GMP, respectively, but retained the ability to bind the second messenger (16, 17).

The transmembrane protein LapD of *Pseudomonas fluorescens* belongs to the last category. It is a modular protein with tandem, enzymatically degenerate GGDEF and EAL domains, a HAMP domain and a periplasmic output domain (17). The output and HAMP domains are separated by a transmembrane domain, which spans the inner bacterial membrane. LapD regulates the localization and maintenance of a large adhesin protein LapA on the bacterial cell surface, in response to changing pools of c-di-GMP in the cytosol, which in turn is modulated by exogenous inorganic phosphate levels (17-19). The periplasmic output mechanism involves a protease, LapG, which is encoded in the same operon as LapD and LapA. LapG is a member of the super family of bacterial transglutaminase like cysteine proteases (BTLCPs) and proteolyzes the

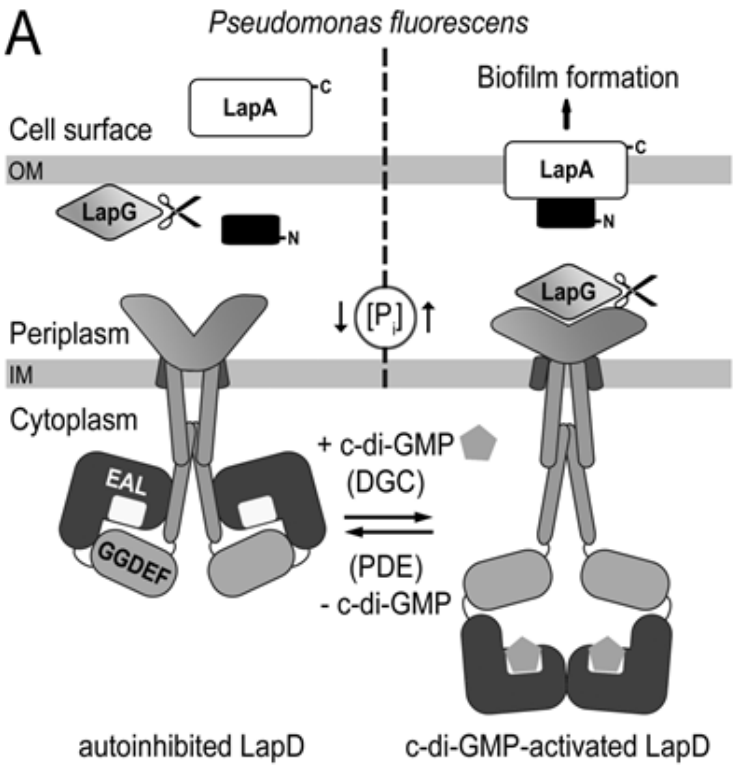
amino terminus of LapA in the presence of calcium, causing LapA release from the cell surface, thereby leading to biofilm dispersal (20, 21). Previous studies have shown that LapD switches between two conformations: It resides in an autoinhibited state at low cellular c-di-GMP levels (the 'off' state), and in an active state when bound to c-di-GMP (the 'on' state) (19). In particular, the dinucleotide binds specifically to the EAL domain of LapD, which releases autoinhibitory interactions and triggers a conformational change in the HAMP domain concomitant with changes in its periplasmic domain (19). The activity states correspond to conformations that differ in their ability to bind to the protease LapG. In the 'off' state, LapD has a low affinity for LapG, enabling the protease to reach its substrate, LapA, which is localized at the outer membrane and is released into the supernatant as a result of the proteolytic event. In the 'off' state, LapG gets recruited to LapD, sequestering it away from the substrate, resulting in a stabilization of LapA at the cell surface. As a consequence, cells adhere to substrates more stably via the LapA adhesin (18, 19).

Our previous structural characterizations identified hallmarks of the functional states of LapD's cytosolic domains. We also determined the first structure of the periplasmic output domain of LapD, which facilitated the identification of a key residue in LapD critical for LapG binding and aided the identification of orthologous systems in several other bacteria (19). Structure-function studies on the *Legionella pneumophila* LapG provided the first atomic models for a bacterial protease of the DUF920 family and revealed a conserved calcium binding site critical for LapG activity (20). Although the endogenous target of *L. pneumophila* LapG remains elusive, the protease was able to process $LapA^{Nterm}$, a minimal fragment of *P.*

*fluorescens* LapA that serves as a model substrate, thereby proving that the basic mechanism of catalysis and substrate recognition remains conserved across the different bacterial species (20). Yet the switching mechanism of LapD's periplasmic domain remained elusive, complicated by the fact of an unusual domain-swap observed in the LapD output domain structure.

Here we present the crystal structures of the isolated output domain of CdgS9 ($CdgS9^{output}$) from *L. pneumophila* and that of a complex between *P. fluorescens* LapG and *L. pneumophila* $CdgS9^{output}$. Besides further proving the conservation of the LapD-LapG interaction, these structures elucidate the mode of LapG recognition by LapD and the basic mechanism by which LapD output is controlled. The structural analysis further reveals structural features controlling LapG activity and suggests anti-cooperativity as a way of fine-tuning the LapD signaling output.

**Figure 4.1: The LapAGD signaling system.**
**(A)** Model of LapD-mediated regulation of biofilm formation in *P. fluorescens* in response to exogenous $P_i$ availability: Inner-membrane localized LapD senses cytosolic c-di-GMP levels and thereby modulates cell-surface associated large adhesin protein LapA by differential recruitment of the periplasmic protease LapG.
**(B)** Genetic map of the LapDG ortholog-containing operon in *L. pneumophila*.
**(C)** Genetic map of the Lap operon in *P. fluorescens*.




B *Legionella pneumophila Philadelphia 1*


| | |
|---|---|
| *lpg0827* | type I secretion outer membrane protein (TolC) |
| *lpg0828* | LapG-like protease (DUF920) |
| *lpg0829* | LapD-like c-di-GMP receptor (CdgS9) |
| *lpg0830* | thioesterase or lipase (predicted) |
| *lpg0831* | flavin-containing monooxygenase (predicted) |

C *Pseudomonas fluorescens Pf-5*


| | |
|---|---|
| *PFL0130* | DUF920 protease (LapG) |
| *PFL0131* | c-di-GMP receptor (LapD) |
| *PFL0132* | hypothetical protein |
| *PFL0133* | large adhesin (LapA) |
| *PFL0134* | TolC-like type I secretion outer membrane protein (LapE) |
| *PFL0135* | type I secretion system ATPase (LapB) |
| *PFL0136* | type I secretion membrane fusion protein (LapC) |

## 4.3. Results

### *Crystal structure of L. pneumophila CdgS9$^{output}$*

We previously reported the structure of the periplasmic domain of *P. fluorescens* LapD (LapD$^{output}$), which allowed us to identify a main residue ($W^{125}$) that is required for LapG binding and signaling output (19). While the overall structure resembled a dimer of PAS domain folds, we were intrigued by the extensively domain swapped conformation it adopts in the crystals. The unusual structure prevented us from inferring a switching mechanism of LapD that would result in the discrete LapG binding states that were apparent in the biochemical data, namely that LapD would only bind to LapG in the periplasm when bound to c-di-GMP. In order to assess whether this structural feature is physiologically relevant and conserved among the other orthologs of this protein, we determined the crystal structure of the corresponding output domain of the *L. pneumophila* LapD ortholog, CdgS9 (residues 22-152; space group $P6_122$; 1 molecule/asymmetric unit; 2.1 Å resolution) (Fig. 4.2A; Table 4.1). The structure was solved by single-wavelength anomalous dispersion (SAD) phasing on crystals grown with selenomethionine-derivatized protein. Despite moderate sequence identity and similarity with *P. fluorescens* LapD$^{output}$, the tertiary structure of the protein remains highly conserved. Yet, in contrast to the domain-swapped conformation observed for *P. fluorescens* LapD$^{output}$, we observed a canonical PAS domain fold in the case of the *L. pneumophila* ortholog, which comprises the molecule in the asymmetric unit (Fig. 4.2A). Considering crystal symmetry mates, we identified a dimeric state as the putative biologically relevant unit (calculated $\Delta G^{int}$ = -20.2 kcal/mol; total buried surface area of 3020 $Å^2$) (22), with the dimer interface

running along a two-fold symmetry axis. The only crossover point between output domain protomers is observed at the termini of the protein chains, which would tether to the transmembrane domains in the full-length protein. Overall, the dimeric model is in topological agreement with this model.

This dimer adopts a 'closed' conformation as opposed to the open, V-shaped form that we observed in case of *P. fluorescens* LapD$^{output}$. As a consequence, the tryptophan residues ($W^{125}$ in *P. fluorescens*; $W^{126}$ in *L. pneumophila*) that form the distal tip of the PAS fold and are a main anchor point for LapG binding (19) are farther apart in the *P. fluorescens* structure than in the one of the *L. pneumophila* LapD orthologs (50 Å *versus* 35 Å). The dimer interface is moderately conserved considering the sequences of 19 putative LapD orthologs (Fig. 4.2B), and is stabilized by hydrophobic as well as polar interactions. A detailed description of the interface will follow the description of CdgS9$^{output}$-LapG complex crystal structure (see below). Owing to the differences between the quaternary structures of the output domains of *P. fluorescens* LapD and *L. pneumophila* CdgS9, it remains to be seen if the 'open' and 'closed' conformations of the output domain might represent different LapG binding states or ortholog-specific variations.

