Structural elucidation of the multidomain response regulator PleD using X-ray crystallography

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Carmen Chan
aus Grossbritannien

Basel, 2004

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof Tilman Schirmer
Prof Urs Jenal

Basel, den 24. November 2004

Prof Hans-Jakob Wirz
Dekan

Declaration

I declare that I wrote this thesis, **Structural elucidation of the multidomain response regulator PleD using X-ray crystallography**, with the help indicated and only handed it into the Faculty of Science of the University of Basel and to no other faculty and no other university.

Abstract

c-diGMP (bis-(3→5)-cyclic di-guanosine monophosphate) is used extensively in bacteria to control biofilm formation and is lately postulated as a novel secondary messenger. Little is known about the signalling process, nor the control, of this dinucleotide. It is clear, however, that its synthesis is catalysed by the DGC (diguanylate cyclase) domain that contains a conserved GG(D/E)EF sequence motif. Despite its high abundance in bacteria, the structure was until now unknown.

The PleD protein from *Caulobacter crescentus* contains a C-terminal DGC domain, preceded by the input domain D1 and the adaptor domain D2. PleD is a response regulator from the two-component signalling system. The output DGC response relies phosphorylation at the N-terminal D1 input domain. Therefore, the control of c-diGMP signal can be revealed in this multi-domain protein.

The objectives of my PhD work are to (1) reveal the structure of DGC domain, (2) understand the catalytic mechanism of DGC, and (3) understand the regulation of the DGC response through the structure of PleD.

The crystal structure of PleD has been solved in complex with c-diGMP to 2.7 Å. The fold of the DGC domain is similar to adenylate cyclase, but the proposed nucleotide binding mode is substantially different. The crystal packing has suggested that two DGC domains align in a two-fold symmetric way to catalyse c-diGMP synthesis. Hence, PleD is active as a dimer using D1 and D2 domains for dimerisation. The dimer formation is probably caused by phosphorylation at the D1 domain. In addition, the structure shows that two intercalated products bind at the D2-DGC domain interface. Such binding is thought to serve an allosteric purpose by immobilising DGC domain movements and prevent them from forming the active site.

This thesis reports the crystal structure of PleD in complex with c-diGMP, and discusses the implications of the structure on DGC catalysis and on activation and inhibition regulation of DGC activity in PleD. In addition, the thesis describes the preparative investigations and characteri-

sation that have led to structure determination of PleD. These include the design and screening of PleD constructs, the establishment and optimisation of expression and purification, protein characterisation, crystallisation optimisation, and diffraction data collection.

Abbreviations

aa	Amino acid
AC	Adenylate cyclase
AU	Analytical ultracentrifugation
c-diGMP	Cyclic di-guanosine monophosphate
CCP4	Collaborative Computational Project Number 4
CD	Circular dichroism
CF	C-terminal 6xHis fusion of full-length PleD
DGC	Diguanylate cyclase
DLS	Dynamic light scattering
DNA	Deoxynucleic acid
DNAP	DNA polymerase
ϵ	Extinction coefficient
F_c	Calculated structure factor
F_o	Observed structure factor
FOM	Figure of merit
His-tag	Hexahistidine-tag
HPLC	High pressure liquid chromatography
IEF	Isoelectric focussing
MAD	multiwavelength anomalous diffraction
min	Minute
MW	Molecular weight
NCS	Non-crystallographic symmetry
NF	N-terminal 6xHis fusion of full-length PleD
NMR	Nuclear magnetic resonance
OD	Optical density
PDB	Protein Data Bank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulphonyl fluoride
rms	Root-mean-squared

rpm	Rotation per minute
R-factor	Crystallographic residual for working set of reflections
R_{free}	Crystallographic residual for test set of reflections
RR	Response regulator
s	Second
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecylsulphate-polyacrylamide gel electrophoresis
SeMet	Selenomethionine
temp	Temperature
TLS	Translation, libration, screw-rotation

Acknowledgements

I would like to thank my supervisor, Prof Tilman Schirmer, for providing me a very interesting project. I highly appreciate his guidance and support throughout my PhD study, especially during the exciting stage of structure elucidation. I would also like to thank Prof Urs Jenal who initiated the collaboration and made this project possible. Thanks also to Prof Andreas Engel for agreeing on moderating the viva voce.

This project involved people mostly affiliated with the Biozentrum at the University of Basel. I am grateful to Dr Jun-ichi Saito, who was in his sabbatical year at the Biozentrum, for his tremendous help in crystallisation optimisation. Thanks also go to Dr Nicolas Amiot, our collaborator from the Department of Chemistry, Basel, for synthesising the enzymatic product molecules, to Dr Ralf Paul and Dietrich Samoray for their superb contribution to the activity study, Dinesh Palanivelu for collecting diffraction data at the synchrotron facility, and to Dr Paul Jenoe and Ariel Lustig for their expert analyses on mass fingerprinting and analytical ultracentrifugation.

I am indebted to those who were put under time pressure in reading different parts of my dissertation— Dr Thomas Ahrens, Martin Allan, Dr Zora Markovic-Housley, Dr Caroline Peneff, and Dr Jun-ichi Saito. Without their kind support and their painstaking comments and advice, this dissertation would not have been possible. I also thank Dr Thomas Braun for offering me quick help on the computing issues related to the writing up of this dissertation.

A special thank-you to Martin Allan for being most supportive to me during the stage of writing up. I appreciate his company for the long evenings I needed in front of the screen in the lab and his attentiveness in bringing me dinner and snacks to keep me going.

It has been a pleasure in work with the members of the Schirmer Group. Apart from those I have mentioned above, I would like to offer my sincere thank to Dr George Orriss for his unlimited support, both scientifically and emotionally, throughout my PhD. I also treasure the scientific exchanges

that I engaged with many at the Biozentrum, especially with Dr Jochen Köser, Prof Olga Mayans, Mr Kitaru Suda and Dr Patrick Van Gelder.

Times might have been difficult, but thanks to the excellent company I have of Fabian Axthelm, Melanie Boerries, Pilar Garcia-Hermosa, Laurent Kreplak, Sebastian Meier, Reika Watanabe, and many others. Their continuous care and support throughout the years is highly appreciated.

I am grateful to Dr Shy Arkin, the supervisor for my Master's study, for recommending me to come to work in Basel, which turns out to be a fruitful experience. Last but not least, I would like to thank my family for their unconditional support at all times.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

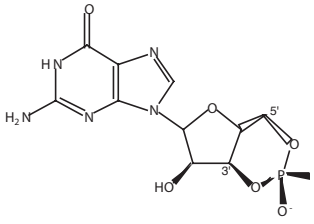
c-diGMP as a novel secondary messenger

Cyclic nucleotides like cyclic adenosine monophosphate (cAMP) or cyclic guanosine monophosphate (cGMP) (Figure 1.1) have been recognised as important low molecular weight signalling molecules. While bacterial pathogens can interfere with cGMP signalling of their eukaryotic host cells [65], prokaryotes in general do not seem to use cGMP for signalling. This suggests an alternative signalling molecule. In fact, the cyclic bis-3→5-dinucleotide c-diGMP has been shown to regulate cell surface associated traits in bacteria including *Caulobacter crescentus* [1, 23], *Escherichia coli*, and the pathogenic bacteria *Pseudomonas aeroginosa* and *Salmonella typhimurium* [56], and community behaviour like biofilm formation in pathogenic bacteria including *Pseudomonas fluorescens* [57] and *Vibrio cholerae* [64]. The recent suggestion that c-diGMP might be a novel secondary messenger [10, 18, 26] has led to mounting interest in the regulation of this cyclic di-nucleotide.

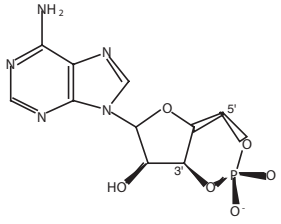
The structure of c-diGMP

The structure of c-diGMP is known [16, 39]. In the crystal structure, two c-diGMP molecules intercalate to form a dimer that is stabilised by a hydrated Mg^{2+} ionon the pseudo two-fold symmetry axis (Figure 1.2). The dimer structure is further stabilised by the alternate stacking of guanines from each monomer. In addition, there is also a set of parallel hydrogen bonds between the guanine and the phosphate of the other monomer.

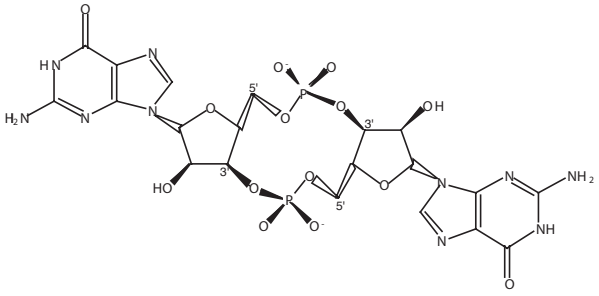
Figure 1.1: Cyclic mono-nucleotides like cAMP and cGMP have an intramolecular phosphodiester linkage, whereas cyclic di-nucleotides like c-diGMP contain two intermolecular phosphodiester linkages to form a large 12-membered cyclic structure.



cGMP

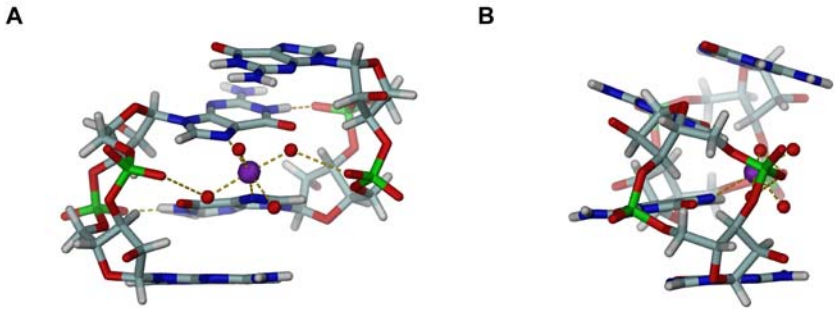


cAMP



c-diGMP

Figure 1.2: Crystal structure of c-diGMP showing two intercalated c-diGMP molecules [16, 39]. A, Front view shows the alternate stacking of the guanine bases coordinated by the hydrated Mg^{2+} ion (purple). Water molecules are in red. B, Side view shows the cyclic structure with the two phosphodiester linkages. All diagrams of the organic and protein molecules in this report are produced in the programme DINO [48].



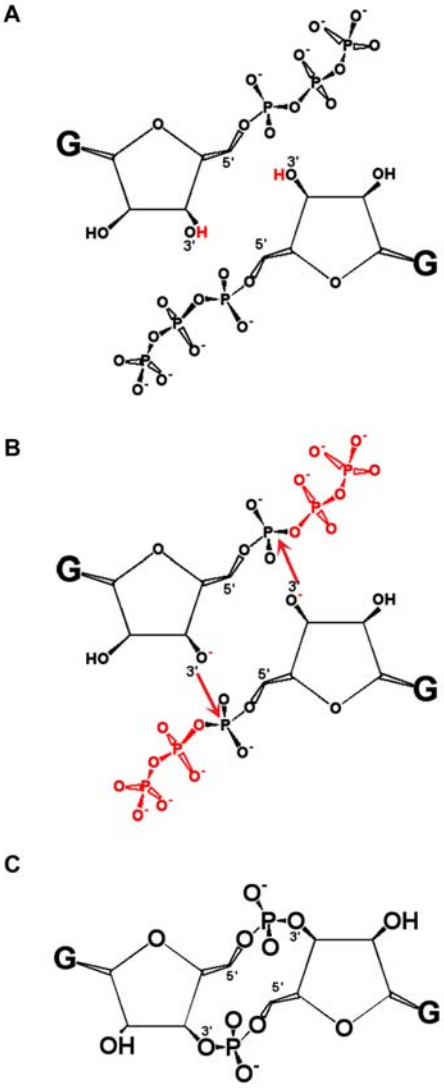
DGC catalyses c-diGMP synthesis

c-diGMP is a cyclic molecule composed of two guanosine monophosphates that are connected by two intermolecular 3'→5'-phosphodiester bonds (Figure 1.1). It is synthesised by the enzyme diguanylate cyclase (DGC) according to the following reaction [54]:

$$2\ \text{GTP} \quad \rightarrow \quad \text{c-diGMP} + 2\ \text{PPi} \tag{1.1}$$

Two guanosine 5'-triphosphate (GTP) substrate molecules are converted by DGC in a condensation reaction to form a c-diGMP molecule and two pyrophosphates (PP_i). The ribosyl 3'-hydroxyl group is first deprotonated to allow for a nucleophilic attack on the α-phosphate of the second GTP molecule. A 3'→5'-phosphodiester bond is formed, and β- and γ-phosphates subsequently leave as pyrophosphate. The c-diGMP synthesis requires mutual attack of the two GTP molecules to give a cyclic di-nucleotide plus two pyrophosphates.

Figure 1.3: c-diGMP synthesis. The letter 'G' denotes guanine. A, Deprotonation of the ribosyl 3' hydroxyl groups of the two GTP substrate molecules. B, Mutual attack of the α-phosphates from the ribosy' 3' oxygens. C, c-diGMP and two pyrophosphates (not shown) as products.