**Table 4.1: X-ray data collection and refinement statistics.**

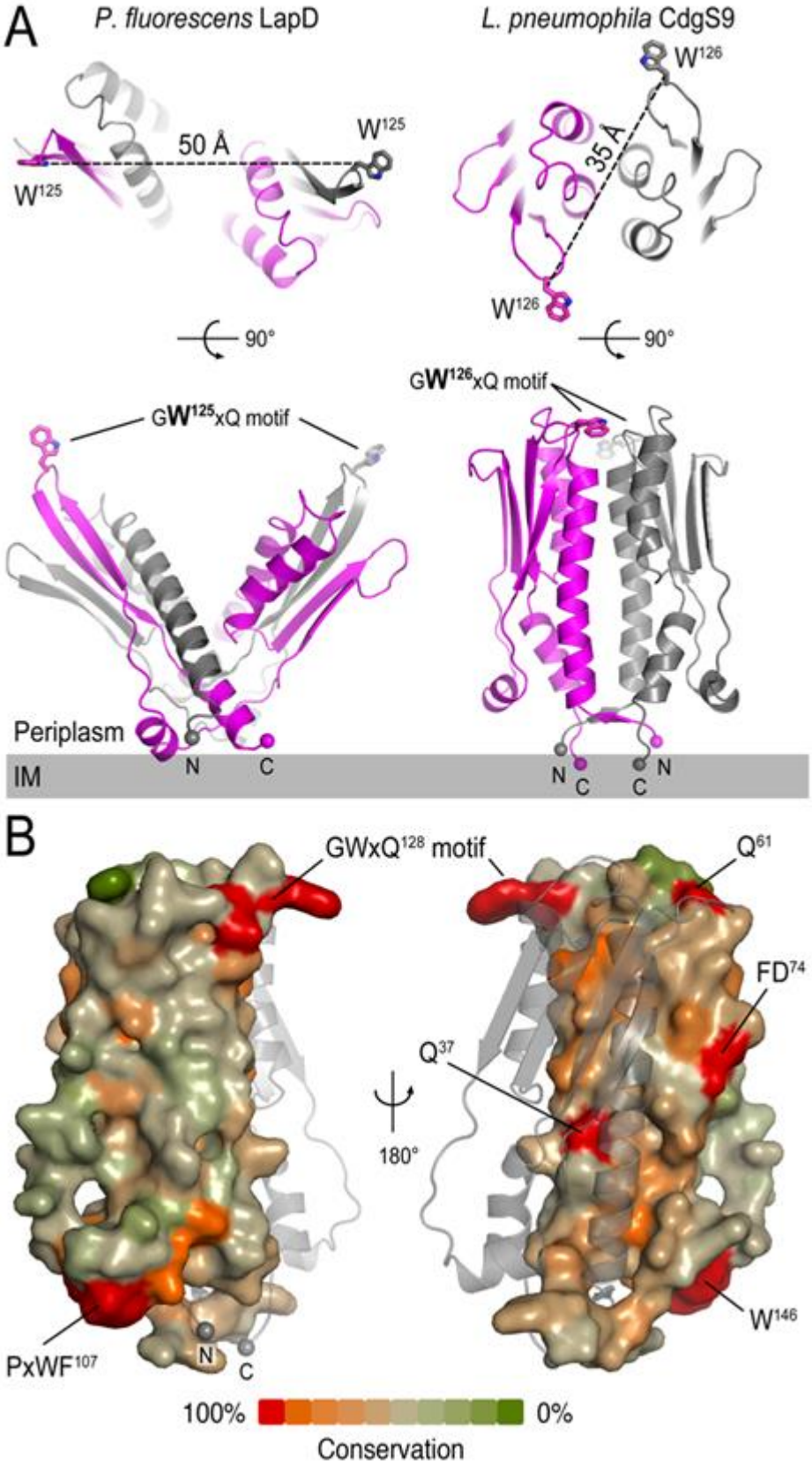
| | *Lpg* CdgS9-*Pfl* LapG | *Lpg* CdgS9-*Pfl* LapG | *Lpg* CdgS9$^{output}$ | *Lpg* CdgS9$^{dual}$ |
|---|---|---|---|---|
| | (selenomethionine) | (native) | (selenomethionine) | (selenomethionine) |
| **Data Collection** | | | | |
| X-ray source | CHESS, A1 | CHESS, A1 | CHESS, A1 | CHESS, A1 |
| Wavelength (Å) | 0.9771 | 0.9771 | 0.9771 | 0.9771 |
| Space group | $P2_1$ | $P2_1$ | $P6_122$ | $P2_12_12_1$ |
| Unit cell | | | | |
| a, b, c (Å) | 59.1, 74.7, 62.9 | 75.9, 73.7, 88.2 | 61.6, 61.6, 147.5 | 49.0, 135.4, 171.4 |
| α, β, γ (°) | 90, 101.8, 90 | 90, 92.9, 90 | 90, 90, 120 | 90, 90, 90 |
| Resolution (Å)[a] | 50-2.3 (2.36-2.28) | 50-2.1 (2.18-2.10) | 50-2.1 (2.22-2.14) | 50-2.2 (2.32-2.24) |
| No. of reflections | | | | |
| Total | 164,096 (12,523) | 238,369 (16,845) | 214,519 (17,521) | 725,165 (40,104) |
| Unique | 22,898 (2,319) | 55,798 (5,264) | 9,799 (942) | 55,207 (5,013) |
| Completeness (%) | 94.1 (95.6) | 98.3 (93.7) | 99.9 (100.0) | 98.8 (91.2) |
| Redundancy | 7.2 (5.4) | 26.1 (25.4) | 21.9 (18.6) | 13.1 (8.0) |
| I/σ(I) | 20.1 (2.3) | 15.7 (2.5) | 35.5 (10.1) | 21.3 (2.6) |
| $R_{meas}$ (%) | 8.9 (67.1) | 8.2 (38.7) | 7.7 (30.9) | 10.3 (53.2) |
| **Refinement[(b)]** | | | | |
| $R_{work}$ / $R_{free}$ (%) | 22.4 / 26.5 | 16.6 / 22.0 | 21.1 / 24.0 | 18.0 / 21.9 |
| rms deviations | | | | |
| Bond length (Å) | 0.009 | 0.008 | 0.008 | 0.022 |
| Bond angles (°) | 1.255 | 1.137 | 1.210 | 1.410 |

(a) Values in brackets are for the highest resolution bin. (b) Work-in-progress.

**Figure 4.2: Structural comparison of periplasmic output domain from *P. fluorescens* LapD and *L. pneumophila* CdgS9**

**(A)** Two orthogonal views of each output domain structure are shown in ribbon representation, with the two protomers colored pink and gray, respectively. *P. fluorescens* $LapD^{output}$ has an open, V-shaped conformation and is extensively domain swapped (19), while *L. pneumophila* $CdgS9^{output}$ adopts a closed conformation with the arms of each protomer held parallel. Distances between the tryptophan residues in the conserved GWxQ motif of each protomer are indicated. The relative position of the inner cell membrane (gray bar) and connection to the flanking transmembrane (TM) helices are indicated.

**(B)** The sequences of 19 orthologs of LapD were aligned and the sequence conservation was mapped on to the accessible surface. The surface is colored according to the degree of conservation of accessible residues.


A
P. fluorescens LapD
L. pneumophila CdgS9
W125
50 Å
W125
W126
35 Å
W126
90°
90°
GW125xQ motif
GW126xQ motif
Periplasm
IM
N
C
N
C
C
N
B
GWxQ128 motif
Q61
FD74
Q37
180°
PxWF107
N
C
W146
100%
0%
Conservation

***Crystal structure of the CdgS9$^{output}$–LapG complex***

We obtained clarity in this regard for the case of CgdS9 by determining the structure of CdgS9$^{output}$ in complex with LapG. In spite of our best efforts, we have so far been unable to either express or crystallize either of the homogeneous complexes. Since we had observed in the past that *P. fluorescens* LapG could bind *L. pneumophila* CdgS9$^{output}$ *in vitro*, we went ahead and determined the crystal structure of the heterologous complex containing *L. pneumophila* CdgS9$^{output}$ bound to *P. fluorescens* LapG by SAD phasing (space group $P2_1$; 2.1 Å resolution) (Table 4.1; Fig. 4.3A). The final model contained two virtually identical complexes per asymmetric unit, with each complex consisting of two CdgS9$^{output}$ domains and one LapG molecule. Based on this structure, we were not only able to visualize the interaction interface between the output domain of CdgS9 (the LapD ortholog) and LapG as well as the first atomic model of *P. fluorescens* LapG, but it also revealed an active state of a LapD ortholog. In comparison with the previously determined apo-CdgS9$^{output}$ structure, a model for the allosteric regulation of periplasmic PAS domains and the control of periplasmic proteolysis via LapG could be deduced. Each of the individual discoveries will be discussed in the following sections.

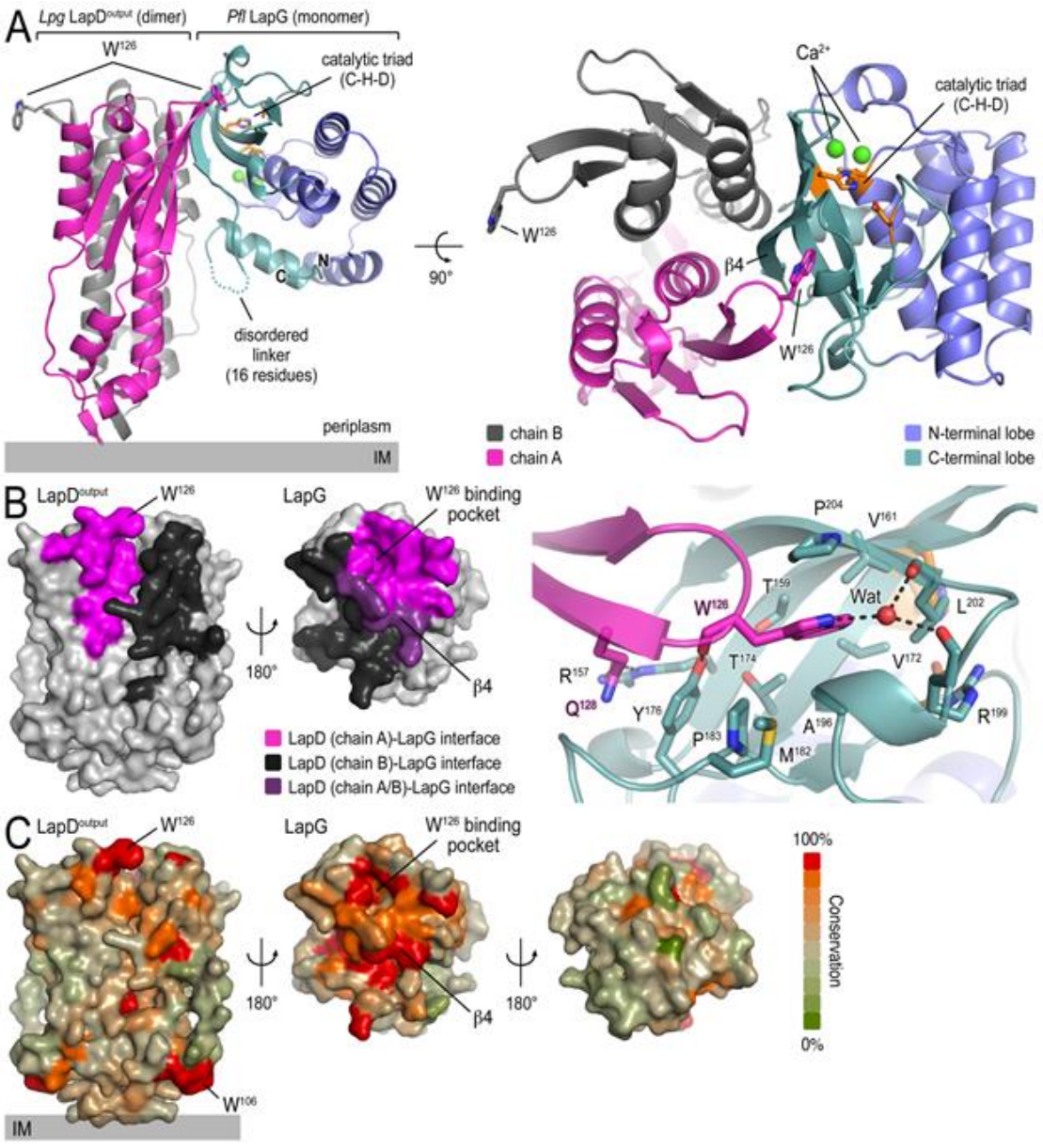
***The output domain-LapG interface***

Similar to the structure of apo-CdgS9$^{output}$, the output domain in the complex also forms dimers, using the same interface and overall topology (differences between the interfaces will be discuss in detail in a later section; see below). Likewise, *P. fluorescens* LapG adopts a bilobal fold with an alpha-helical N-terminal lobe and a C-

terminal lobe consisting of beta-sheets, which is by and large comparable to the structure of the *L. pneumophila* ortholog determined previously (minor mechanistic differences will be discussed below) (20).