High abundance of DGC domain in bacteria yet lack of structural information

DGC activity was recently ascribed to a domain family hitherto known as 'GGDEF' or 'DUF1' domain [45]. The annotation of 'GGDEF' and 'DUF1' was used because this domain family possesses a very conserved GG(D/E)EF sequence motif and was one of the 'domains of unknown function' at the time of domain classification. Now that it is proven to confer diguanylate cyclase activity, we have decided to renew the annotation to 'DGC domain', and I will use this for the rest of this report.

The general importance of the DGC domain is marked by its abundance in the bacterial genomes. A search for DGC in the the SMART domain database [37] showed 1152 hits in September 2004. As shown in Figure 1.4, DGC is mostly found in bacteria and occurs in various combinations with other sensory and/or regulatory modules [18, 19].

Despite the wide distribution and obvious regulatory relevance of DGC domains, *in vitro* functional characterisation of this domain family was only recently carried out [45]. No structural information about this domain family was available, although it was predicted to be homologous to the adenylate cyclase (AC) catalytic domain [61] and the DNA polymerase I (DNAP) palm domain [13] in a threading study [46]. Interestingly, both AC and DNAP I catalyse a very similar chemical reaction as the DGC domain, as the reaction involves a nucleophilic attack from a ribosyl 3' oxygen to a 5' phosphate on a nucleoside triphosphate [62]. However, unlike the intermolecular attack in the DGC catalysis, the attack carried out in the AC reaction is intramolecular, so that the product is a cyclic mono-nucleotide. This is different in the DNAP reaction in which an intermolecular attack is involved in extending the replicating DNA strand with a incoming nucleotide.

PleD from *C. crescentus* as a model for DGC study

We were interested in understanding the catalytic mechanism of the DGC domain. Unravelling the structure of this significant signalling domain would yield information at the molecular level. We have chosen the signalling protein PleD from *Caulobacter crescentus*, which has a DGC domain, as a model system. The asymmetric cell cycle of *C. crescentus* involves a cell differentiation step in which the motile, flagellum-containing swarmer cell loses its flagellum, before a stalk can be produced at the same cell pole to allow the cell to transform into a stalked cell. The transition of the swarmer cell to

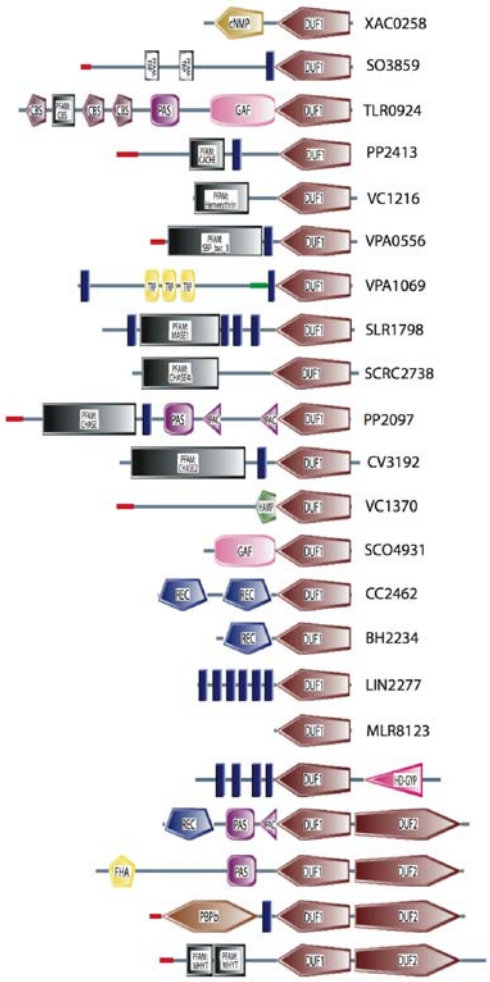


Figure 1.4: A non-exhaustive list of DGC domain containing proteins shown in terms of their domain architecture. The domain symbols were taken from the SMART database, i.e. DGC domain is denoted DUF1 domain. The identity of the protein is listed on the right of the protein representation. PleD from Caulobacter crescentus is labelled CC2462 and consists of two REC (CheY homologous receiver) domains and a C-terminal DGC / DUF1 domain. Domain abbreviations include: CACHE, Ca^{2+} channels and chemotaxis receptors; CBS, Cystathionine β-synthase, prototype for a family of repeats; CHASE, cyclases/histidine kinases associated sensory extracellular; cNMP, cNMP binding domain; DUF2 (EAL), presumable cyclic diguanylate phosphodiesterases; FHA, Forkhead-associated domain; GAF, cGMP-specific and -stimulated phosphodiesterases/ adenylate cyclases (*Anabaena*)/ FhlA (*E. coli*); HAMP, Histidine kinases, adenylyl cyclases, methyl-accepting proteins, phosphatases; HD-GYP, metal-dependent phosphohydrolase; Hemerythrin, oxygen binding protein; MASE, membrane associated sensor (MASE1 and MASE2); MHYT, integral membrane sensor domain; PAC, PAS C-terminal motif; PAS, Drosophila period clock, aryl hydrocarbon receptor, and single-minded proteins; PBPb, high affinity periplasmic solute-binding protein of ABC-type amino acid transport system; Pfam, protein family (Pfam database of alignments and HMMs, $http://www.sanger.ac.uk/Software/Pfam/$); Rec, receiver domain of response regulators; SPB_bac_3, bacterial extracellular solute-binding proteins, family 3; TRP, tetratricopeptide repeat, involved in proteinprotein interaction. The blue vertical bars symbolise membrane-spanning domains. This figure is adapted from [26].

the stalked cell is regulated by PleD as proved by the observed supermotility of the swarmer cell and the inability of the cell to grow proper stalk in a *pleD*$^{-}$ mutant [1, 23].

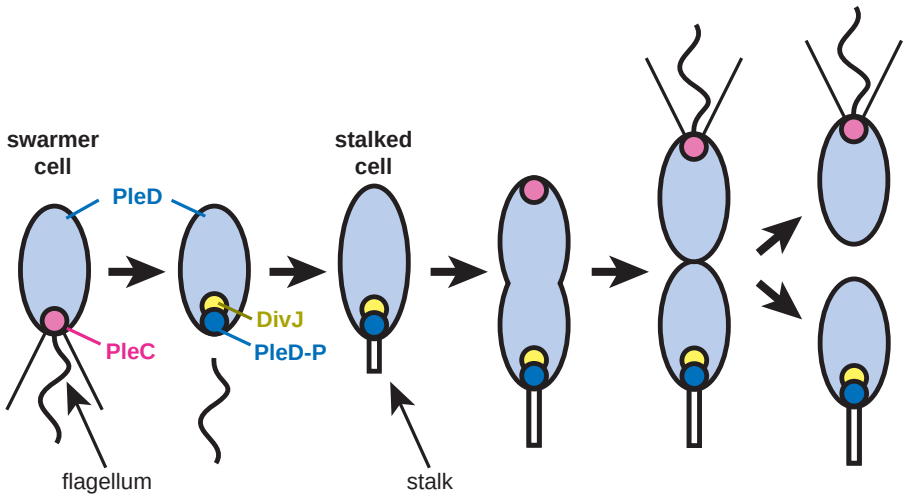


Figure 1.5: PleD temporally and spatially controls the cell differentiation from swarmer cell to stalked cell in the cell cycle of *C. crescentus*. PleD is in coloured in light blue and is distributed throughout the cell. Phosphorylated PleD is coloured in blue and is localised to the stalk pole. Its cognate histidine kinase DivJ is coloured in yellow, and its phosphatase PleC in pink.

Unphosphorylated PleD is inactive and widely distributed over the cell. Upon phosphorylation by its sensor histidine kinase partner DivJ, it becomes activated, sequesters to the differentiating pole of the cell and catalyses the conversion of two GTP molecules to c-diGMP [45]. Thus, PleD activity is temporally and spatially controlled through the coupling of activity and its cellular localisation.

PleD belongs to the two-component signal transduction pathway

PleD belongs to the two-component signal transduction pathway prevalently used in bacteria [50]. In a typical two-component system, information is transferred from the first component, a histidine kinase, to a second component, a response regulator, through a phosphoryl group (Figure 1.6). Upon

receiving an input signal at the N-terminal sensory domain, the C-terminal kinase domain of histidine kinase autophosphorylates on a conserved histidine. The response regulator catalyses the transfer of this phosphoryl group to a conserved aspartate on its N-terminal receiver domain, which subsequently leads to an output signal from its C-terminal effector domain.

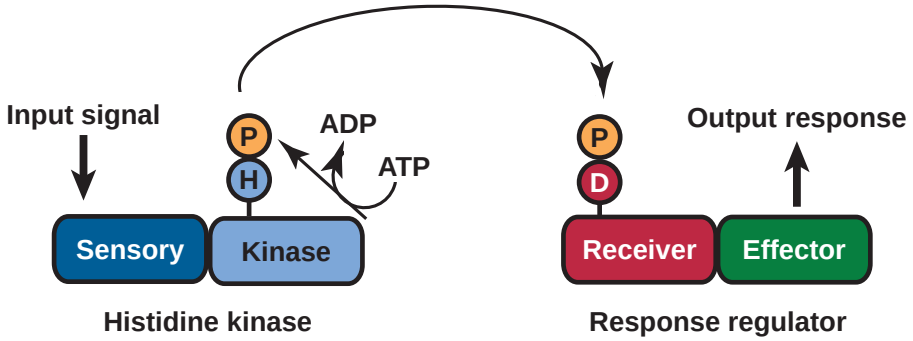


Figure 1.6: Typical domains utilised in phosphotransfer in the two-component signal transduction pathway. This involves the transfer of a phosphoryl group (circled 'P') from a conserved histidine (circled 'H') in the kinase domain to a conserved aspartate (circled 'D' in the receiver domain in the response regulator.)

PleD is an unorthodox response regulator

The response regulators constitute a large protein family. They typically contain a conserved CheY-like receiver domain [35, 58] and a DNA-binding effector domain that functions as a transcription factor [4, 41, 51]. All structurally characterised receiver domains share the structural features of the chemotaxis protein, CheY, from *E. coli* that comprises a doubly-wound, five-stranded parallel sheet structure (Figure 1.8). PleD is an unorthodox response regulator in that it consists of three domains instead of two (Figure 1.7). The N-terminal domain D1 is CheY-like and carries the phosphoaccepting aspartate D53. The middle domain D2 is also CheY-like but it lacks the phosphoacceptor aspartate. The C-terminal effector domain is the domain of interest, DGC. Apart from the methylesterase CheB structure [12], PleD represents the second structure of a multidomain response regulator with proven enzymatic activity.

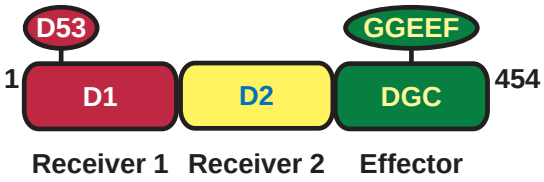


Figure 1.7: Unusual domain arrangement in the multidomain protein PleD. It consists of two receiver domains named D1 and D2, and an effector domain which catalyses c-diGMP formation. The D1 domain carries the phosphoaccepting aspartate D53. The DGC domain carries the conserved sequence motif GGEEF. The starting and the finishing residue numbers are marked.

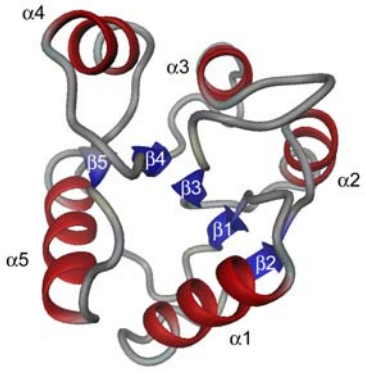


Figure 1.8: Typical response regulator receiver domains adopt the $(\beta/\alpha)_5$ fold as the CheY protein (PDB code 2CHE [59]). Secondary structures are labelled. All helices are annotated as α and all strands as β in this study.

Communication between receiver and effector domains is not well understood

In response regulators, the conformational changes invoked by phosphorylation of the receiver domain have been thoroughly studied on the single-domain CheY protein [35] and other receiver domains like FixJN [4], N-Spo0A [38] and NtrC^r [29]. A common mechanism seems to be employed to propagate structural changes from the phosphorylation site to a large surface covering mainly the C-terminal portion of the domain, involving helices $\alpha 3$, $\alpha 4$ and $\alpha 5$, and β-strands $\beta 4$ and $\beta 5$. Interestingly, subsets of this surface in different response regulators are identified to be involved in protein-protein or domain-domain interactions, some of which might help regulate the function of the effector domain.

For example, the $\alpha 4$-$\beta 5$-$\alpha 5$ surface represents the interaction surface between activated CheY and the N-terminal peptide of its effector protein FliM [36]. The same surface represents the dimerisation interface in the phosphorylated FixJN [4]. As for the few known structures of intact multidomain response regulators, subsets of the $\alpha 4$-$\beta 5$-$\alpha 5$ surface provides the domain interface in CheB [12], DrrB [51] and in NarL [2]. In unphosphorylated CheB, this $\alpha 4$-$\beta 5$-$\alpha 5$ surface obstructs the methylesterase catalytic triad in the effector domain, and thus, suggests an activation mechanism by the relief of active site obstruction [12]. It is, however, less clear in DrrB how this surface might be related to a possible dimerisation mechanism of the C-terminal DNA-binding domain as suggested by the complex structure of a dimeric PhoB effector domain bound with its target DNA [5]. In NarL, the receiver domain uses this surface to block the DNA recognition helix of the effector domain [2], hence preventing it from binding to the target DNA. Phosphorylation of its receiver domain leads to the disruption of this domain interface [17] in a comparative study using NMR. This is to allow the C-terminal effector domain to dimerise to bind to the target DNA [41].

Despite the distinct mechanisms that have been proposed to activate some of the response regulators, molecular details that describe such mechanisms are missing. There is a need for structures of a multidomain response regulator in both its active and inactive forms.

Objectives of this project

Based on our interest in the role of DGC domain on the regulation of c-diGMP signalling, we asked the following questions:

- What is the structure of DGC?
- What is the catalytic mechanism of DGC?
- How does the domain architecture in PleD help in translating a phosphorylation signal into a DGC response?

The main aim of this project was to address these questions by determining the 3-D structure of the DGC domain or the intact response regulator PleD using X-ray crystallography. The method of X-ray crystallography has been chosen due to its strength in dealing with medium-size proteins like PleD (predicted MW of about 50 kDa) or larger complexes, which are otherwise too big for nuclear magnetic resonance, a method that is generally limited to proteins under 30 kDa. To this end, the prerequisite was to obtain a PleD or DGC domain construct that can be over-expressed, purified to high purity while being soluble and stable in solution, and is amenable to crystallisation to give diffraction-quality crystals. Previous constructs made in the laboratory of our collaborators suffered from poor expression and high aggregation. I have, therefore, set the following objectives:

- Design PleD constructs that enable studies on both the full-length protein and the DGC domain.
- Screen PleD constructs for their level of expression and solubility.
- Establish the purification procedures to obtain PleD/DGC of high purity for characterisation studies and crystallisation.
- Characterise PleD constructs to obtain physicochemical information on PleD that might provide clues useful for crystallisation.
- Carry out limited proteolysis study on the full-length PleD constructs to define the domain boundaries, which might be useful for designing new constructs for certain domains.
- Screen for optimal crystallisation conditions to obtain well diffracting crystals of PleD / DGC.
- Solve the structure using X-ray crystallography.

Many of these objectives were met at the end of the project. The highlight is our structure of the full-length PleD in complex with its enzymatic product c-diGMP. The complex structure provides detailed information on

the product-enzyme interactions, hence offering insight into the catalytic mechanism of DGC. The full-length structure of the multidomain protein sheds light on the mechanism of activation and product inhibition of PleD. Detailed information regarding the structure is described in our manuscript (attached at the end of this report) which has been accepted by the Proceedings of the National Academy of Sciences, USA.

This report details my attempts in meeting the objectives and evaluates which objectives were achieved. I have divided the chapter of 'Results' into three sections. In the section of 'Selection of PleD constructs', I will describe the basis of the design of constructs, their selection basing on their expression level and solubility, the development of the purification procedures and some characterisation of the constructs. The second section of 'Structural determination of PleD by X-ray crystallography' focusses on crystallisation screens of the constructs, procedures involved in structural elucidation, and analysis of crystal packing. The final section of 'Structural analysis' provides structural information about the complex structure of the full-length PleD. I will explain in the chapter of 'Discussion' our model of the activation mechanism of PleD and the allosteric inhibition of the enzyme by its product.

Chapter 2

Materials and Methods

Bioinformatics analysis

Analysis of the primary protein sequences, including pI, molecular weight, extinction coefficient calculations / predictions, were carried out using the Proteomics and Sequence Analysis Tools at the ExPASy Proteomics Server [20]. The secondary structure of PleD was predicted using the PredictProtein Server, also from ExPASy. Analysis of the domain arrangement in PleD was carried out using SMART [37] and InterPro linked to ExPASy.

The protein sequences of PleD homologues were searched using the EMBnet server (www.ch.embnet.org). Multiple sequence alignment was carried out using the default settings in ClustalW [63] and the results were presented using ESPript [22].

The overall and residue-by-residue geometry of the PleD structure was analysed using the programme PROCHECK [32].

Subcloning

Full length hexahistidine fusions of wildtype *pleD* were subcloned using the plasmid pRP45 [45], which bears a silent G489T mutation in the *pleD* gene it carries, as template. Full length hexahistidine fusions of *pleD** [45] were subcloned using the plasmid pRP60 [45] as template. Truncated His_6 fusions were subcloned from the full length constructs.

DNA encoding PleD (TrEMBL code: Q9A5I5) was amplified by PCR using the plasmid pRP45 as template with PleD specific primers containing an NdeI restriction site at the 5' end and an EcoRI restriction site at the 3' end. Hexahistidine codons were inserted in the PCR primer after the

NdeI restriction site for the production of an N-terminal hexahistidine fusion, or before the EcoRI restriction site for the production of a C-terminal hexahistine fusion. The PCR fragments were digested with NdeI and EcoRI and subcloned independently into NdeI-EcoRI cut pRUN expression vectors that were derived from the pBR322 vector. The hexahistidine fusion proteins were overexpressed in the *Escherichia coli* strain BL21 (DE3) pLysS.

Overexpression

Cell cultures were grown in LB medium with the addition of 0.1 mg / mL ampicillin at 30 °C until they reached an OD_{600} of 0.5. They were induced with 0.4 mM IPTG and were left to grow for a further three hours. The cells were then harvested and washed in 1/25 culture volume of TNCP pH 8.0 buffer (20 mM TrisHCl, pH 8.0, 500 mM NaCl, 'Complete', EDTA-free protease inhibitor cocktail tablet used according to manufacturer's instruction (Roche Diagnostics AG, Rotkreuz, Switzerland), 0.01 % phenylmethylsulphonyl fluoride (PMSF)) before finally resuspended in the same amount of the same buffer. The cell resuspension was immediately frozen for overnight storage at -80 °C.

Purification

Cells were thawed in a water bath and lysed using a small probe sonicator (Misonix Inc, New York, USA) with pulses for 6 x 30 sec at 50 % output. The cell lysate was clarified by centrifugation in a TFT 70.38 rotor at 50 000 rpm at 4 °C for 30 min (Kontron Instruments, Switzerland). The supernatant was then loaded onto a HiTrap chelating HP column (Amersham Biosciences, Freiburg, Germany) charged with 0.1 mM $NiSO_4$C and pre-equilibrated with TNMI pH 8.0 buffer (20 mM TrisHCl, pH 8.0, 500 mM NaCl, 5 mM β-mercaptoethanol, 50 mM imidazole). PleD was eluted on the ÅKTA Purifier at approximately 200 mM imidazole using an gradient of 50-500 mM. The fractions referring to the 200 mM elution were pooled and dialysed against TND pH 8.0 buffer (20 mM TrisHCl, pH 8.0, 100 mM NaCl, 1 mM DTT) using Spectra-Por 7 dialysis membrane of cut-off 25 kDa at 4 °C overnight. The protein solution was concentrated to no more than 40 mg/ mL using an Amicon Ultra device with a cut-off of 10 kDa (Millipore AG, Volketswil, Switzerland). After clarification in a Beckman TLA 100.2 rotor (Beckman Coulter GmbH, Krefeld, Germany) at 50 000 rpm for 15 min, the sample was loaded on a Superdex 200 HR 10/30 gel

filtration column (Amersham Biosciences Europe, Freiburg, Germany) pre-equilibrated with the same buffer. PleD appeared as monomer and fractions contributing to this peak were pooled. The protein was concentrated in the same way as before for crystallisation.

Selenomethionine-substituted protein preparation

Selenomethionine-substituted PleD was expressed using the metabolic inhibition pathway as described previously [14]. Briefly, cell culture was grown in M9 medium with the addition of 0.1 mg / mL ampicillin at 30 °C.

At 15 minutes before the OD_{600} reached 0.5, amino acid supplements were added to the culture which included: L-Lysine, L-Phenylalanine and L-Threonine to 100 mg / mL, L-isoleucine, L-Leucine, L-Valine and L-Selenomethionine to 50 mg / mL. The cells were allowed to grow for 15 min before induction as for the native material. The purification procedures were the same as for native PleD.

Protein concentration determination

Based on the Beer-Lambert relation [15], the concentration of purified PleD and its co-purified c-diGMP molecules were calculated from the UV absorption measurements at 280 and 253 nm using the equations 3.4 and cp. The derivation of the equations is explained on page 31.

Polyacrylamide gel electrophoresis and Western blot

SDS-PAGE on 12-20 % gradient gels was performed according to the method of Laemmli [31]. Protein bands were visualised with Coomassie Brilliant Blue staining and the molecular weights they referred to were compared to the LMW-SDS Marker (Amersham Biosciences, Freiburg, Germany). Native-PAGE was performed on 12-20 % gradient gels using the same buffer as for SDS-PAGE but omitting SDS and β-mercaptoethanol.

For western blot analysis, protein samples were electroblotted onto nitrocellulose membranes (BA85, Schleicher & Schuell, Dassel, Germany) using the Bio-Rad Mini-PROTEAN 2 Electrophoresis / Mini Trans-Blot Module. The immunoanalysis was carried out according to the manufacturer's protocol of ECL Western Blotting Analysis System (Amersham Biosciences, Freiburg, Germany). Briefly, non-specific binding sites on the blotting membrane were blocked with Tris-buffered saline (TBS) (20 mM Tris-HCl pH7.6,

137 mM NaCl) containing 5 % skimmed milk powder and 0.1 % Tween. The blot was washed twice for 10 min in TBS-Tween before incubating in TBS containing the primary antibody, monoclonal anti-polyHistidine antibody in mouse (Sigma Chemie, Buchs, Switzerland), at 1:3000 dilution for 1 hour at room temp. The blot was washed twice in TBS-Tween, each time for 10 min, and then incubated in TBS containing the secondary antibody, anti-mouse IgG (Fc specific) peroxidase conjugate (Sigma), at 1:5000 dilution. The blot was twice washed as before and incubated in ECL Western Blotting Detection Reagents (Amersham Biosciences, Freiburg, Germany) mixed at the ratio of 1:1 for 1 min. The blot was subsequently exposed on Kodak X-OMAT XAR-5 radiography film for 15 s to 15 min until protein bands appeared.

Limited proteolysis by trypsin

Purified CF at 5 mg/ mL in the storage buffer was proteolysed by trypsin at the w/w ratio of 5000:1 on ice. Aliquots of 22 μL were removed from the reaction mixture at 10, 30, 60 and 90 min and added to 2.4 μL 1 % PMSF (final concentration 0.1 %) to stop the reaction. The stopped reaction mixtures were then subjected to SDS-PAGE and Western blot analysis aganist the His-tag.

Mass spectrometry

Liquid chromatography (LC) / Mass spectrometric (MS) analysis of the CF protein and its digests were carried out on 100 mm i.d. capillary columns packed with C_{18} material (5 mm particle size, MONITOR, Column Engineering, Ontario, USA). Bound peptides were eluted with a linear 30 min gradient from 0.05 % TFA to 60 % acetonitrile containing 0.05 % TFA at a flow rate of 1.5 ml / min into a micro ion source of a TSQ7000 instrument (Thermo Finnigan, San Jose, USA). A voltage of 1300 V was applied to initiate spraying. The instrument was scanned between 200 to 2000 Da m/z in 3 s at unit resolution. All mass spectrometric measurements were performed by the group of Dr Paul Jenoe, Biozentrum, Basel.

Mass fingerprinting

Trypsinsed fragments of CF were separated using SDS-PAGE. Fragments in gel representing proteolysis intermediates were further trypsinised. Trypsinised

fragments belonging to each intermediate were subsequently subjected to matrix-assisted laser desorption ionisation-time of flight (MALDI-TOF) mass spectrometric analysis. MALDI-TOF mass spectra were acquired on a Bruker Reflex III instrument (Bruker Daltonik, Bremen, Germany). Peptides were analysed either in linear or in reflector mode by using a-cyano-4-hydroxycinnamic acid (1 mg / ml in 80 % acetonitrile / 0.1% TFA) as matrix. Samples were prepared by mixing 1 ml peptide solution with 1 ml matrix solution and 300 nl were deposited onto anchor spots of a Scout 400 mm / 36 sample support (Bruker Daltonic, Bremen, Germany). The droplet was left to dry at room temperature. The instrument was calibrated with angiotensin II, substance P, bombesin, and ACTH-18-39. For each proteolysis intermediate, the molecular weights of the ejected species were measured and searched in the sequence library of the theoretically trypsinised CF using the programme MASCOT [47]. These fragments were used to map the boundary of the proteolysis intermediate in the protein. The mass fingerprinting analysis was carried out by the group of Dr Paul Jenoe, Biozentrum, Basel.

N-terminal peptide sequencing

For protein identification with N-terminal sequencing , protein samples were blotted from SDS-PAGE onto polyvinylidene difluoride membranes. Bands corresponding to the protein of interest were cut and sent to Analytical Research and Services, University of Bern, for N-terminal sequencing.