The complex between $CdgS9^{output}$ (residues 22-152) and LapG (residues 50-251) is asymmetric with a stoichiometry of 2:1, i.e. a $CdgS9^{output}$ dimer interacts with a LapG monomer in the crystal lattice. The interfacial surface area buried between the two proteins in the complex is rather expansive (6560 $Å^2$ with a calculated $\Delta^iG$ = -52.3 kcal/mol) (22). In the crystal structure of the complex, β4 in the C-terminal lobe of LapG is buttressed against the $CdgS9^{output}$ dimer interface, in particular between the β-hairpin loop that presents the conserved $GW^{126}xQ$ motif of one output domain protomer and α2 in chain B of the other protomer. The highly conserved tryptophan residue in $CdgS9^{output}$ ($W^{126}$), which was previously shown to be integral for LapG binding in the *P. fluorescens* ortholog, fits snugly into a predominantly hydrophobic binding pocket at the bottom of LapG's C-lobe. There are several accessory hydrogen bond involving both side chain as well as backbone atoms and spanning a large surface LapG and $CdgS9^{output}$ (Fig. 4.3B). Also, while sequence conservation on the LapG-interacting surface of $CdgS9^{output}$ is moderate but apparent, the concentration of conserved residues on the corresponding interfacial area on LapG suggests that this interface is relevant for the 20 orthologous systems used to generate the sequence alignment, which formed the basis for this analysis (Fig. 4.3C). Although the surface on the LapD ortholog appears less conserved, a pronounced shape complementarity between the interacting surfaces may contribute to the stable recruitment of LapG to the periplasmic output domain of LapD orthologs.

**Figure 4.3: Crystal structure of *P. fluorescens* LapG and *L. pneumophila* CdgS9$^{output}$ complex.**
**(A)** The LapG-CdgS9$^{output}$ complex is shown as a ribbon representation, with the two protomer chains of CdgS9$^{output}$ colored in pink and gray, while the N-and C- terminal lobes of LapG are shown in slate and cyan respectively. The conserved catalytic triad (cysteine-histidine-aspartate; C-H-D) is shown as sticks with the carbon atoms in orange. The highly conserved tryptophan residue ($W^{126}$) in each protomer of CdgS9$^{output}$ is presented as sticks and calcium ions are shown as spheres in green. The relative position of the inner cell membrane (gray bar) is indicated. Two orthogonal views are shown.
**(B)** Surface representation of the interaction interface between CdgS9$^{output}$ and LapG: The interfaces between LapG and the CdgS9 half-sides (pink and black) as well as a groove presented at the CdgS9$^{output}$ dimer interface (purple) were mapped onto the accessible surface area of the individual proteins (left panel). The binding pocket for CdgS9 $W^{126}$ in LapG is shown as a close-up view (right panel).
**(C)** Surface conservation of CsgS9$^{output}$ and LapG interaction interface: Based on the alignment of 19 sequences of LapD and LapG orthologs, the sequence conservation was mapped on to the accessible surface of each protein. The surface is colored according to the degree of conservation of accessible residues. For LapG, two views, separated by a 180° rotation, are shown.


A
Lpg LapDoutput (dimer)
Pfl LapG (monomer)
W126
catalytic triad
(C-H-D)
N
C
disordered
linker
(16 residues)
periplasm
IM
90°
Ca2+
catalytic triad
(C-H-D)
W126
β4
W126
chain B
chain A
N-terminal lobe
C-terminal lobe
B
LapDoutput
W126
LapG
W126 binding
pocket
180°
β4
LapD (chain A)-LapG interface
LapD (chain B)-LapG interface
LapD (chain A/B)-LapG interface
P204
V161
T159
W126
Wat
L202
T174
V172
R157
Q128
Y176
P183
A196
M182
R199
C
LapDoutput
W126
LapG
W126 binding
pocket
180°
β4
180°
100%
Conservation
0%
W106
IM

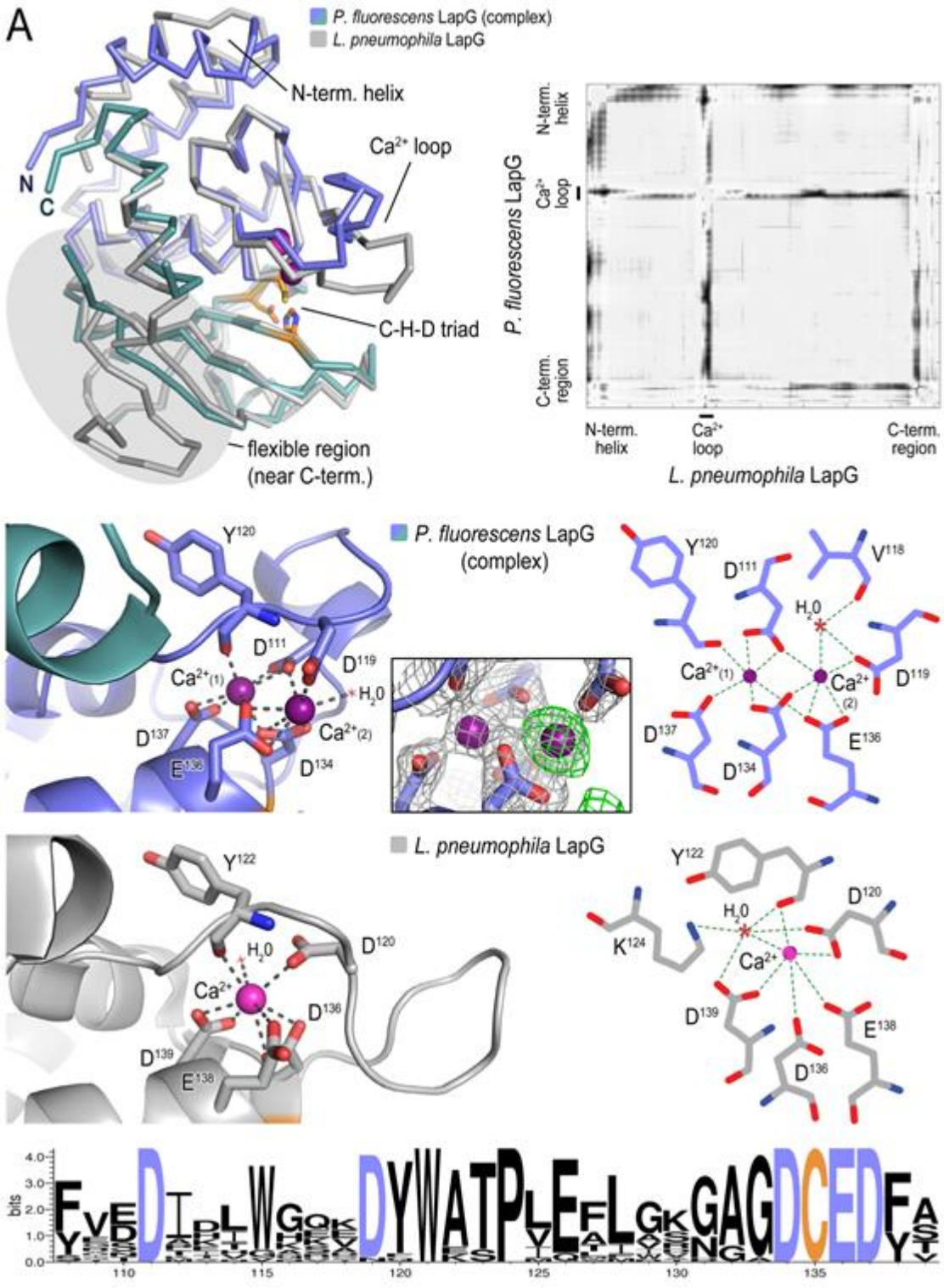
***P. fluorescens* LapG vs. *L. pneumophila* LapG**

In order to ascertain the degree of conservation of the LapG orthologs and to detect any potential conformational changes that may occur upon binding to a LapD ortholog, the crystal structure of $Ca^{2+}$-bound *L. pneumophila* LapG was superimposed onto that of *P. fluorescens* LapG from the complex structure (Fig. 4.4). Despite moderate sequence conservation between the two orthologs, the two structures are very similar with regard to the overall fold ($C_\alpha$ rmsd of 1.09Å over 136 atoms). However, there are a few significant differences between the two. Based on the crystal structure of *P. fluorescens* LapG in the complex, we were able to identify a rare di-$Ca^{2+}$ binding motif in the N-terminal lobe of LapG. This was rather unusual because two positively charged species when placed in such close proximity are expected to repel each other. In *P. fluorescens* LapG however, the two $Ca^{2+}$ ions are enclosed in a tight co-ordination shell, which compensates for any repulsive forces. No additional $Ca^{2+}$ was added during purification or crystallization, indicating that both $Ca^{2+}$ ions co-purified with the complex. In contrast, we identified only a single $Ca^{2+}$ ion at the same site in the *L. pneumophila* LapG structure, which was in agreement with binding studies using this ortholog. This apparent difference may be explained by the reorganization of a region in LapG we termed $Ca^{2+}$-loop. In *P. fluorescens* LapG, an additional negatively charged residue ($D^{111}$) coordinates to the $Ca^{2+}$ ion that is observed in both structures, thus allowing $D^{119}$ to adopt an alternate side chain orientation and thereby accommodating the second $Ca^{2+}$ ion. At this point it remains unclear as to whether this type of di-$Ca^{2+}$ binding is specific to *P. fluorescens* LapG or it is associated with $CdgS9^{output}$-LapG complex formation. Alternatively, the second

$Ca^{2+}$ binding site in *L. pneumophila* LapG could be cryptic, requiring a particular chemical environment to enable its occupation by divalent cations. This notion is supported by the overall sequence conservation of residues that coordinate both $Ca^{2+}$ ions in *P. fluorescens* LapG.