Absorption spectroscopy

Protein concentrations were determined with absorption spectroscopy at 280 nm. The extinction coefficients ϵ of the constructs were calculated from the amino acid sequence [21] and listed in Table 3.1. The extinction coefficient of c-diGMP was experimentally determined to be 16 000 M^{-1} at 254 nm by Dietrich Samoray, Biozentrum, University of Basel.

Isoelectric focussing

Isoelectric focussing (IEF) of CF was performed during the EMBO PEPC workshop, EMBL Hamburg, 2002. A Bio-Rad mini, pre-cast IEF gel of the pH range 3-10 was run using the Bio-Rad Mini-PROTEAN 2 system.

Circular dichroism

Circular dichroism (CD) measurements were carried out using a Cary 61 spectropolarimeter equipped with a thermostatted quartz cell (Hellma, Muellheim, Germany). Spectra were recorded at 20 °C in a quartz cell with a path length of 1 mm with 400 μL protein of 0.1 mg / mL in 20 mM NaPi, pH 8.0, 100 mM NaCl. Scans were recorded over the range of 200-259 nm at the rate of 0.15 nm / s.

For the thermal stability experiment, CD was measured at the fixed wavelength of 221 nm while changing the temperature at the rate of 1 °C / min using a water bath (Lauda RC3) together with a temperature programmer. The temperature was raised from 4 to 90 °C and lowered to 4 °C. The dead time of the cell was 5 s.

Analytical ultracentrifugation

Analytical ultracentrifugation analysis was performed on NF and CF samples in the protein storage buffer. Sedimentation velocity (SV) and sedimentation equilibrium (SE) runs were carried out in a Beckman XLA analytical ultracentrifuge equipped with absorption optics. The SV runs were performed at 54 000 rpm at 20 °C using a 12 mm double sector cell. The SE runs were performed at 12 000 and 18 000 rpm at at 20 °C. The SE results were analysed using a floating baseline computer programme that adjusts the baseline absorbance to obtain ln A versus r^2, where A is the absorbance and r the radial distance. A specific volume of 0.73 cm^3 g^{-1} and a solution density of ρ=1.003 g cm^{-3} was assumed. All analytical ultracentrifugation experiments were performed by Ariel Lustig, Biozentrum, University of Basel.

Dynamic light scattering

Dynamic light scattering of CF was performed using a Protein Solutions device in the EMBO PEPC workshop, EMBL Hamburg, 2002. Measurement of 50 μL of CF at 1 mg / mL sample in the storage buffer was recorded.

Chemical synthesis of c-diGMP

c-diGMP was chemically synthesised by Dr Nicolas C. Amiot, Department of Organic Chemistry, Basel, according to the procedures described by Ross

and co-workers [53]. c-di-GMP was purified by semi-preparative reversed phase HPLC on a Merck LiChrospher 100 RP18 endcapped (10 μm) column (Merck KgaA, Darmstadt, Germany) at 37 °C. 0.1 M triethyl ammonium carbonate buffer (TEAC) pH 7.0 containing 7.5 % methanol (isocratic conditions) was used as mobile phase at a flow rate of 7.5 mL / min. The separation was achieved on an HP1050 Series and detected at 252 nm.

Reversed phase high pressure liquid chromatrography

Purified CF samples were analysed using 80 μL aliquots loaded onto a Merck LiChrospher 100 RP18 endcapped (5 μm) HPLC column at 37 °C. 0.1 M TEAC pH 7.0 containing 7.5 % methanol (isocratic conditions) was used as mobile phase at 1 mL / min on a Waters Alliance 2690 Separative Module. A Waters 2487 ultraviolet detector was used at 252 nm as detection device (Waters AG, Rupperswil, Switzerland).

Crystallisation

Crystals were obtained at room temperature by vapour diffusion using the hanging drop method. PleD at a nominal concentration of 200 μM (assuming an ϵ_{280} of 9200 $M^{-1}cm^{-1}$ equivalent to 10 mg/mL) in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT, 2 mM $MgCl_2$ and 0.8 mM c-diGMP was mixed with the reservoir (1.0 M glycine pH 9.2, 2 % dioxane and 14.5 % polyethylene glycol 20 000 (PEG 20 k)) at a ratio of 1:1. SeMet-substituted crystals were obtained in the same manner, but using a reservoir solution containing 1.0 M TAPS pH 9.0, 2 % dioxane and 11 % PEG 20 k.

For the native protein crystallisation, clover-leaf like crystals obtained by direct vapour diffusion were crushed on the drop. They were picked up as seeds by streaking with piece of hair and were seeded into a clear drop that was already equilibrated.

Data collection

Cryoprotectants contained the mother liquor with an additional 2-5 % PEG 20 k and 5-15 % ethylene glycol. The crystal was soaked successively in cryoprotectants containing 5 %, 10 % and 15 % ethylene glycol. Each soaking

lasted for 5-10 s. After the last soaking the crystal was flash frozen in liquid nitrogen.

Diffraction data were collected from a single native crystal (about 0.015 mm in diameter) and a single SeMet-substituted crystal (0.060 mm) at the beamline X06SA (PX) at Swiss Light Source, Villigen, Switzerland using cryo-conditions. Before collecting data on the native crystal, the crystal was allowed to anneal by blocking the liquid nitrogen stream for 3 s.

Structural elucidation

Diffraction data were processed with MOSFLM/SCALA [44]. 18 selenium positions were identified using SHELXD [55]. Phase refinement was performed by SHARP/SOLOMON [11] and was followed by two-fold averaging and phase extension using DM [8]. The model was built using interactive graphics in the programme 'O' and refined by using REFMAC5 [44] imposing strict non-crystallographic symmetry (NCS) constraints for the two copies in the asymmetric unit except residues 117, 164, 168, and 404. The entire mainchain was defined by final electron density except residues 137-146, 282-288, and the C-terminal His-tag. The scheme of structural elucidation is summarised in Figure 3.18. The structure was elucidated under the close supervision by Prof Tilman Schirmer, Biozentrum, University of Basel.

Chapter 3

Results

Design of PleD constructs

Recombinant method

For the study of the bacterial PleD protein and its DGC domain, we have applied the recombinant method. Recombinant method allows tailoring of the protein according to the researcher's desire and is, thus, well suited for the study of domains. It offers the possibility of attaching purification tags to the protein which facilitates the purification procedures. Hexahistidine-tag (His-tag) was chosen for easy affinity chromatography [24]. Compared to other affinity tags like glutathione-S-transferase which is a protein on its own, the small size of a hexahistidine tag, which consists of only six amino acids, does not interfere with function and crystallisation ability of many recombinant proteins, and hence, might shorten the protein preparation procedures of having to cleave the tag. In our study, two His-tagged constructs were made for each protein sequence of interest, one of them was tagged at the N-terminus and the other at the C-terminus.

Another advantage of using recombinant proteins is that their production can be boosted by the use of an efficient expression system, which is normally a better alternative to protein extraction from the native source, if the goal is to produce proteins in a large quantity to meet the need of crystallisation. In the case of bacterial PleD, the expression system of choice was *E.coli*.

PleD constructs

All constructs produced in this study are summarised in Table 3.1. They were divided into three categories: constructs covering the full-length PleD,

the DGC domain, and D2-DGC domains. The nomenclature of these constructs are as follows. They were annotated as Nx or Cx, where 'N' annotated N-terminal His-tagged fusion and C annotated C-terminal His-tagged fusion. The 'x' represents the type of protein sequence–'F' represents the full-length protein, and a number signifies the starting amino acid position of that construct. For example, the construct N137 refers to the N-terminal His-tagged domain construct starting at residue 137 and finishing at the last residue 454.

Table 3.1: A summary of the physical properties of the various full-length and domain constructs of PleD. MW refers to the theoretic molecular weight calculated using PeptideMass at the ExPASy Proteomics Server [20] whereas ϵ refers to the theoretical extinction coefficient at 280 nm calculated using the Peptide Property Calculator at www.basic.nwu.edu/biotools/proteincalc.html.

Category	Construct	His_6-tag	Residue range	No of aa	MW (Da)	pI	ϵ
Full-length	wt PleD	-	1-454	454	49593.04	5.68	9200
	NF	N	1-454	460	50415.39	6.04	9200
	CF	C					
	PleD*	-	1-454	454	49607.05	5.87	10480
	NF*	N	1-454	460	50429.40	6.18	10480
	CF*	C					
DGC	N287	N	287-454	175	19021.77	6.18	3960
	C287	C					
	N319	N	319-454	143	15526.93	5.95	2680
	C319	C					
D2-DGC	N137	N	137-454	325	35538.60	6.65	6520
	C137	C					
	N150	N	150-454	312	34215.83	6.19	6520
	C150	C					
	N153	N	153-454	309	33930.70	6.32	6520
	C153	C					

Full-length constructs

The first category belongs to the full-length PleD, which includes the wild-type and the constitutively active mutant PleD* [45]. The wild-type PleD sequence has the Q9A5I5 entry in the SwissProt/TrEMBL database [6]. The constitutively active mutant PleD* contains the following mutations– T120N, T214A, E220A, P234H and N357Y. It leads to elongated stalks [1], has a dominant negative effect on motility of the cell, and localises to the cell pole in the absence of its kinase DivJ and phosphatase PleC [45]. It shows a higher DGC activity than the wild-type regardless of the disruption of

the phosphorylation site at D53 in the PleD*D53N mutant, thereby proving that PleD* mimicks the activated state of the wild-type PleD. Nevertheless, initial investigation on PleD* showed that it suffered from very poor expression and high aggregation in solution, and was, therefore, shelved. In this report, I only focus on wild-type PleD.

DGC constructs

The second category belongs to DGC constructs which include the N237 / C237 and N319 / C319 constructs. The N319 / C319 constructs were designed based on sequence alignment and threading experiments by Pei and Grishin [46]. They observed a weak structural homology between the GGDEF-motif containing sequences and the adenylate cyclase (AC) catalytic domain after an extensive iterative search in the protein sequence database. From the result of a multiple sequence alignment they concluded that GGDEF-motif containing sequences share the same fold as AC catalytic domain. Accordingly, the DGC domain of PleD would start at residue 319.

Beside the N319 / C319 constructs, we have designed other DGC constructs starting from residue 287 with the following considerations. The integrated domain database InterPro assigned the domain arrangement of 'response regulator receiver'–'response regulator receiver'–'GGDEF' to PleD ('GGDEF' domain is equivalent to 'DGC' domain as explained before). The prediction of the domain boundaries was as follows: The first response regulator receiver (RRR) domain ranges from residue 1 to residue 120 or 130; the second RRR domain ranges from residue 160 to residue 280 or 290; the DGC domain ranges from residue 280, 290 or 330 to residue 454. The prediction of the DGC domain was particularly unclear.

A search using the programme METAMOTIF from the EMBnet server (www.ch.embnet.org) for protein sequences with the arrangement of two 'RRR' domains and a 'DGC' domain resulted in seven hits. As the structure of CheY was known, we wanted to use the CheY sequence as a ruler to phase the RRR and the DGC domains. Considering CheY proteins contain around 120 residues, we arbitrarily assigned residues 1- 150 in each homologous sequence as the first putative RRR domain and residues 130-300 as the second putative RRR domain. Four of these putative RRR sequences were randomly selected and then multiply aligned (Figure 3.1).

It was found that the second putative RRR sequences could be aligned with the first RRR sequences starting from residue 150. This suggested the second RRR domain started at around residue 150. The second finding was that the C-terminal helix of CheY could be mapped to a helix pre-

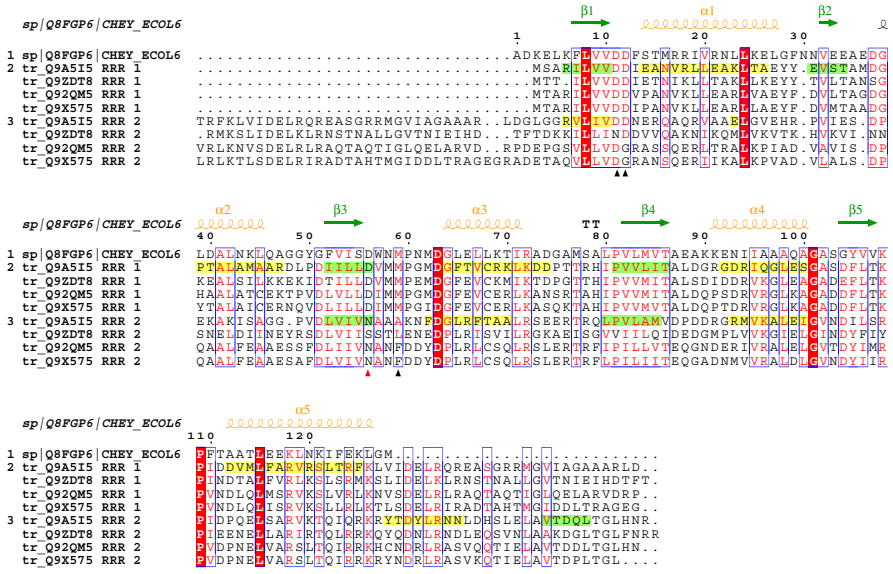


Figure 3.1: Prediction of DGC domain boundary in PleD by multiple sequence alignment and secondary structure prediction. CheY from *E.coli* (sequence group 1) was aligned with the CheY-like domains of PleD (SwissProt/TrEMBL: Q9A5I5) and other PleD homologues (SwissProt/TrEMBL: Q9ZDT8, Q92QM5, Q9X575). The sequences aligned fall into three groups: group 1, CheY; group 2, the first putative RRR domain, or RRR 1, of the PleD and PleD homologous sequences; and group 3, the second putative RRR domain, RRR 2. The numbering of the sequence and the secondary structures shown at the top of the alignment are from the CheY protein (PDB accession code 1DJM). Secondary structures predicted for PleD are shown in the background of the PleD sequence: yellow for helix and green for β-strand (http://cubic.bioc.columbia.edu/predictprotein/). The strictly conserved residues are in red background and the conserved residues are in red letters. The phosphoacceptor D56 is annotated with a red arrow and the acidic cluster, including M59 whose mainchain amide is used for phosphorylation, is annotated with a black arrow at the bottom of the alignment. The figure was produced by ESPript [22].

dicted in PredictProtein (http://cubic.bioc.columbia.edu/predictprotein/), ranging from residues 274 to 280 in the PleD sequence. We concluded that D2 domain would finish at residue 280 and that DGC domain would start at the next predicted secondary structure, which was a β-strand at residue 289. There were two polar residues, D282 and E287, in the putative D2-DGC domain linker. We designed the DGC constructs to start at E287.

D2-DGC constructs

The third category belongs to the D2-DGC constructs. Characterisation of the full-length PleD protein using limited proteolysis found two fragments that were resistant to proteolysis. The combination of mass fingerprinting and N-terminal sequencing identified the two fragments to be the D1 domain and a C-terminal fragment starting at residue 156 and thus comprising the D2 and DGC domains. The linker connecting these two experimentally defined domains ranged from residue 130 to 156. We designed constructs that started at the polar residues E137 and D150 and the flexible glycine G153 on this linker for an experimental screen on expression and solubility. More details about the design of these constructs follow in the next section.

Overexpression, purification and characterisation of PleD

Expression and solubility test of full-length PleD

A trial was carried out to test the expression level and the soluble yield of the full-length NF and CF constructs as a function of growth temperature, time of induction and duration of expression. The bacteria were grown at 30 °C and 37 °C, induced at an optical density at 600 nm (OD_{600}) of 0.4, 0.7 and 1.0 for 1, 2, 3 and 4 hours. The cells were lysed and pelleted to estimate the amount of inclusion bodies and the yield of the soluble fraction.

Higher growth temperature, induction at higher OD, and longer expression time led to more protein going into inclusion bodies. NF and CF were found to give the best soluble yield when grown at 30 °C and induced at OD_{600} of 0.5 for 3 hr.

Purification of full-length PleD

NF and CF were purified in two chromatography steps using nickel affinity chromatography and gel filtration. The supernatant of cell lysate was first purified on a nickel affinity column. For CF purification, the protein was eluted with an imidazole gradient from 50 to 500 mM. Two elution peaks were observed. Peak A was eluted with 120 mM imidazole and appeared to be 80 % pure by estimation from the SDS gel (Figure 3.2A). Peak B was eluted at an imidazole concentration of 250 mM and was smaller and less pure as seen from the contaminants of various sizes shown on the SDS gel. This peak also had a tendency to aggregate during the subsequent overnight dialysis process.

In the second chromatography step of gel filtration, both elution peaks were run individually on a Superdex 200 gel filtration column. This column has a fractionation range from 10 to 600 kDa, and is thus, well suited for resolving full-length PleD which has a calculated monomeric MW of 49.6 kDa. For the run of Peak A from the nickel affinity column, two peaks were eluted. The first peak was eluted at the void volume and showed a major band of CF with contaminants, suggesting this peak was due to protein aggregates (Figure 3.2B). The second peak was eluted at the monomer MW, was bigger than the first peak, and showed a clean band on the gel.

As for the run of Peak B, two peaks were also eluted at similar elution time as observed from the run of peak A. For this run, however, the first peak corresponding to the protein aggregates was much more intense than the

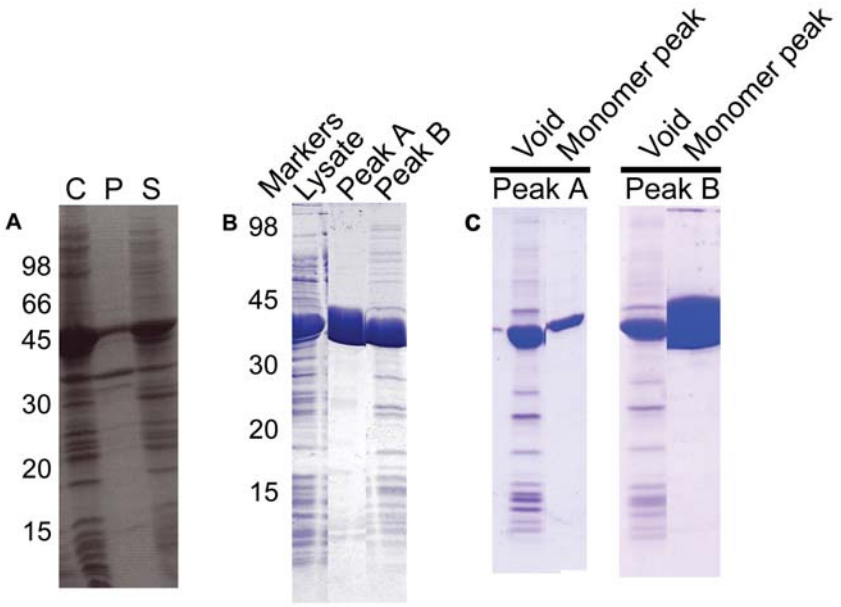


Figure 3.2: SDS-PAGE analysis of CF expression and purification. A, CF was over-expressed in the cells (C) which were spun down to remove the pellet (P) and give the supernatant (S). B, CF lysate supernatant was purified on a nickel affinity column which gave two elution peaks. C, Gel filtration showed that peak A consisted of more monomer proteins than aggregates. Peak B was the reverse.

second monomeric peak. The gel filtration chromatography thus confirmed the difference between the two elution peaks from nickel affinity column. In conclusion, CF existed as a monomer in solution and only Peak A, the peak eluted at 120 mM imidazole, corresponded to monomeric CF in solution.

It was observed that the protein concentration process before gel filtration was crucial in alleviating aggregation. By keeping the protein concentration below 40 mg / mL the elution in the gel filtration run could be shifted towards the monomer peak. Figure 3.3 shows the elution profile of the optimised gel filtration run.

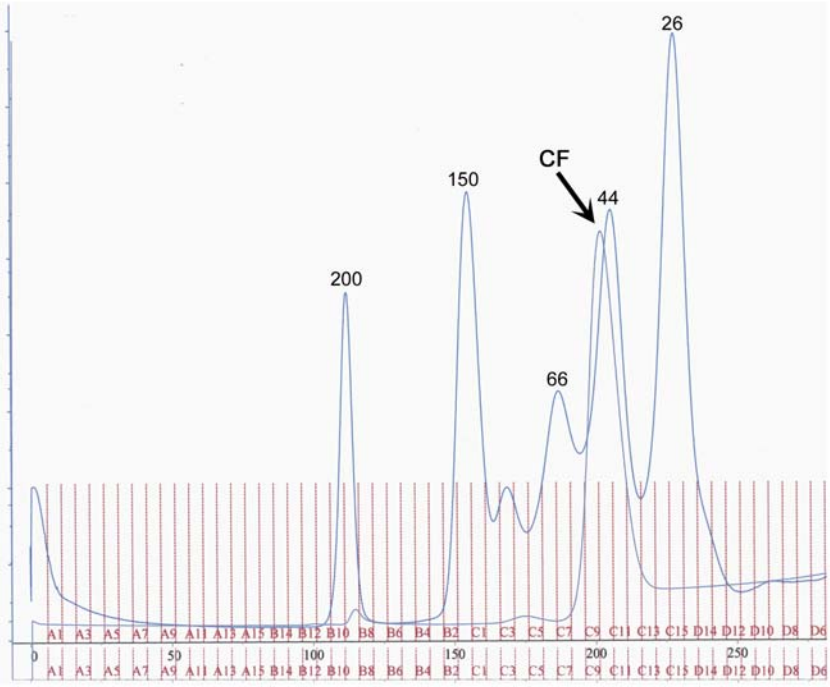


Figure 3.3: Elution profile of the optimised run of CF (peak A from nickel affinity column) on Superdex 200 gel filtration column with an overlaid chromatogram of the calibration run. Elution peaks are labelled according to their respective protein masses of 26, 44, 66, 150 and 200 kDa. The 200 kDa elution peak gives the void of the column. A trapped bubble has contributed to the first elution peak at 0 mL.

For the purification of NF on a nickel affinity column, the elution started already at 100 mM imidazole. Nevertheless, the elution pattern it displayed was very similar to CF (Figure 3.4).

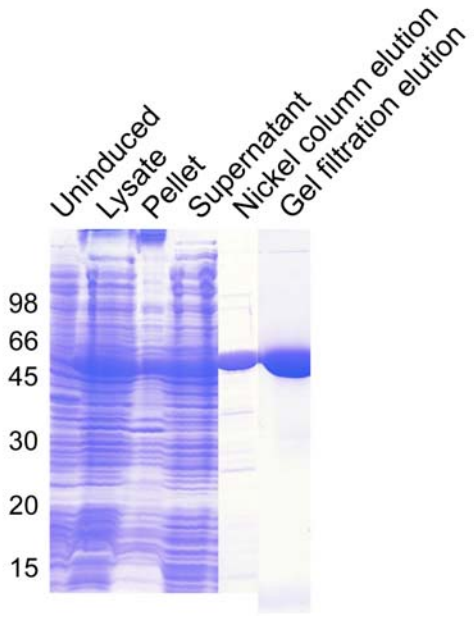


Figure 3.4: SDS-PAGE analysis of NF expression and purification.

Verification of the identity of CF construct

Purified CF was confirmed using mass spectrometric analysis and N-terminal peptide sequencing. Using LC-MS, the molecular weight (MW) of CF was determined to be 50 276 Da, which is 139 Da smaller than the theoretical value. N-terminal sequencing confirmed the N-terminal sequence of CF but found the first methionine missing. This explains the difference in the calculated and experimental values of MW, which is close to the mass of a methionine which is 149 Da. In addition, the identity of CF was also indirectly proven by the positive signal on the Western blot probed with an anti-histidine antibody (Figure 3.5A).

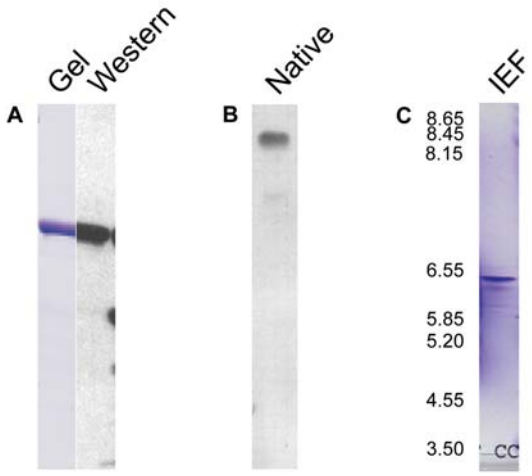


Figure 3.5: Characterisation of the CF construct. A, Western blot of purified monomeric CF showed positive signal when probed with anti-histidine antibody. B, CF gave a single band on the native gel. C, CF showed an experimental pI of 6.3 using isoelectric focussing.

Identification of c-diGMP bound to CF

The UV absorption spectrum of purified CF showed a peak around 260 nm in addition to the peak at 280 nm which is expected for proteins due to absorption contribution by aromatic residues (Figure 3.6B). Because PleD is a di-nucleotide cyclase, we thought the absorption peak at 260 nm might be due to the presence of bound nucleotides, most probably being GTP, c-diGMP or the linear reaction intermediate, diguanosine tetraphosphate (pppG3'p5'G), in the DGC reaction [54]. These nucleotides can be identified

according to their polarities using RP-HPLC. To identify the ligand co-purifying with CF, the CF sample was analysed using RP-HPLC and the chromatogram was compared to the reference chromatograms.

Purified CF gave two elution peaks from the HPLC column (Figure 3.7). The first peak corresponded to that given by c-diGMP, with a retention time of 7.3 min. This indicated that CF carried c-diGMP throughout the purification procedures. The second peak appeared at a later time at 12.4 min. The possibility of the substrate GTP or the linearised dinucleotide intermediate were ruled out since they were probably more polar and could only be eluted at an earlier time. So the peak might correspond to the CF protein itself.