**Figure 4.4: Structural identification of a rare di-$Ca^{2+}$ binding motif in *P. fluorescens* LapG.**

**(A)** On the left, a ribbon representation of superimposed atomic models of LapG from *L. pneumophila* and *P. fluorescens* is shown, with the former colored gray and the latter slate and cyan, respectively. The catalytic triad residues (C-H-D) are shown as sticks with the carbon atoms colored in orange. $Ca^{2+}$ ions are indicated as magenta and pink spheres in the structures of *P. fluorescens* and *L. pneumophila* LapG, respectively. Differential orientations of the $Ca^{2+}$ binding loop are indicated. In the middle panels, close-up views and LIGPLOT analysis of the $Ca^{2+}$ binding loop reveal that $D^{119/120}$ and $E^{136/138}$ adopt alternate side chain rotamers in the two structures, explaining the ability of *P. fluorescens* LapG to bind two calcium ions as opposed to one. Due to the rearrangement of the loop in *P. fluorescens* LapG, a third residue $D^{111}$ is now placed proximally to one $Ca^{2+}$ ion, thereby stabilizing this motif. The central inset shows uninterpreted electron density that was observed only in the structure of *P. fluorescens* LapG. The density has amplitudes of $|F_o|$ to $|F_c|$ and is contoured at 3 $\sigma$ prior to the inclusion of calcium at that site in the refinement. In the right-top panel, distance difference matrices based on $C_\alpha$ positions of the two structures were used to compare the conformations in a reference frame-independent manner. Difference matrices were regularized using a Z-score analysis and shaded in grey indicating the magnitude of conformational difference (white, no difference; grey, conformational difference between the two orthologs). Each entry in the matrix depicts the difference in distance between corresponding $C_\alpha$ atoms in the two structures. In the bottom panel, the sequence logo is shown for the LapG family of proteases, indicating conservation of the residues in the di-$Ca^{2+}$ binding.


A
P. fluorescens LapG (complex)
L. pneumophila LapG
N-term. helix
Ca2+ loop
N
C
C-H-D triad
flexible region
(near C-term.)
P. fluorescens LapG
L. pneumophila LapG
N-term. helix
Ca2+ loop
C-term. region
Y120
D111
D119
Ca2+(1)
H2O
Ca2+(2)
D137
E136
D134
V118
Y122
D120
K124
D139
E138
D136
bits

The second region of conformational change between CdgS9$^{output}$-bound *P. fluorescens* LapG and the *L. pneumophila* LapG ortholog apparent in the superposition localizes to the site of interaction with LapD orthologs (Fig. 4.5). While the overall binding pocket for the critical tryptophan residue in the output domain of *L. pneumophila* CdgS9 (W$^{126}$) is preserved in apo-*L. pneumophila* LapG, output domain binding requires the peeling away of β5, a short beta blade connecting to the C-terminal helix α6 via a flexible linker. In the complex, both β5 and the linker are disordered, whereas β5 and part of the linker are ordered in the apo form.

***Apo vs complexed CdgS9$^{output}$***

Considering that the core LapG-binding motif, the conserved GW$^{126}$xQ loop (GW$^{125}$xQ in *P. fluorescens* LapD$^{output}$), is surface exposed in the apo-CdgS9$^{output}$ structure and potentially available for engaging LapG, it was not clear how differential LapG binding would be achieved. This issue could be addressed through a structural comparison between apo-CdgS9$^{output}$ and its LapG-bound state. This analysis revealed the conformational changes that create a high-affinity binding site on CdgS9 for LapG.

Superposition using a single protomer chain of the structures of apo- and LapG-bound CdgS9$^{output}$ as the reference revealed that the periplasmic output domain undergoes a piston-like conformational change concomitant with a 5 Å-sliding motion along the output domain dimer interface (Fig. 4.6). This concerted change creates a cradle for the aforementioned β4-blade of LapG, which does not exist in the apo-state of the output domain (Fig. 4.6A). The conformational change is realized through the central helices of the output domain of CdgS9 and the residues that line the interface

between them (Fig. 4.6B). In particular, the two helices of each PAS domain fold (H1 and H2) come together in the dimeric output module to form a 4-helix bundle in the apo-state (Fig. 4.6B and C). The overall structure appears bulged, which is a consequence of a perpendicular stacking of four phenylalanine residues ($F^{33}$ and $F^{34}$) at the base of the output module (closer to the membrane) as well as the burying of two arginine residues ($R^{75}$), contributed by each symmetry related protomer (Fig. 4.6C). The charge on arginine residue $R^{75}$ is partially neutralized by serine residue $S^{48}$, which form hydrogen bonds with $R^{75}$.

Due to the conformational change in $CdgS9^{output}$ associated with LapG binding, the symmetry within the output domain module is broken. The tilting of one PAS domain fold relative to the other in the output domain module pushes the central helices H1 closer together, permitted by smaller residues at the interface, a perpendicular-to-parallel stacking transition of one $F^{33}$-$F^{34}$ residue pair, and the removal of $R^{75}$ from the interface. The net result is a transition from a 4-helix bundle with poor shape complementarity with LapG to a 2-helix bundle and an overall shape that is able to accommodate LapG (Fig. 4.6C).

Taken together, the crystal structures of apo-$CdgS9^{output}$ and a $CdgS9^{output}$-LapG complex yield a detailed model for the activation of LapD by c-di-GMP binding to its cytosolic module, which releases an autoinhibitory interaction, which like triggers a conformational change in the periplasmic domain through the juxtamembrane HAMP domain and transmembrane segments. LapG binding to the off-state of LapD is prevented by an apparent surface mismatch. In contrast, activated

LapD displays a surface that allows LapG to interact with both half-sides of the output domain module.

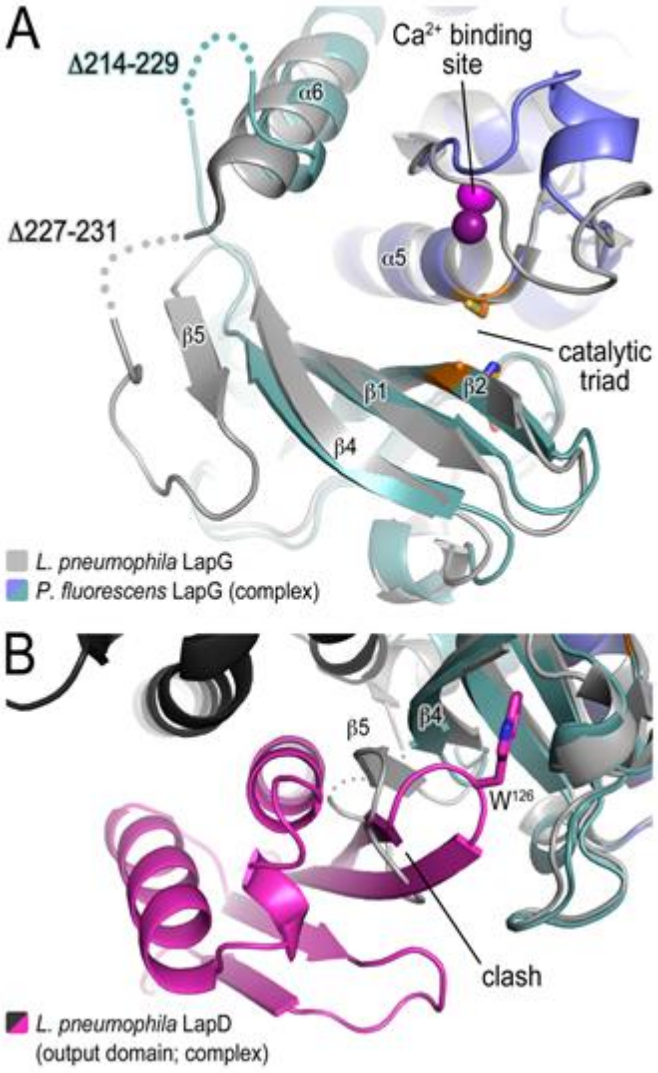



**Figure 4.5: Conformational changes between apo- and CdgS9$^{output}$-bound LapG.** **(A)** A close-up view of the C-terminal lobe of LapG is shown. The structure of *L. pneumophila* LapG shown in gray is overlaid on that of *P.fluorescens* LapG, colored in slate and cyan. Disordered regions not resolved in the crystal structures are shown as dots. Calcium ions are represented as magenta spheres. The catalytic triad residues (C-H-D) are shown as sticks with the carbon atoms colored in orange.
**(B)** A close-up view of the interaction surface between LapG and CdgS9$^{output}$ identifies a *L. peumophila* LapG conformation that is incompatible with CdgS9 binding.

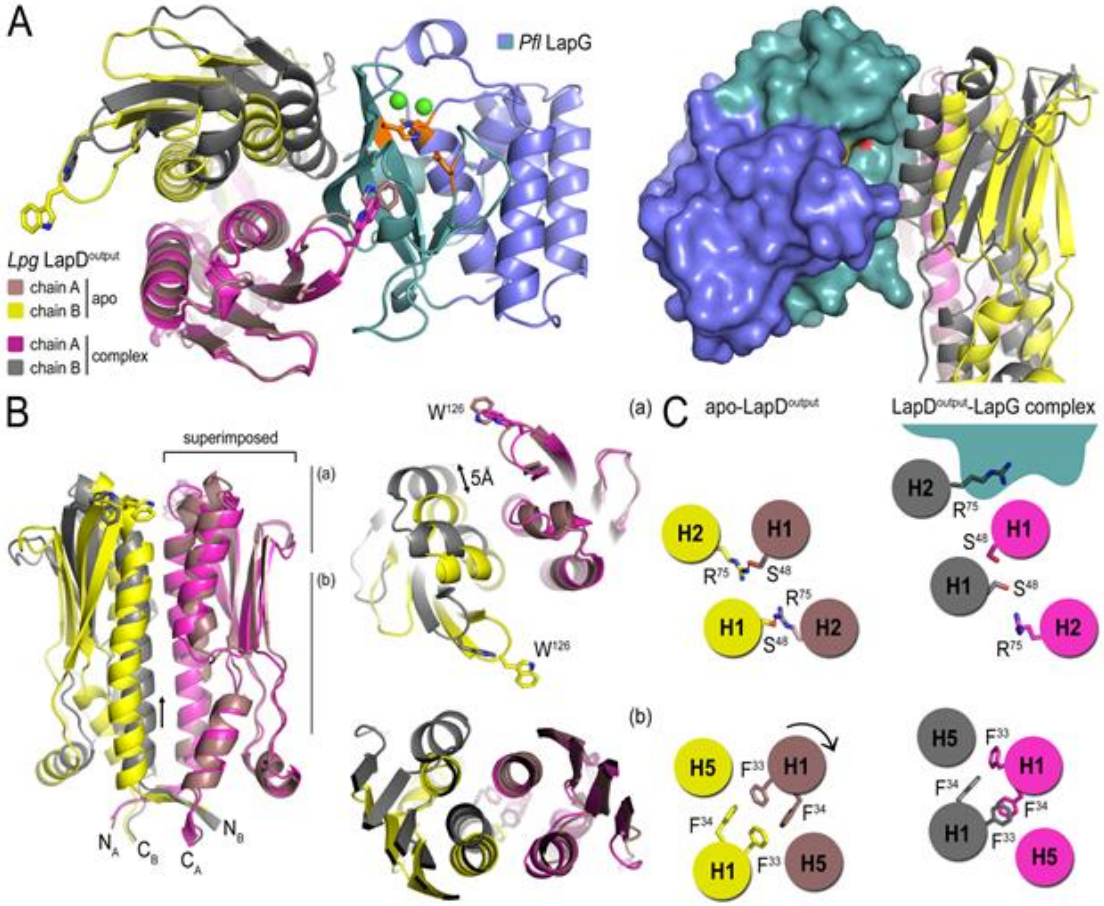



**Figure 4.6: Conformational changes in CdgS9$^{output}$ between the apo-state and in complex with LapG.**

**(A)** Top and side views of the superposition of apo-CdgS9$^{output}$ and LapG-bound CdgS9$^{output}$ with chain A as the reference. Protomers A and B of apo-CdgS9$^{output}$ are colored tan and yellow, while chains A and B of LapG-bound CdgS9$^{output}$ are colored pink and gray, respectively. Both cartoon (left panel) and surface (right panel) representations of LapG are shown.

**(B)** Detailed view of the conformational differences between the two output module states. Two perpendicular views are shown. The right panel shows slices of the top view.

**(C)** Cartoon model of the structural transition from an apo-state to a LapG binding-competent state of LapD orthologs based on the analysis shown in panels (A) and (B).

***Stoichiometry and asymmetry of the LapG-CdgS9$^{output}$ complex***

We were intrigued by the 2:1 stoichiometry of the LapG-CdgS9$^{output}$ complex in the crystal structure. Since full-length LapD is considered constitutively dimeric (17-19) owing to its HAMP and output domains (19), and the apo-CdgS9$^{output}$ structure is symmetric, we wondered if the formation of a symmetric, 2:2 complex was theoretically possible. Inspection of the complex structure revealed an asymmetric structure in which the conformational changes were not equivalent in the two half-sites of the output domain dimer (Fig. 4.6C). Furthermore, superimposing a CdgS9$^{output}$ domain that engages LapG via its GW$^{126}$xQ loop onto the output domain of the complex structure, which is not bound to LapG, reveals poor shape complementarity at the putative second LapG binding site (Fig. 4.7A). While this could be due to a substoichiometric co-expression of the complex components or an incomplete transition of the output domain to a fully active, symmetric state, it may also indicate mechanisms that would favor an asymmetric 2:1 complex.

While we cannot distinguish between these possibilities based on our current data, we can assess the stoichiometry of the complex in solution. We used size-exclusion chromatography-coupled multi-angle light scattering (SEC-MALS), a method that yields the absolute molecular weight of a protein or complex as well as its polydispersity index, as it is eluted from a gel filtration column (Fig. 4.7B). First, we analyzed the complexes obtained from co-expression of LapG and LapD$^{output}$ orthologs. The [*P. fluorescens* LapG - *L. pneumophila* CdgS9$^{output}$] complex sample used for crystallization eluates predominantly as a 2:1 complex although a slight upwards trend at the left tail-end of the peak may hint at a fraction of higher-order or

2:2 complexes. Next we used the complex between homologous proteins ($CdgS9^{output}$ and LapG) from *L. pneumophila*. The *L. pneumophila* complex obtained through co-expression also purifies predominant 2:1 species. While this observation ruled out any discrepancies that may arise from studying a heterologous complex, it did not answer whether a 2:2 complex can be achieved. To ascertain this possibility, we spiked the co-expressed and purified complex with an excess purified *L. pneumophila* LapG. If anything, the size distribution of the complexes got even tighter, with only a 2:1 species being observed.

Together, these results indicate the existence of mechanisms that prevent the formation of symmetric 2:2 complexes between LapG and LapD, which are apparent from the structural characterization. Preliminary data using the full-length receptor even suggest anti-cooperativity as a mode for fine-tuning complex formation (data not shown).

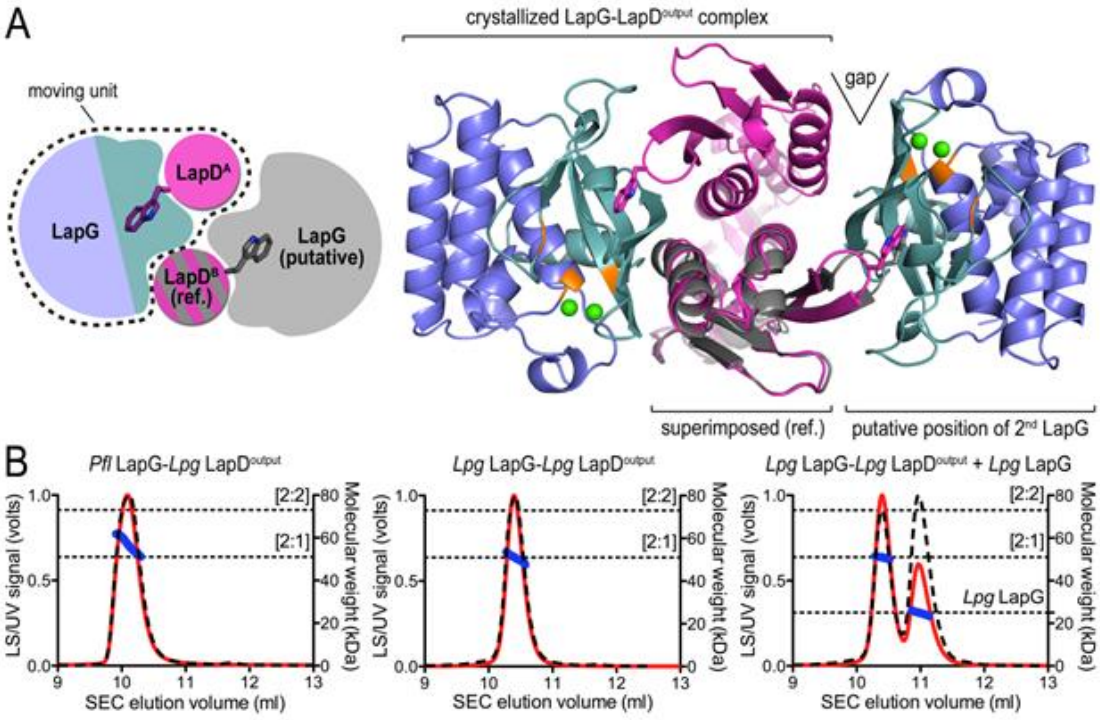



**Figure 4.7: Stoichiometry of CdgS9$^{output}$-LapG complex.**
**(A)** Superimposition of a CdgS9$^{output}$ domain that engages LapG via its GW$^{126}$xQ loop onto the output domain of the complex structure, which is not bound to LapG, reveals poor shape complementarity at the putative second LapG binding site.
**(B)** Stoichiometry of CdgS9$^{output}$-LapG complex in solution. CdgS9$^{output}$-*P. fluorescens* LapG elutes as a 2:1 complex with a slight tendency to form a higher ordered species. CdgS9$^{output}$-*L. pneumophila* LapG is predominantly a 2:1 stoichiometric species and is not responsive to excess *L. pneumophila* lapG.

## 4.4. Discussion

The structural analyses described above of the output domain of the *L. pneumophila* CdgS9 ortholog in its apo-state and when bound to a LapG ortholog revealed the molecular mechanism by which LapD regulates proteolysis in the periplasm by sequestering the protease LapG. As an extension, we predict that the structural transitions observed in the isolated output domain are triggered by an allosteric mechanism involving the cytoplasmic domains of LapD in the context of the full-length, transmembrane protein. As such, we argue that the isolated domain is in a deregulated state and is able to react to LapG binding since regulatory, autoinhibitory domains are lacking, that would lock the receptor in its off-state (the apo state). This is consistent with the observation that the isolated output domain is a poor competitor for the interaction between LapG and full-length LapD (data not shown). Nevertheless, the conservation of the involved interfaces and motifs in the otherwise fairly divergent LapD family of receptors suggest that the proposed mechanism is more generally applicable. Validation of the working model by using homology modeling and structure-guided site-directed mutagenesis in *P. fluorescens* is currently under way. Also, while we currently favor the possibility that the domain-swapped *P. fluorescens* $LapD^{output}$ structure (Fig. 4.2A) is an artifact, we are in the process of formally assessing the quaternary structure arrangement.

LapD was initially discovered in *P. fluorescens* as a crucial regulator of cell adhesion, and the underlying regulatory network remains one of best-understood signaling systems with regard to LapD receptor function but also considering other c-di-GMP-dependent pathways (17-21). LapD-like signaling nodes were validated in a

few other biofilm-forming bacteria. Unfortunately, while we have strong evidence that they are regulated in a similar fashion, we know very little about the biological function of other orthologs, especially when it comes to pathogens. Here, we used the *L. pneumophila* LapD ortholog, CdgS9, which facilitated the structural studies. Although genetic knock-outs have no obvious phenotype in *L. pneumophila*, overexpression of CdgS9 has an effect on intracellular growth of the pathogen (48). Future studies will have to address the physiological pathways that LapD and LapG orthologs regulate. Until then, these proteins remain orphan c-di-GMP receptors and proteases, respectively.

Previous studies on the switching of HAMP domains and periplasmic domains in bacterial proteins have yielded several models for the conformational changes associated with changes in activity state of these proteins. The proposed structural rearrangements range from piston motions to scissor-type motions (29, 30). Here we describe the conformational changes associated with the transition of a ligand-free LapD receptor output domain to a LapG-bound state. The overall motion has components of a piston-type model, where one half-sides shift vertically along the second half-side of the dimeric output domain module, and a scissor-like model encompassing a sliding motion along the dimer interface. Considering the current model for HAMP domain-switching which involves undergoing small amplitude translations or rotations during signal transduction (29, 30), we predict that the output domain follows the conformational changes in the HAMP as a direct response. At the same time, the apparent asymmetry of the complex suggests mechanisms inherent to the output domain that counteract an activating signal and ultimate contribute to a anti-

cooperativity. The current working model predicts that (transient) occupation of the second LapG binding site on the dimeric output domain of LapD would release LapG at the first, high-affinity site.

In summary, we propose a detailed model for how c-di-GMP binding to the cytoplasmic domain of LapD sets off a conformational change that is communicated through the HAMP domain to the periplasmic domain. This allosteric trigger allows LapD to bind to LapG in the periplasm and as a result prevents LapA cleavage. Based on the conservation of binding events and key sequence motifs, we predict that the model may be universal for LapD-like protein, which may also include receptors that signaling in the opposite direction (‘outside-in’) by translating environmental signals emerging in the periplasms into altered activity states in the cytoplasm. Future studies will ascertain the generality of our current model and the potential variations thereof.