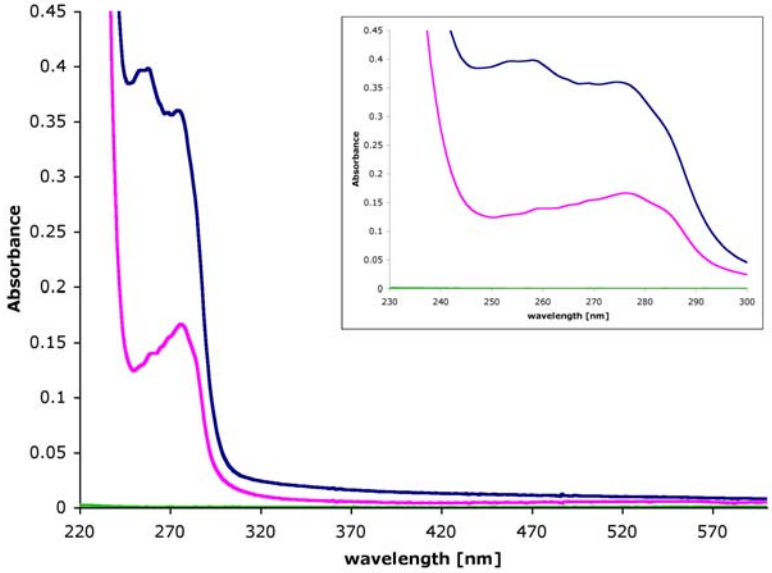


Figure 3.6: UV absorption spectra of purified CF before (red) and after (blue) dialysis at 4°C. The buffer baseline is in green.

Quantification of amount of bound c-diGMP on CF

The concentration of purified PleD was determined using the Beer-Lambert relation [15] by measuring UV absorption.

$$A_n = \epsilon_n * c * l \tag{3.1}$$

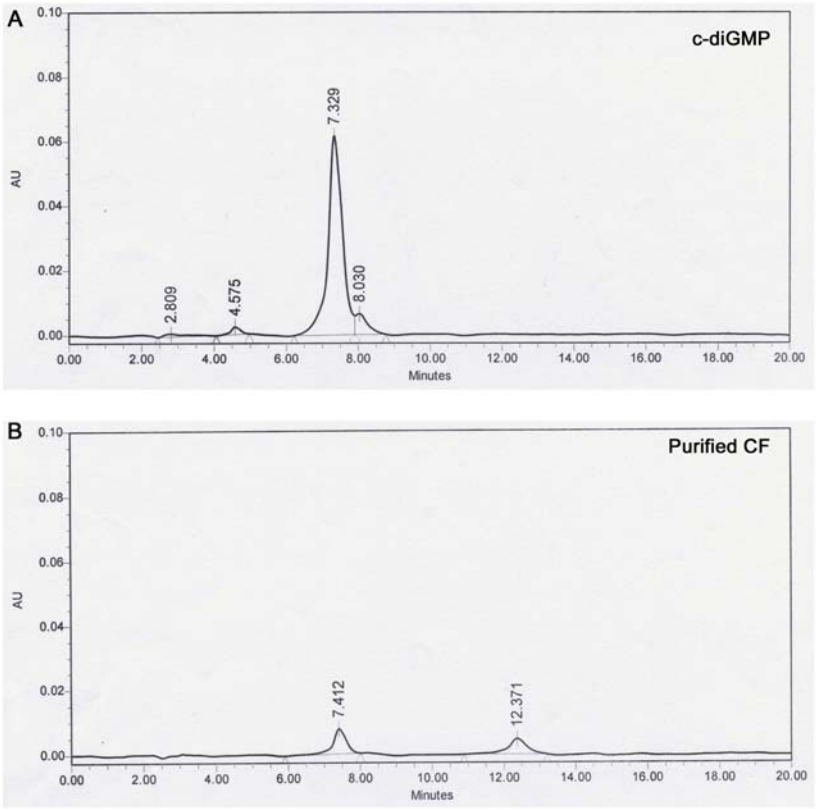


Figure 3.7: Elution profiles on HPLC measured at 252 nm. A, c-diGMP. B, Purified CF sample. The same elution peak at around 7.4 min was obtained.

where A_n = absorbance at n nm
ϵ_n = molar extinction coefficient at n nm
l = pathlength

Both PleD and its co-purified c-diGMP contribute to UV absorption. Therefore, the total absorption of the purified PleD sample is given by the sum of the absorption of both protein and ligand. Since the absorption maxima of PleD and c-diGMP are at 280 and 253 nm respectively, we measure A_{280} and A_{253}:

$$A_{280} = \epsilon_{280p} * c_p * l + \epsilon_{280l} * c_l * l \quad (3.2)$$

$$A_{253} = \epsilon_{253p} * c_p * l + \epsilon_{253l} * c_l * l \quad (3.3)$$

where A_{280} = measured absorbance at 280 nm
A_{253} = measured absorbance at 253 nm
ϵ_{280p} = 9200 $M^{-1}cm^{-1}$ theoretical value for PleD at 280 nm
ϵ_{253p} = 5878 $M^{-1}cm^{-1}$ theoretical value for PleD at 253 nm
ϵ_{280l} = 9600 $M^{-1}cm^{-1}$ for c-diGMP at 280 nm
ϵ_{253l} = 16160 $M^{-1}cm^{-1}$ for c-diGMP at 253 nm
l = 1 for this experiment

Let the molar ratio of ligand : protein be x, i.e. $c_l/c_p = x$, and the measured absorption ratio of A_{253} / A_{280} be R. By dividing equation 3.3 by equation 3.3, x is expressed in function of R, which is measured experimentally, and the extinction coefficients, which are known values.

$$x = \frac{R * \epsilon_{280p} - \epsilon_{253p}}{\epsilon_{253l} - R * \epsilon_{280l}} \quad (3.4)$$

For a purified CF sample, x is around 1. As shown by the complex structure of PleD with c-diGMP which was solved later, two ligand molecules probably co-purify with PleD, hence, the occupancy of bound c-diGMP is x/2. The concentration of the protein can then be deduced by substituting x into equation 3.3, as shown in equation 3.5. The concentration of the ligand is $c_p * x$.

$$c_p = \frac{A_{280}}{\epsilon_{280p} + \epsilon_{280l} * x} \quad (3.5)$$

Hydrodynamic characterisation of CF

CF protein was shown to be a monomer in solution from the purification step of gel filtration chromatography (Figure 3.3). This was confirmed by the following measurements including dynamic light scattering (DLS) and analytical ultracentrifugation (AU).

DLS observes the fluctuations in the intensity of light scattered by particles in solution [9]. The fluctuations reflect translational diffusion that is dependent on the shape, MW and concentration of the particles in the sample. DLS is a useful tool to investigate the physical homogeneity of a purified protein sample. For the measurement of CF sample, a monomodal distribution was observed which indicated the presence of a single species. The measured diffusion coefficient of the species was 630 x10^{-9} cm / s^2, which corresponded to a spherical protein of 61 kDa. This observation was consistent with a monomeric PleD of MW 50 kDa that had a molecular shape deviating from a perfect sphere.

Sedimentation measurements were carried out using AU [33, 49]. Sedimentation velocity run of CF showed a sedimentation coefficient of 3.9 S. Sedimentation equilibrium runs at 18 000 rpm of a sample with the concentration of 0.45 mg / mL gave an estimated MW of 53 kDa.

Conformation and thermal stability of CF

Native-PAGE separates proteins in their native state according to their net charge, mass and shape. The CF protein ran as one band on the native gel (Figure 3.5A). Together with the gel filtration result, this suggested that CF was a monomer in solution.

In circular dichroism (CD) spectra of proteins, peptide bonds dominate the far-UV region, and are useful for characterisation of secondary structure [9]. Figure 3.8A shows the CD spectrum of CF in the far-UV region. The characteristic minimum at 221 nm was found, which was indicative of a structure with considerable α-helical content. Upon heating, the molar ellipticity at 221 nm showed a monophasic transition from -1400 to -800 with a sharp increase of ellipticity around 49 °C. On cooling down, the molar ellipticity remained stable at the value observed at 90 °C. This suggested that the CF structure was irreversibly unfolded, or melted, at around 49 °C.

Isoelectric point determinaton

Isoelectric focussing (IEF) allows the separation of proteins according to their isoelectric point, pI, in the presence of a continuous pH gradient. The

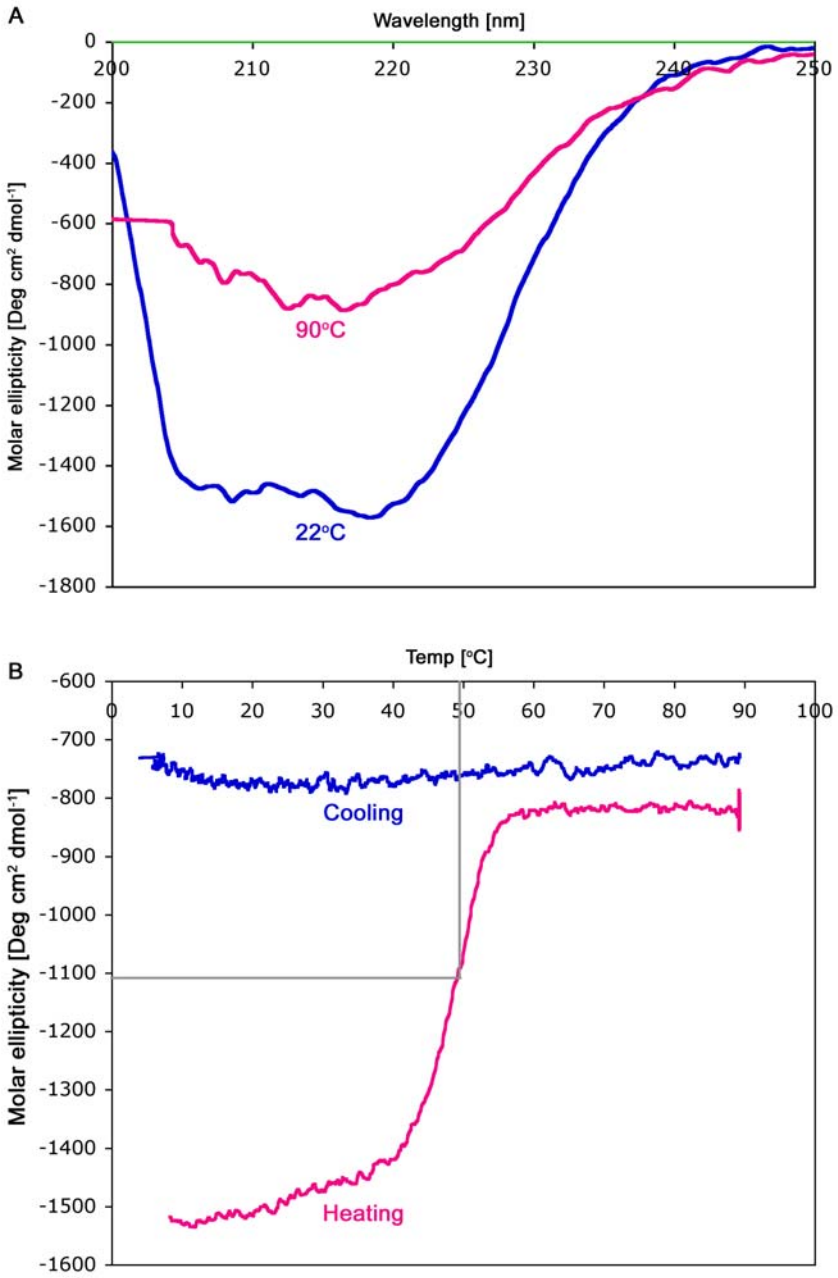


Figure 3.8: Conformation characterisation of CF using CD. A, CF showed the characteristic trough of α-helix at 221nm at 22 °C (blue) which collapsed at 90 °C (pink). The buffer baseline is in green. B, CF showed monophasic transition at around 49°C upon heating (pink). Once melted, it could not renature by cooling (blue).

pI of CF was experimentally determined to be 6.3 (Figure 3.5 B), compared to the theoretical value of 6.0 calculated using the programme PeptideMass on the ExPASy website.