## 4.5. Material and Methods

### *Protein expression and purification*

The periplasmic output domain of the *L. pneumophila* LapD ortholog CdgS9 (*lpg0829*; residues 22-152) and *P. fluorescens* LapG lacking the signal peptide (*Pfl01_0130*; residues 50-251) were amplified from genomic DNA by standard PCR and cloned separately into a pET28a based vector (Novagen) that adds an N-terminal, cleavable $His_6$-SUMO tag, as well as together into a bacterial dual expression vector, pETDuet-1 (Novagen). With regards to the duet vector, *L. pneumophila* $His_6$-SUMO-$CdgS9^{output}$ and *P. fluorescens* LapG were cloned into MCS-1 and MCS-2, respectively.

Native and selenomethionine-substituted proteins were over-expressed in *E. coli* BL21 T7 Express or T7 Express Crystal cells (New England Biolabs), respectively. For the expression of native proteins, cultures were grown at 37°C in Terrific Broth (TB) media supplemented with 50 μg/ml kanamycin. At an $OD_{600}$ of ~1, the temperature was reduced to 18°C, and protein expression was induced by adding 1 mM IPTG. Selenomethionine-labeled proteins were expressed in cells grown at 37°C in M9 minimal media supplemented with the desired antibiotic, vitamins (1 μg/ml thiamin and 1 μg/ml biotin), carbon source (0.4% glucose), trace elements, and amino acids (50 μg/ml of each of the 20 amino acids with selenomethionine substituting for methionine). Protein expression was induced at an $OD_{600}$ corresponding to ~0.6. In both cases, protein expression was allowed to proceed for 16 hours at 18°C, after which cells were harvested by centrifugation, re-suspended in Ni-NTA buffer A (25 mM Tris-HCl [pH 8.5], 500 mM NaCl and 20 mM imidazole), and flash frozen in

liquid nitrogen. Cell suspensions were thawed and lysed by sonication. Cell debris was removed by centrifugation and the clear lysates were incubated with NiNTA resin (Qiagen) that was pre-equilibrated with Ni-NTA buffer A. The resin was washed with 20 column volumes of buffer A, followed by protein elution with 5 column volumes of NiNTA buffer B (25 mM Tris-HCl [pH 8.5], 500 mM NaCl and 300 mM imidazole). The eluted proteins were buffer-exchanged into a low-salt buffer (25 mM Tris-HCl [pH 8.5] and 150 mM NaCl) on a fast desalting column (GE Healthcare). Proteins were subjected to size exclusion chromatography on a Superdex 200 column (GE Healthcare) pre-equilibrated with gel filtration buffer (25 mM Tris-HCl [pH 8.5] and 150 mM NaCl). Where indicated, the $His_6$-SUMO moiety was cleaved off by using the yeast protease Ulp-1 following desalting. Ulp-1, uncleaved protein and the cleaved fusion tags were removed by NiNTA affinity chromatography prior to the final gel filtration.

For co-purification of the $CdgS9^{output}$-LapG complex, the procedure remained identical barring the use of low salt buffers, i.e. instead of 500mM NaCl, Buffers A and B had 200 mM salt, while the final gel-filtration buffer contained 200mM NaCl. Purified proteins were concentrated on Amicon filters with an appropriate size cutoff to concentrations >30 mg/ml, flash frozen in liquid nitrogen and stored at -80°C. Site-directed mutagenesis was carried out using the Quikchange kit (Agilent Technologies) following the manufacturer's instructions, followed by validation through DNA sequencing.

***Crystallization, data collection and structure solution***

Crystals were obtained by hanging drop vapor diffusion mixing equal volumes of protein (10-30 mg/ml) and reservoir solution followed by incubation at 20°C. For the native and selenemethionine-derivatized *L. pneumophila* CdgS9$^{output}$ crystals, the reservoir solution contained 0.1M Bis-Tris (pH=5.0), 14% PEG3350 and 4% v/v 2,2,2-Trifluoroethanol. Crystals of the native CdgS9$^{output}$-LapG complex were obtained from a reservoir solution containing 0.1M Bis-Tris (pH=6.0) and 0.1M Magnesium Formate. Our initial efforts directed towards experimental phase determination proved intractable due to extremely low expression levels of the selenomethione-substituted CdgS9$^{output}$-LapG complex. Hence phases were determined by harvesting crystals generated by *in vitro* mixing of *P. fluorescens* LapG expressed as the selenomethionine-derivatized protein, with native *L. pneumophila* CdgS9$^{output}$ from the same condition as that for the native ones. For cryo-protection, crystals were soaked in reservoir solution supplemented with 20-25% 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 Cornell High Energy Synchrotron Source (CHESS, Ithaca).

Data reduction was carried out with the software package HKL2000 (23). Experimental phases for the initial structure determination were obtained from Single Anomalous Diffraction (SAD) experiments on crystals grown from selenomethionine-derivatized proteins by using the software package PHENIX (24). Refinement in PHENIX and COOT (25) yielded the final models. Data collection and refinement statistics are summarized in Table 4.1. Illustrations were made in Pymol

(Schrödinger). Alignments were generated using ClustalW2 (26). Sequence logos were generated using WebLogo (27, 28).

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

SEC-MALS measurements were carried out by injecting purified proteins (100 µM) onto a Phenomenex gel filtration column pre-equilibrated with gel filtration buffer (25 mM Tris-HCl [pH 7.4] and 100 mM NaCl). The SEC system was coupled to an 18-angle, static light scattering detector and a refractive index detector (DAWN HELEOS-II and Optilab T-rEX, respectively; Wyatt Technology). Data was collected at 25°C every second at a flow rate of 1 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.

## References


1. Hall-Stoodley L, Costerton J. W, Stoodley P. Bacterial biofilms: from the natural environment to infectious diseases. 2004. Nat Rev Microbiol. 2:95–108.

2. Parsek M. R, Singh P. K. Bacterial biofilms: an emerging link to disease pathogenesis. 2003. Annu Rev Microbiol. 57:677–701.

3. Mah TF, O'Toole GA. 2001. Mechanisms of biofilm resistance to antimicrobial agents. Trends Microbiol. 9(1):34-9.

4. Spellberg B, Powers JH, Brass EP, Miller LG, Edwards JE. 2004. Trends in antimicrobial drug development: Implications for the future. Antimicrobial Research and Development • CID. 38:1279–86.

5. Hengge R. Principles of c-di-GMP signalling in bacteria. 2009. Nat Rev Microbiol. 7:263–273.

6. Schirmer T, Jenal U. Structural and mechanistic determinants of c-di-GMP signalling. 2009. Nat Rev Microbiol. 7:724–735.

7. Ryan R. P, Fouhy Y, Lucey J. F, Crossman L. C, Spiro S, et al. Cell-cell signaling in*Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. 2006. Proc Natl Acad Sci U S A. 103:6712–6717.

8. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T. 2004.Structural basis of activity and allosteric control of diguanylate cyclase. Proc. Natl. Acad. Sci. U. S. A. 101:17084 –17089.

9. Tchigvintsev A, Xu X, Singer A, Chang C, Brown G, Proudfoot M, Cui H, Flick R, Anderson WF, Joachimiak A, Galperin MY, Savchenko A, Yakunin AF. 2010. Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases. J. Mol. Biol. 402:524–538.

10. Lovering AL, Capeness MJ, Lambert C, Hobley L, Sockett RE. 2011. The structure of an unconventional HD-GYP protein from Bdellovibrio reveals the roles of conserved residues in this class of cyclic-di-GMP phosphodiesterases. MBio. 2(5).

11. Ryjenkov DA, Simm R, Romling U, Gomelsky M. The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. 2006. J Biol Chem. 281:30310–30314.

12. Krasteva P. V, Fong J. C, Shikuma N. J, Beyhan S, Navarro M. V, et al. *Vibrio cholerae*VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. 2010. Science. 327:866–868.

13. Hickman J. W, Harwood C. S. Identification of FleQ from *Pseudomonas aeruginosa* as a c-di-GMP-responsive transcription factor. 2008. Mol Microbiol. 69:376–389.

14. Sudarsan N, Lee E. R, Weinberg Z, Moy R. H, Kim J. N, et al. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. 2008. Science. 321:411–413.

15. Leduc J. L, Roberts G. P. Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of *Xanthomonas axonopodis pv. citri*. 2009. J Bacteriol. 191:7121–7122.

16. Navarro M. V, De N, Bae N, Wang Q, Sondermann H. Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX. 2009. Structure. 17:1104–1116.

17. Newell P. D, Monds R. D, O'Toole G. A. LapD is a bis-(3′,5′)-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0-1. 2009. Proc Natl Acad Sci U S A.106:3461–3466.

18. Newell PD, Boyd CD, Sondermann H, O'Toole GA. 2011. A c-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage. PLoS Biol. 9:e1000587.

19. Navarro MV, Newell PD, Krasteva PV, Chatterjee D, Madden DR, O'Toole GA, Sondermann H. 2011. Structural basis for c-di-GMP mediated inside-out signaling controlling periplasmic proteolysis. PLoS Biol. 9(2):e1000588.

20. Chatterjee D, Boyd CD, O'Toole GA, Sondermann H. 2012. Structural characterization of a conserved, calcium-dependent periplasmic protease from Legionella pneumophila. J. Bacteriol. 194(16):4415-25.

21. Boyd CD, Chatterjee D, Sondermann H, O'Toole GA. 2012. LapG, required for modulating biofilm formation by *Pseudomonas fluorescens* Pf0-1, is a calcium-dependent protease. J. Bacteriol. 194:4406–4414.

22. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. 2007. J Mol Biol. 372:774–797.

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

24. Adams P, Grosse-Kunstleve R, Hung L, Ioerger T, McCoy A, et al. (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 58: 1948-1954.

25. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography 60: 2126-2132.

26. 24. Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

27. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188 –1190.

28. Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18:6097– 6100.

29. Airola MV, Sukomon N, Samanta D, Borbat PP, Freed JH, Watts KJ, Crane BR. HAMP domain conformers that propagate opposite signals in bacterial chemoreceptors. 2013. PLoS Biol. 11(2):e1001479.

30. 30. Hulko M, Berndt F, Gruber M, Linder JU, Truffault V, Schultz A, Martin J, Schultz JE, Lupas AN, Coles M. The HAMP domain structure implies helix rotation in transmembrane signaling. 2006. Cell. 126(5):929-40.