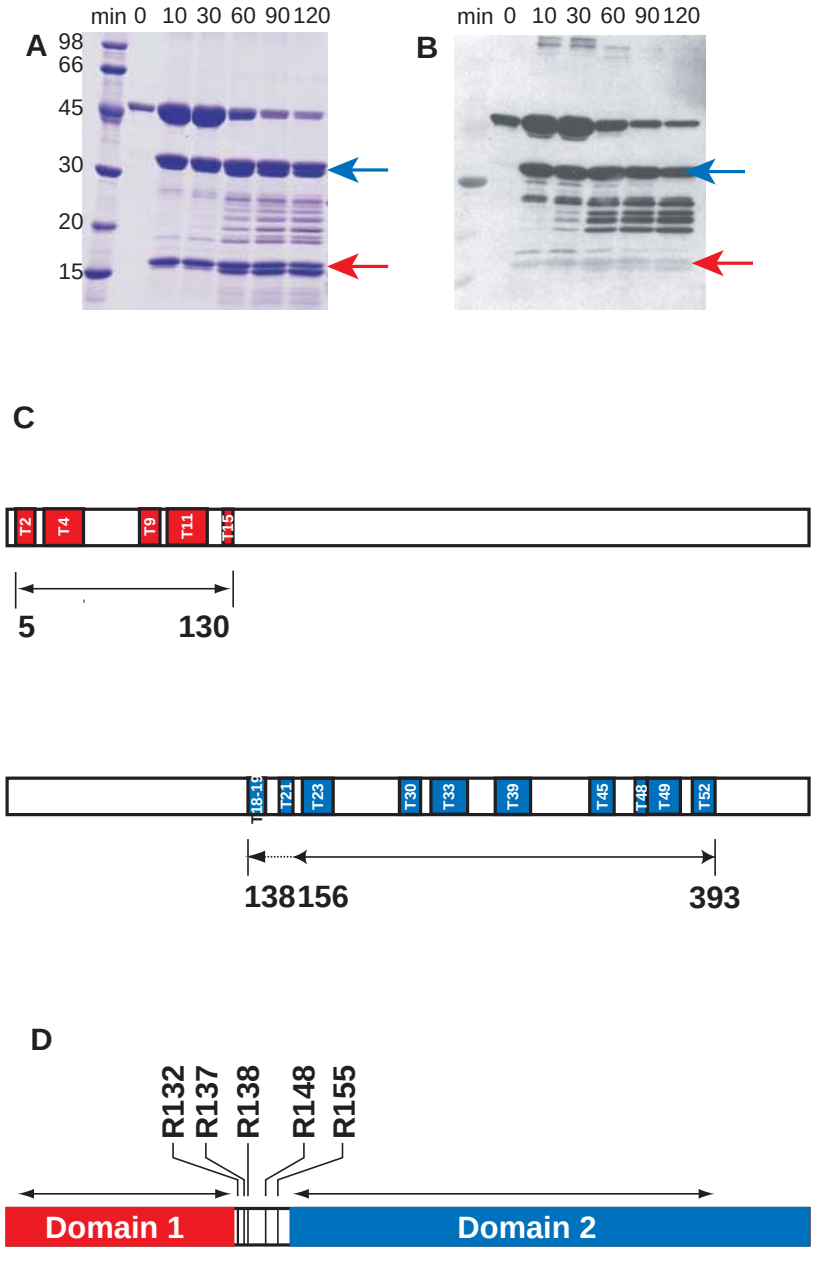
Domain boundaries delineation of PleD

CF was subjected to limited trypsinolysis which yielded two groups of protease-resistant fragments. The larger fragments were around 32 kDa (pointed with blue arrow in Figure 3.9A) and the smaller about 15 kDa (red arrow) as shown in the SDS-PAGE. Each group of intermediates contained two visibly distinctive bands. The bands were isolated and further trypsinised to allow for mass fingerprinting. In this technique, the trypsinised fragments of each band, being small enough for MALDI, were analysed by mass spectrometry. The molecular weights of these fragments were then searched in the sequence library of the theoretically trypsinised CF using the programme MASCOT [47]. The two bands of MW around 15 kDa resulted in tryptic fragments that were mapped to the N-terminal region between residues 5 and 130 on CF (red segments in Fig 3.9 C). On the other hand, the bands belonging to the larger intermediate of around 30 kDa resulted in tryptic fragments that covered the C-terminal region of CF (blue segments in Fig 3.9 C). The lower band was mapped to residues 138-393, whereas the upper band was mapped to residues 156-393. N-terminal sequencing of the upper band confirmed the starting residue to be residue 156.

Western blot of the same SDS gel followed by immunological detection using anti-histidine antibodies showed positive signal for the bands belonging to the larger intermediates, showing that they both carried the C-terminal His-tag from the C-terminal His-tagged CF construct (Fig 3.9 B). This agreed with the finding from electrospray-mass spectrometry analysis of the solution sample of the trypsinised mixture. Two masses of 15 218 Da and 35 117 Da were measured, which were similar to those of the protease-resistant fragments estimated from the gel. The smaller mass of 15 218 Da was closest to the mass of a tryptic fragment of residues 1-138 whereas the larger mass of 35 117 Da was closest to a tryptic fragment of residues 138-460 covering the C-terminal His-tag.

According to above findings, there exists two possible stable fragments (Fig 3.9 D). A conservative prediction was that the first fragment encompassed residues 1-130, covering the D1 domain, and the second fragment

Figure 3.9: Domain delineation by limited proteolysis and mass fingerprinting. A, SDS-PAGE showing a time-controlled proteolysis of CF from the intact protein (single band) to two stable intermediates (arrows). B, The corresponding western blot showed that only the larger intermediates (blue) were recognised by anti-histidine antibody. C, The location of the tryptic fragments that were identified by mass fingerprinting on CF. D, The inferred domain arrangement of CF consisted of two domains separated by a linker that contained the tryptic cleavage sites R132, R137, R138, R148 and R155.



would start at any position between residues 130 and 156 and finally end at residue 460 to cover the D2 and DGC domains. The stretch of amino acids connecting these two putative domains is rather hydrophobic. We have designed three domain constructs starting at R137, D150 and G153. R137 and D150 were chosen since they are polar and would probably favour exposure to the solvent environment. G153 was chosen due to its flexibility.

A follow-up experiment was carried out on one of these domain constructs to check if smaller protease-resistant fragments existed. However, limited proteolysis of the N137 construct did not show any detectable stable intermediates (Figure 3.10).

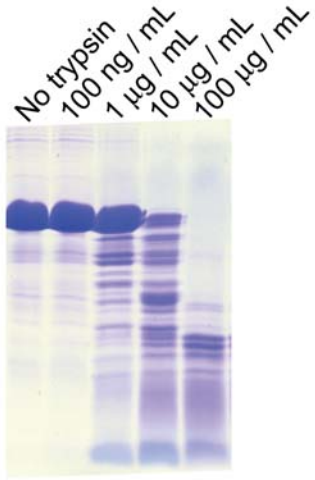


Figure 3.10: SDS-PAGE analysis of the limited proteolysis of N137. Limited proteolysis of the N137 construct using different amounts of trypsin did not show any stable fragments.

Expression and solubility test of D2-DGC constructs

A small-scale expression and solubility screen was performed on the domain constructs N/C137, N/C150 and N/C153. All constructs were expressed in the same way and purified on a nickel affinity column to assess their solubility. Under the tested conditions, all constructs but C153 expressed to yield soluble proteins that could be purified (Figure 3.11). Apart from the 153 constructs, the N-terminal constructs gave 2-3x higher yield than their C-terminal counterparts.

Due to the general positive results of the expression and solubility of the D2-DGC constructs, there was a plan to perform crystallisation using these constructs, particularly N/C137 and N/C150. However, diffraction of the CF protein was obtained at that time, and thus the plan was shelved.

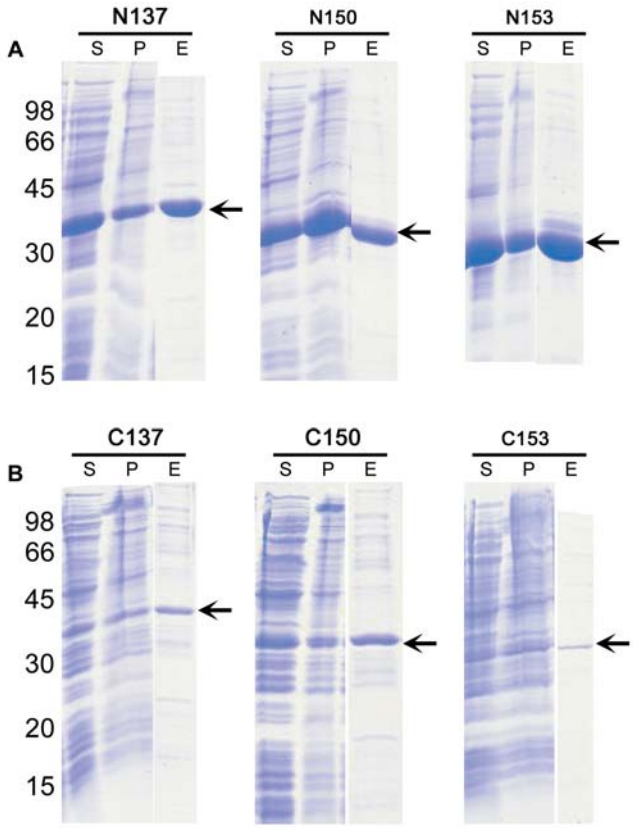


Figure 3.11: Expression and solubility test of D2-DGC constructs. S denotes supernatant, P denotes pellet, and E denotes elution from nickel affinity column. The arrow shows the expected size of the construct on the gel. A, N-terminal His-tagged constructs. B, C-terminal His-tagged constructs.

Expression and solubility test of DGC constructs

The C287 construct was purified in a similar way as for the full-length PleD constructs using both nickel affinity and gel filtration chromatography. As

in the CF preparation, the cell lysate was passed through a nickel affinity column. Gradient elution with imidazole showed elution peaks at 200 mM and 500 mM. The 500 mM elution was highly aggregated and the aggregation could be observed by eye straight after elution. The 200 mM elution was 50 % pure (Figure 3.12B) and tended to precipitate during the protein concentration step (Figure 3.12C). Approximately 20 % of the soluble fraction was recovered and further purified on Superdex 200. Out of this minute fraction of proteins, the majority eluted as void or as contaminants of variable sizes. Only a very minute fraction corresponded to the monomer MW of C287 and could only barely be seen on a silver-stained gel (not shown).

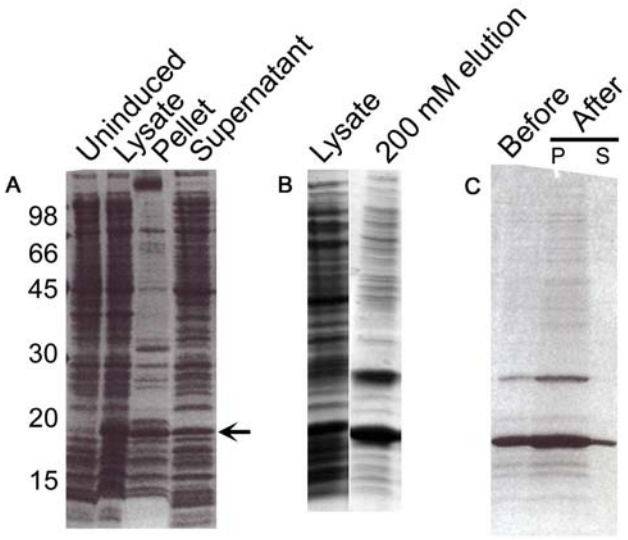


Figure 3.12: Expression and Purification procedures of C287. A, C287 showed moderate level of expression. 50 % was soluble as seen from the lysate supernatant. B, Lysate supernatant was purified on a nickel affinity column which gave an elution at 200 mM imidazole that contained plenty of contaminants. C, Concentrating this peak resulted in the majority going into pellet.

The other DGC constructs, N319 / C319, inferred from the Pei and Grishin study [46] also suffered from poor solubility and could not be purified on the nickel column (Figure 3.13).

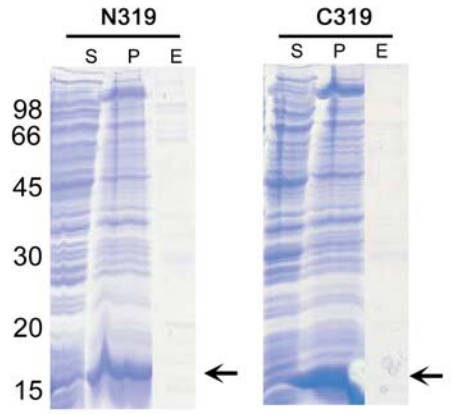


Figure 3.13: Expression and solubility test of N319 / C319 constructs. S denotes supernatant, P denotes pellet, and E denotes elution from nickel affinity column. Both constructs tended to form inclusion bodies and could not be purified to the amount observable by SDS-PAGE.