# CHAPTER 5

## Conclusions

Bacteria have the ability to form surface-attached communities, called biofilms, in both free-living environmental habitats and during pathogenic colonization in infectious diseases (1). Many of the cellular processes contributing to biofilm formation are regulated by the molecule bis-(3'-5')-cyclic dimeric guanosine monophosphate or or c-di-GMP. This RNA-based dinucleotide has emerged as a broadly conserved intracellular second messenger in bacteria, regulating adhesion, motility, biofilm formation, differentiation and cell cycle progression, while exerting control at transcriptional, translational and post-translational levels (2-7). Cyclic-di-GMP is synthesized from two molecules of GTP by diguanylate cyclases (DGCs) containing the GGDEF domain (8). Degradation of c-di-GMP is accomplished by phosphodiesterases (PDEs) containing either an EAL (9, 10) or a HD-GYP domain (11). Many GGDEF and EAL domains occur in multi-domain proteins, fused with diverse regulatory domains common to bacterial signaling proteins (12). These regulatory domains can modulate synthesis or degradation of c-di-GMP in response to a broad range of signals such as light (13), oxygen (14), or through post-translational modifications such as phosphorylation (15) or proteolysis (13). Up until now, only a few of these environmental and cellular cues that modulate c-di-GMP mediated signaling pathways have been identified. In order to fully appreciate the multidimensional influence of these signaling networks on bacterial physiology and

pathogenesis, it is important that we strive to unveil novel sensory inputs and their mechanism of downstream signal propagation.

In contrast to c-di-GMP turnover domains (GGDEF, EAL and HD-GYP) that are readily identifiable based on preliminary sequence analysis, the staggering diversity in modes of c-di-GMP binding has made the recognition of its effector components quite challenging. Yet, identification of these receptors is instrumental to fully gauge the entire gamut of cellular processes regulated by this this cyclic nucleotide. Receptors of c-di-GMP identified thus far include proteins with PilZ domains (17, 20, 21) or degenerate GGDEF/EAL domains, transcription factors and riboswitches (5, 7, 22, 43), and have been shown to impact exopolysaccharide (EPS) synthesis (16-18), motility (19-21), transcription (22-24), and sub-cellular (25) or cell-surface protein localization (26). However, out of these only a few studies have revealed the molecular details of how these c-di-GMP effectors control their final outputs (22, 25-31).

The existence of an astonishingly large number of predicted c-di-GMP signaling proteins in certain bacterial genomes (there are 43 and 22 in our study organisms, *Pseudomonas fluorescens* Pf0-1 and *Legionella pneumophila*, respectively) raises the possibility that c-di-GMP signaling networks are extraordinarily complex. Recent studies have provided strong evidence that c-di-GMP signaling proteins (DGCs or PDEs) can produce distinct outputs through differential subcellular localization of their targets (32-34). Other studies have shown that c-di-GMP signaling proteins in the same pathway often physically interact with one another or with their targets (35). These data, combined with the well-established role of oligomerization in regulating

DGC and PDE activities, points to direct protein-protein interactions and multi-protein complexes as a likely means to exert spatial and temporal compartmentalization of specific c-di-GMP pools. It is possible that some of these c-di-GMP signaling proteins may be expressed only under specific growth conditions or in response to certain environmental variations. For example, *rapA*, which encodes a c-di-GMP specific PDE in *P. fluorescens*, is expressed only under conditions of exogenous inorganic phosphate ($P_i$) starvation (36). However, it has still not been completely understood as to how such spatiotemporal sequestration can be achieved.

Furthermore, the apparent imbalance between the number of c-di-GMP metabolizing enzymes and its sensors suggests that new classes of c-di-GMP-binding proteins are only waiting to be uncovered. Exciting new high-throughput methods of identification of potential c-di-GMP receptors have recently been proposed. One such method involves use of a c-di-GMP-affinity resin to capture target proteins from cell lysates (40), while the other relies on utilizing an *E. coli* based over-expression system, thereby circumventing the need of protein purification (41). The success of the proposed techniques would contribute to a better characterization and understanding of the complex c-di-GMP signaling network.

Additionally, understanding how c-di-GMP receptors participate in generating biological outputs has also remained a poorly understood process, yet essential to appreciating a critical component of the bacterial lifestyle, that is, the transition between motile and sessile states. This is particularly important because it is estimated that approximately 80% of human infections involve bacterial biofilms. Since biofilm dwelling bacteria is much more tolerant to antibiotic treatment as compared to

planktonic cells, it has become extremely difficult to remove these microbial aggregates in clinical as well as industrial settings. Thus, in order to come up with effective measures to prevent and/or disrupt biofilms, we need to attain a fundamental understanding of the molecular mechanisms and regulatory pathways that govern this developmental process in bacteria.

It has been established that c-di-GMP is necessary for biofilm formation and maintenance in bacteria. Hence, decreasing intracellular c-di-GMP levels should lead to biofilm dispersal. This indeed is true in case of *P. fluorescens* biofilms, where detachment is initiated upon upregulation in the expression of a c-di-GMP specific PDE RapA. In *P fluorescens*, stable surface attachment and commitment to a sedentary lifestyle in response to nutrient availability in the environment, is dependent on the appropriate localization of an outer-membrane embedded, large adhesin protein LapA (36). LapA plays a critical role in an early step in biofilm development called "irreversible attachment", during which bacterial cells attach to a surface via their longitudinal axis (36), and its secretion requires LapEBC, an ATP-binding cassette type transporter (37). So how does the adhesin LapA impact biofilm development?

It has been shown that LapD, an inner-membrane localized c-di-GMP receptor that binds the dinucleotide via its catalytically degenerate EAL domain, plays an important role in the localization and maintenance of LapA, by differential recruitment of a periplasmic cysteine protease LapG, in response to changing pools of c-di-GMP in the cytoplasm, which in turn is regulated by a system (Pst) that monitors inorganic phosphate levels in the environment (26, 29, 30, 36, 37). LapD is a multi-domain protein with a cytoplasmic HAMP-GGDEF-EAL module and a periplasmic output

module (26). In order to gain a molecular insight into the mechanism of LapA regulation by LapD, we conducted X-ray crystallographic analyses of the structure of the intracellular and periplasmic modules of LapD.

Our structural analyses of LapD revealed an autoinhibited conformation of the cytosolic GGDEF-EAL dual domain module ($LapD^{dual}$) in the apo-state, a dimeric form of c-di-GMP-bound EAL domains in the active state ($LapD^{EAL}$·c-di-GMP), and an extensively interwoven, domain-swapped dimer architecture of the periplasmic output domain ($LapD^{output}$) that is competent for LapG binding. The above data, coupled with extensive biochemical and genetic analyses, helped us formulate a hypothesis for the mechanism of c-di-GMP mediated LapA regulation by LapD (29, 30, Chapter II).

When exogenous inorganic phosphate availability is severely limited, intracellular c-di-GMP levels are low and LapD is held in an autoinhibited conformation or an "off" state, in which access of c-di-GMP to the EAL domain is partially blocked. In this conformation, LapD's output domain is unable to bind LapG in the periplasm. Thus, LapG gains access to and cleaves the amino-terminus of LapA, releasing this critical biofilm adhesin from the cell surface and thereby leading to biofilm dispersal. In contrast, when there is abundant inorganic phosphate in the growth medium, c-di-GMP levels in the cytosol increases and it binds to the EAL domain of LapD, disrupting the autoinhibitory state and inducing a conformational change that is detected by the HAMP domain and relayed on to the periplasmic output domain causing sequestration of the periplasmic protease LapG, thereby preventing

the proteolysis of LapA, and ultimately leading to LapA retention on the outer membrane and hence stable surface attachment.

By revealing key motifs for the regulation of LapD, we identified similar signaling systems in many other bacterial strains that might control periplasmic protein processing events in a similar regulatory fashion (29, 30). While functionally important motifs and the overall domain organization were conserved, the overall sequence identity was rather low. Hence, we decided to establish the conservation of the basic receptor function and autoinhibition based on the isolated, soluble domains of the Lap orthologs from *Legionella pneumophila*, the causative agent of Legionnaires' disease in humans (46).

Our work on the LapG and LapD orthologs from *L. pneumophila* indeed established that they form a complex as part of a conserved regulatory system, with a significant degree of cross-specificity between the two systems. In order to shed light on the molecular details of LapG function, we also determined the crystal structure of *L. pneumophila* LapG in its apo- as well as $Ca^{2+}$-bound states. This in turn provided the first atomic models of bacterial proteases of the DUF920 family and revealed a conserved calcium-binding site important for LapG function (38, 39, Chapter III). Although the identity of the endogenous substrate for the *L. pneumophila* ortholog of LapG remains unknown, we have shown that the enzyme can proteolyze the corresponding *P. fluorescens* substrate, indicating a common catalytic mechanism and substrate recognition (38). In future, it would be interesting to validate, structurally and functionally, the hypothesis that LapG recognizes discrete features of the N-terminal domain of LapA that allows this protease to specifically target the adhesin.

Additionally, efforts are underway to identify the corresponding LapG substrates in the pathogenic strains. A detailed understanding of the proteolytic cleavage mechanism will be instrumental for exploiting this signaling system as a therapeutic target in the battle against infectious diseases.

Our structural analyses of the output domain of CdgS9 ($CdgS9^{output}$) from *L. pneumophila* and that of a complex between *P. fluorescens* LapG and *L. pneumophila* $CdgS9^{output}$, has furthered our mechanistic understanding of c-di-GMP regulated LapD-LapG mediated signal propagation in the periplasm. Breaking this interaction pharmacologically might provide a route to prevent or dispel bacterial biofilms and possibly host-pathogen interactions, and could potentially yield a solid basis for structure-guided drug design (Chapter IV).

In summary, our work has led to the identification and characterization of a novel yet well-conserved "inside-out" signal relay mechanism in bacteria wherein cytosolic levels of a nucleotide based second messenger, namely c-di-GMP, can regulate the localization of a cell-surface associated protein LapA (29, 30) via an inner-membrane anchored c-di-GMP receptor LapD. This is the first structure-function analysis of a complete c-di-GMP signaling module in bacteria, from the primary signal (environmental $P_i$) to the final cellular output (biofilm dispersal). Our X-ray crystallographic studies on the different components of this pathway has provided molecular snapshots of the ongoing signaling events and has steered our investigation towards attempts to elucidate the function of these proteins as signal broadcasters in the big picture of prokaryotic signaling networks.

Additionally, the identification of analogous signaling cascades in several other clinically relevant pathogenic species including *L. pneumophila*, *V. cholerae*, *P. aeruginosa* etc. implies that the results discussed here might have broader implications for receptors predicted to mediate either inside-out or outside-in signaling involving the bacterial second messenger c-di-GMP. While LapA regulation happens to be the ultimate output of this pathway in *P. fluorescens*, it is possible that orthologous systems in other pathogens identified through bioinformatics analysis, might involve regulation of secretion of hemolysin-like toxins, which are predicted to have the same N-terminal domain architecture as that of LapA. If so, this would indeed be a novel mechanism of control of microbial pathogenesis by c-di-GMP.

While we have made tremendous progress in understanding some of the key regulatory and architectural features of the LapD-LapG signaling module, several questions still remain unanswered. In particular, while we have successfully characterized the autoinhibited apo-state and the active c-di-GMP-bound state of cytoplasmic LapD, the exact role of the HAMP domain in inside-out signal propagation in the context of the full-length receptor remains poorly understood. The HAMP domain found in LapD occurs in a large number of bacterial proteins, especially histidine kinases and DGC/PDEs (20). Their structure has been studied in isolation, yet little is known about the mechanism of signal relay by this dimeric protein domain. In case of *P. fluorescens* LapD, we have identified both the input signal transmitted by the HAMP domain (corformational rearrangement induced by c-di-GMP binding), and the cellular output regulated by HAMP-mediated changes in protein structure (LapA processing by LapG). Furthermore, with the identification of

more than 300 proteins that contain this HAMP-GGDEF-EAL module, many of which even have intact active site sequences, it is likely that future mechanistic studies on LapD signaling will not only shed light on the unique inside-out signaling mechanism described for LapD homologs, but will also be more generally applicable to proteins involved in outside-in signaling where they work as environmental sensors. In this regard, future attempts towards crystallization of full-length LapD could potentially unravel the signaling mechanisms in these HAMP-containing proteins.

It is also not clear as to how intracellular signaling specificity is achieved. From published and unpublished work it is clear that only a subset of DGCs and PDEs feed into LapD, while the others, who in spite of catalyzing the same basic enzymatic reaction do not alter LapD's activity. Some of this may be explained by differential expression, but not all of the apparent specificity can be explained by it. There is an emerging body of literature suggesting that direct protein-protein interactions between the c-di-GMP turnover enzymes and the c-di-GMP effectors (4, 42, 47), creates local gradients or microcompartments of the dinucleotide within the cytosol, thereby eliciting a specific response. Thus, understanding how multiple c-di-GMP signaling proteins interact with each other or remain isolated in order to generate a biological response is critical to unraveling the intricacies of c-di-GMP signaling in the bacterial kingdom.

## References


1. Hall-Stoodley L, Costerton JW, Stoodley P. 2004. Bacterial biofilms: from the natural environment to infectious diseases. Nat. Rev. Microbiol.2:95–108.

2. D'Argenio DA, Miller SI. 2004. Cyclic di-GMP as a bacterial second messenger. Microbiology 150:2497–2502.

3. Jenal U. 2004. Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria? Curr. Opin. Microbiol. 7:185–191.

4. Jenal U, Malone J. 2006.Mechanisms of cyclic-di-GMP signaling in bacteria. Annu. Rev. Genet.40:385– 407.

5. Römling U, Gomelsky M, Galperin MY. 2005. c-di-GMP: the dawning of a novel bacterial signalling system. Mol. Microbiol. 57:629–639.

6. Römling U, Amikam D. 2006.Cyclic di-GMP as a second messenger. Curr. Opin. Microbiol.9:218 –228.

7. Hengge, R. 2009. Principles of c-di-GMP signalling in bacteria. Nat. Rev. Microbiol. 7**:**263-73.

8. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, Schirmer T. 2004. Structural basis of activity and allosteric control of diguanylate cyclase. Proc. Natl. Acad. Sci. U. S. A. 101:17084 –17089.

9. Christen M, Christen B, Folcher M, Schauerte A, Jenal U. 2005.Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. J. Biol. Chem. 280: 30829–30837.

10. Tiscler AD, Camilli A. 2004. Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation. Mol. Microbiol. 53:857-869.

11. Ryan RP, Fouhy Y, Lucey JF, Crossman LC, Spiro S, Y. He W, L. Zhang H, Heeb S, Camara M, Williams P, and Dow JM. 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.103:6712-7.

12. Galperin MY, Nikolskaya AN, and Koonin EV. 2001. Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol Lett. 203:11-21.

13. Tarutina M, Ryjenkov DA, Gomelsky M. 2006. An unorthodox bacteriophytochrome from *Rhodobacter sphaeroides* involved in turnover of the second messenger c-di-GMP. J. Biol. Chem. 281:34751–34758.

14. Sawai H, Yoshioka S, Uchida T, Hyodo M, Hayakawa Y, Ishimori K, and Aono S. 2010. Molecular oxygen regulates the enzymatic activity of a heme-containing diguanylate cyclase (HemDGC) for the synthesis of cyclic di-GMP. Biochim Biophys Acta 1804:166-72.

15. Paul R, Abel S, Wassmann P, Beck A, Heerklotz H, and Jenal U. 2007. Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. J Biol Chem 282:29170-7.

16. Lee VT, Matewish JM, Kessler JL, Hyodo M, Hayakawa Y, and Lory S. 2007. A cyclic-di-GMP receptor required for bacterial exopolysaccharide production. Mol Microbiol 65:1474-84.

17. Merighi M, Lee VT, Hyodo M, Hayakawa Y, and Lory S. 2007. The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in *Pseudomonas aeruginosa*. Mol Microbiol 65:876-95.

18. Weinhouse H, Sapir S, Amikam D, Shilo Y, Volman G, Ohana P, Benziman M. 1997. c-di-GMP-binding protein, a new factor regulating cellulose synthesis in *Acetobacter xylinum*. FEBS Lett. 416(2):207-11.

19. Christen M, Christen B, Allan MG, Folcher M, Jeno P, Grzesiek S, Jenal U. 2007. DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*. Proc. Natl. Acad. Sci. U.S.A. 104:4112– 4117.

20. Pratt JT, Tamayo R, Tischler AD, Camilli A. 2007. PilZ domain proteins bind cyclic diguanylate and regulate diverse processes in *Vibrio cholerae*. J. Biol. Chem. 282:12860 –12870.

21. Ryjenkov DA, Simm R, Römling U, Gomelsky M. 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. 281:30310 –30314.

22. Krasteva PV, Giglio KM, Sondermann H. 2012. Sensing the messenger: the diverse ways that bacteria signal through c-di-GMP. Protein Sci. 21: 929–948.

23. Leduc JL, Roberts GP. 2009. Cyclic di-GMP allosterically inhibits the CRP-like protein (Clp) of *Xanthomonas axonopodis* pv. *citri*. J. Bacteriol.191:7121–7122.

24. Hickman JW, Tifrea DF, and Harwood CS. 2005. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. Proc Natl Acad Sci U S A 102:14422-7.

25. Duerig A, Abel S, Folcher M, Nicollier M, Schwede T, Amiot N, Giese B, and Jenal U. 2009. Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. Genes Dev 23:93-104.

26. Newell PD, Monds RD, O'Toole GA. 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. 106:3461–3466.

27. Boehm A, Kaiser M, Li H, Spangler C, Kasper CA, Ackermann M, Kaever V, Sourjik V, Roth V, and Jenal U. 2010.Second messenger-mediated adjustment of bacterial swimming velocity. Cell.141:107-16.

28. Fang X, and Gomelsky M. 2010. A post-translational, c-di-GMP-dependent mechanism regulating flagellar motility. Mol Microbiol 76:1295-305.

29. Navarro MV, Newell PD, Krasteva PV, Chatterjee D, Madden DR, O'Toole GA, Sondermann H. 2011. Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. PLoS Biol. 9(2):e1000588.

30. Newell PD, Boyd CD, Sondermann H, and O'Toole GA. 2011. c-di-GMP effector system controls cell adhesion by inside-out signaling and surface protein cleavage. PLoS Biol 9(2):e1000587.

31. Paul K, Nieto V, Carlquist WC, Blair DF, and Harshey RM. 2010. The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism. Mol Cell 38:128-39.

32. Guvener, Z. T., and C. S. Harwood. 2007. Subcellular location characteristics of the *Pseudomonas aeruginosa* GGDEF protein, WspR, indicate that it produces cyclic-di-GMP in response to growth on surfaces. Mol Microbiol 66:1459-73.

33. Kazmierczak BI, Lebron MB, Murray TS. 2006. Analysis of FimX, a phosphodiesterase that governs twitching motility in *Pseudomonas aeruginosa*. Mol. Microbiol. 60:1026 –1043.

34. Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, Giese B, and U. Jenal. 2004. Cell cycle dependent dynamic localization of a bacterial response

regulator with a novel di-guanylate cyclase output domain. Genes & Dev 18:715-727.

35. Chin KH, Lee YC, Tu ZL, Chen CH, Tseng YH, Yang JM, Ryan RP, Y. McCarthy JM, Dow A, Wang H, and Chou SH. 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 396:646-62.

36. Monds RD, Newell PD, Gross RH, and O'Toole GA. 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 63:656-79.

37. Hinsa SM, Espinosa-Urgel M, Ramos JL, O'Toole GA. 2003. Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. Molecular Microbiology. 49(4), 905–918.

38. Chatterjee D, Boyd CD, O'Toole GA, Sondermann H. 2012. Structural characterization of a conserved, calcium-dependent periplasmic protease from Legionella pneumophila. J Bacteriol.194(16):4415-25.

39. Boyd CD, Chatterjee D, Sondermann H, O'Toole GA. LapG, required for modulating biofilm formation by Pseudomonas fluorescens Pf0-1, is a calcium-dependent protease. J Bacteriol.194(16):4406-14.

40. Düvel J, Bertinetti D, Möller S, Schwede F, Morr M, Wissing J, Radamm L, Zimmermann B, Genieser HG, Jänsch L, Herberg FW, Häussler S. 2012. A chemical proteomics approach to identify c-di-GMP binding proteins in *Pseudomonas aeruginosa*. J. Microbiol. Methods 88:229–236.

41. Roelofs KG, Wang J, Sintim HO, Lee VT. 2011. Differential radial capillary action of ligand assay for high-throughput detection of protein-metabolite interactions. Proc. Natl. Acad. Sci. U. S. A. 108:15528 –15533.

42. Weinhouse H, Sapir S, Amikam D, Shilo Y, Volman G, Ohana P, Benziman M. 1997. c-di-GMP-binding protein, a new factor regulating cellulose synthesis in Acetobacter xylinum. FEBS Lett. 416(2):207-11.

43. Jenal U, Malone J. 2006. Mechanisms of cyclic-di-GMP signaling in bacteria. Annu, Rev. Genet.40:385-407.

44. Smith KD, Lipchock SV, Ames TD, Wang J, Breaker RR, Strobel SA. 2009. Structural basis of ligand binding by a c-di-GMP riboswitch. Nat. Struct. Mol.Biol.16:1218 –1223.

45. Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR. 2008. Riboswitches in eubacteria sense the second messenger cyclic-di-GMP. Science 321:411– 413

46. Diederen BM. Legionella spp. and Legionnaires' disease. 2008. J Infect. 56(1):1-12.

47. Lindenberg S, Klauck G, Pesavento C, Klauck E, Hengge R. The EAL domain protein YciR acts as a trigger enzyme in a c-di-GMP signalling cascade in E. coli biofilm control. 2013. EMBO J. doi: 10.1038/emboj.

48. Levi A, Folcher M, Jenal U, Shuman HA. 2011. Cyclic diguanylate signaling proteins control intracellular growth of *Legionella pneumophila*. mBio 2:e00316–00310.