Structural determination of PleD by X-ray crystallography

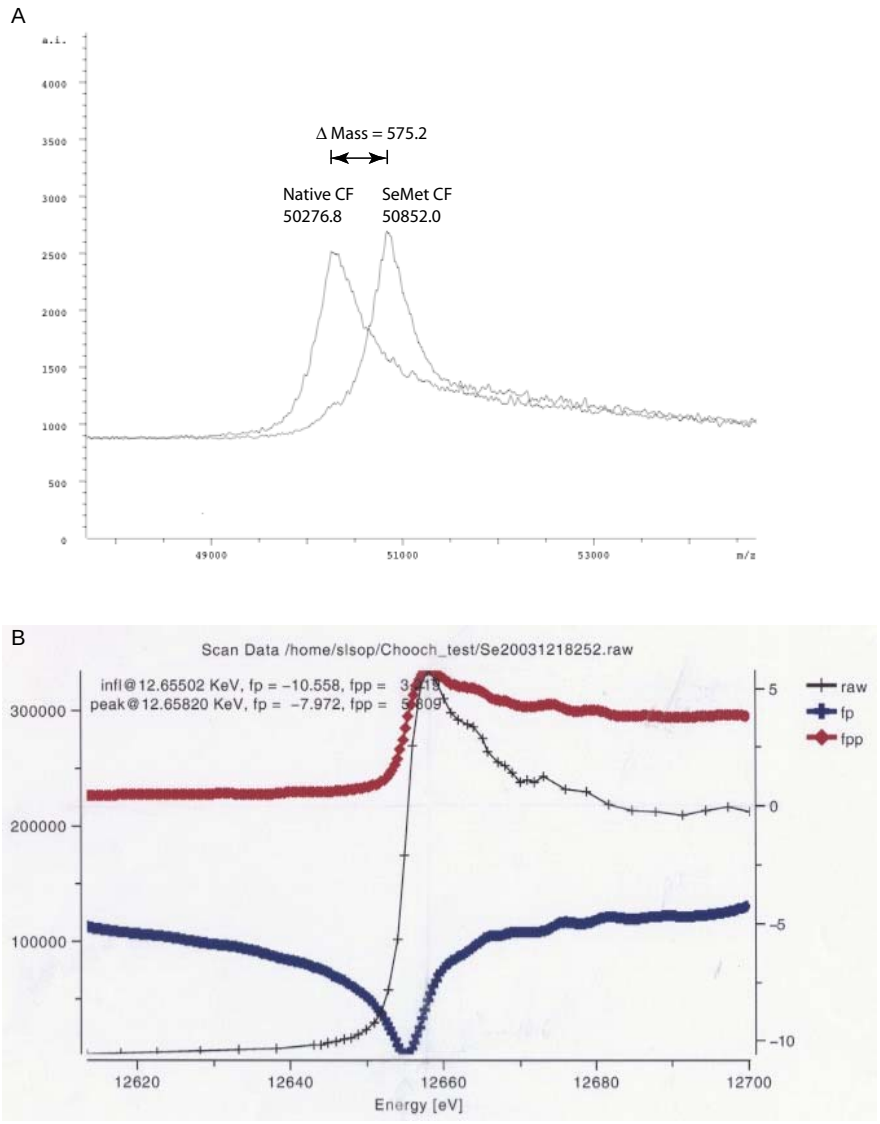
Verification of SeMet substitution in CF

The substitution of Met by SeMet in the CF protein was verified by mass spectrometry and absorption spectrometry. SeMet-substituted CF has a mass of 50852 Da as determined by mass spectrometry. It is 575 Da bigger than the native CF protein (Figure 3.14A). Considering the replacement of a sulphur atom (mass = 32.965) with a selenium atom (mass = 78.96) leads to a mass increase of 46.895 Da, this MW difference in the proteins accounts for SeMet substitution at 12.3 out of 13 sites per PleD monomer. Thus, the occupancy of SeMet is 95 %. The characteristic absorption edge of Se displayed by the SeMet CF crystal quantitatively verified this substitution in the protein (Figure 3.14B).

CF-product crystal

The type of constructs and ligands used in crystallisation, and crystal manipulation affected the quality of PleD crystals. The crystals that were

Figure 3.14: Verification of SeMet substitution in CF using mass spectrometry and absorption spectroscopy. A, Overlaid mass spectra of native and SeMet CF show a mass increase of 575.2 Da in the SeMet CF. This is equivalent to the replacement of 12.3 Met with SeMet. B, Absorption spectrum of the SeMet CF crystal showed the characteristic absorption edge of Se at 12.6582keV.



used in structure determination were the CF crystals co-crystallised with c-diGMP. These CF-(c-diGMP) crystals were cloverleaf-like when directly obtained from vapour diffusion. They belonged to the space group of $P4_22_12$ and diffracted to 3.5 Å. Microseeding was required to produce needle-like crystals that gave an improved diffraction to 2.7 Å (Figure 3.15A, C).

They belonged to the same space group. Assuming two molecules in the asymmetric unit, the Matthews coefficient V_m was determined to be 3.88 Å^3 / Da using equation 3.7, which lies within the common range of 1.66-4.0 Å^3 / Da for soluble proteins [42].

$$\begin{aligned} V_m &= \frac{\text{cell volume}}{\text{total weight of protein in unit cell}} \\ &= \frac{abc}{mnZ} \\ &= 3.88 \text{ Å}^3 \text{ / Da} \end{aligned} \tag{3.6}$$

where $a = b = 135.9$ Å, $c = 169.2$ Å; m is the molecular weight of CF, 50278 Da; n is the number of molecules in the asymmetric unit, 2 in this case; and Z is the number of asymmetric units in the unit cell, 8 for $P4_12_12$.

The solvent content was derived from the following equation:

$$\begin{aligned} \%\text{ solvent} &= 1 - \frac{1.23}{V_m} \\ &= 68.3\ \% \end{aligned}$$

SeMet CF protein was also crystallised in complex with c-diGMP. But when complexed with c-diGMP, crystals in the form of cloverleaf and needles appeared in the same crystallisation drop directly from vapour diffusion. Only the needle form was measured and it diffracted to 3.0-3.2 Å in the multiwavelength anomalous diffraction (MAD) experiment (Figure 3.15B). The SeMet CF-(c-diGMP) crystal belonged to the same tetragonal space group as the native crystal but with slightly smaller cell constants.

Other PleD crystals

Two other PleD crystal forms were also obtained. Co-crystallisation of CF with GMP-PNP formed cloverleaf-like crystals (Figure 3.16). These crystals belonged to the space group of $P4_12_12$ and had a very long cell constant along the c*-axis ($a = b = 86.3$ Å, $c = 295.8$ Å). There were two molecules in the asymmetric unit. The diffraction was only up to 6.6 Å. On the other

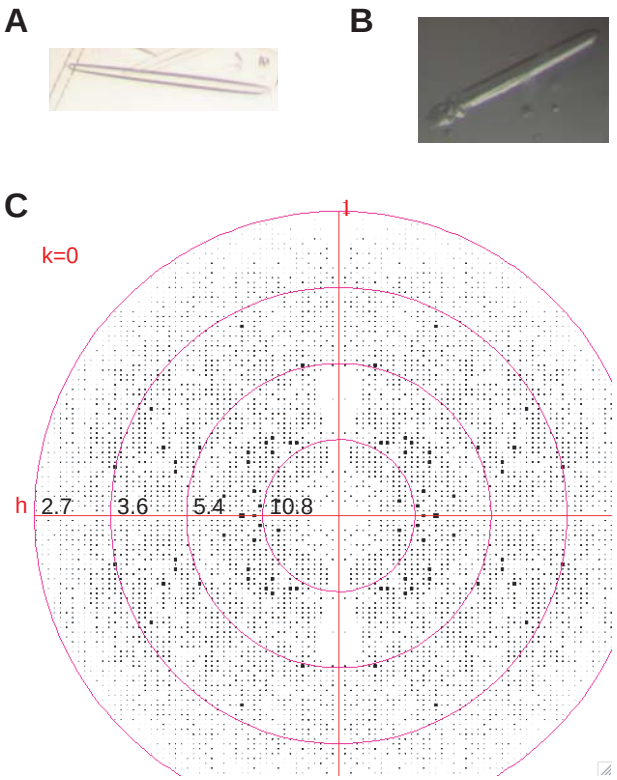


Figure 3.15: CF-c-diGMP crystals and diffraction. A, The native needle crystal of width 15 μm. B, SeMet CF-c-diGMP crystal in needle form. C, Integrated and scaled reflections from the native crystal. Reflections with high intensities are shown as big spots. The diffraction is isotropic and reaches 2.7 Å.

hand, co-crystallisation with GTPγS did not give crystals at all, and that co-crystallisation with PPi gave only salt crystals.

The other PleD crystal form was obtained from the apo-NF crystals. It is interesting that CF did not crystallise in its apo form. The NF crystals were bipyramidal (Figure 3.17). They belonged to the hexagonal space group of $P6_222$ with a unit cell of the dimensons $a = b = 94.2$ Å, $c = 187.4$ Å. They diffracted to 4.5 Å. In contrast to the CF-ligand crystals, the apo-NF crystal form contained only 1 molecule per asymmetric unit.

Data collection

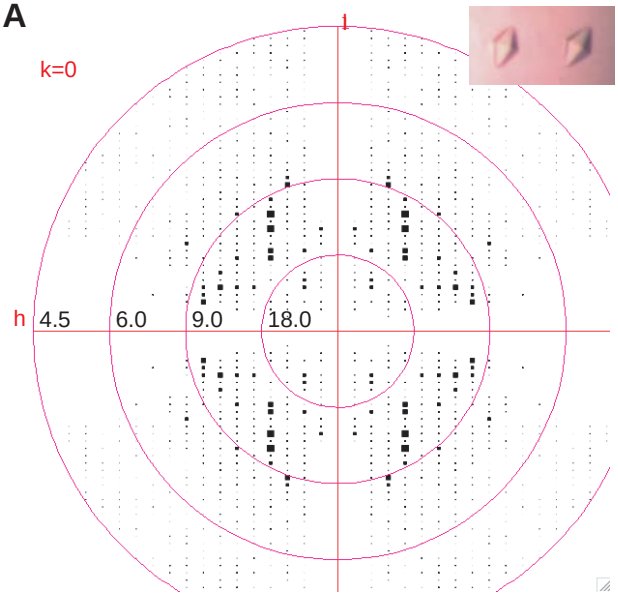
The method of anomalous scattering exploits the anomalous difference in Friedel-related reflection intensities when heavy atoms are excited close to their absorption edge. The sites of the heavy atoms can be located and the phases for the structure factors calculated [60]. We have used the MAD method in which multiple datasets were collected at several wavelengths to maximise the anomalous signal. These datasets were collected from the same crystal to avoid non-isomorphism.

A native dataset from a native CF crystal and three MAD datasets from a SeMet-substituted CF crystal were collected at the synchrotron facility at Swiss Light Source in Villigen, Switzerland. Cryo-conditions were used to prevent ordered ice formation and to reduce the radiation damage to crystals, which is a major concern when using the synchrotron source. Flash freezing was achieved with CF crystals pre-soaked in a mother liquor that was cryoprotected with ethylene glycol. This resulted in a clear drop in the mounting loop and diffraction patterns that were free of ice rings. A complete dataset was collected from a small, single native crystal of 0.015 mm by width by shifting the X-ray beam along the length of the crystal. Likewise, the three MAD datasets were all collected from a single SeMet-substituted crystal by scanning along the length of the crystal.

Overall scheme in structural elucidation of CF

The overall scheme in determining the CF structure is represented as a flowchart in Figure 3.18. Significant results at intermediate steps of structure determination are stated in the corresponding stage in the flowchart for quick reference. Statistics obtained from data collection, phasing and refinement are summarised in Figure 3.19.

A



B

Data collection	
Wavelength (Å)	0.91840
Resolution (Å)	4.5
Highest shell (Å)	4.74 – 4.5
Space group	$P6_222$
a = b (Å)	94.2
c (Å)	187.4
Unique reflections	2864
Average redundancy[a]	4.1 (4.2)
I / σ [a]	8.4 (4.5)
Completeness[a] (%)	89.5 (89.2)
R_{merge} [a,b] (%)	6.3

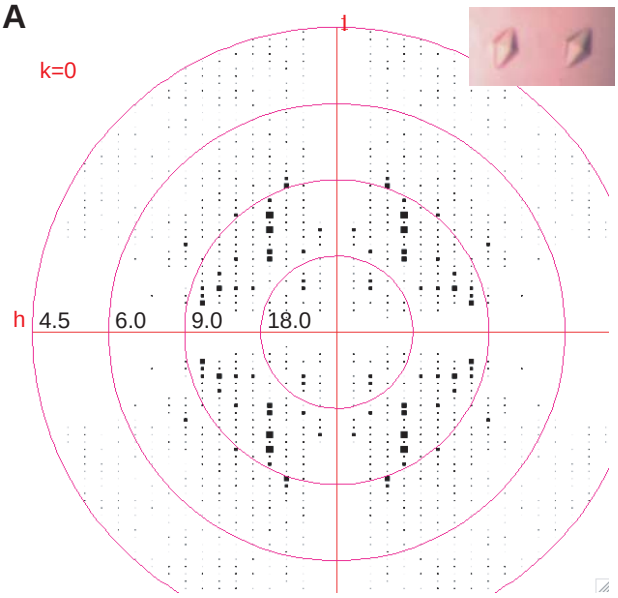
Packing	
Vm	2.39
Solvent content (%)	48.5
No of molecules in AU	1

[a]The numbers in parentheses refer to the highest shell.

[b]$R_{merge} = \Sigma| I - \langle I \rangle | / \Sigma I$, where I is the scaled intensity of a given reflection.

Figure 3.16: Diffraction and statistics of CF-GMP-PMP crystals. A, Integrated and scaled reflections of the cloverleaf-like crystal (top right) to 6.6 Å. B, Statistics of data collection.

A



B

Data collection	
Wavelength (Å)	0.91840
Resolution (Å)	4.5
Highest shell (Å)	4.74 – 4.5
Space group	$P6_222$
a = b (Å)	94.2
c (Å)	187.4
Unique reflections	2864
Average redundancy[a]	4.1 (4.2)
I / σ [a]	8.4 (4.5)
Completeness[a] (%)	89.5 (89.2)
R_{merge} [a,b] (%)	6.3

Packing	
Vm	2.39
Solvent content (%)	48.5
No of molecules in AU	1

[a]The numbers in parentheses refer to the highest shell.

[b]$R_{merge} = \Sigma| I - \langle I \rangle | / \Sigma I$, where I is the scaled intensity of a given reflection.

Figure 3.17: Diffraction and statistics of the apo-NF crystals. A, Integrated and scaled reflections of the bypyramid crystal (top right) to 4.5 Å. B, Statistics of data collection.

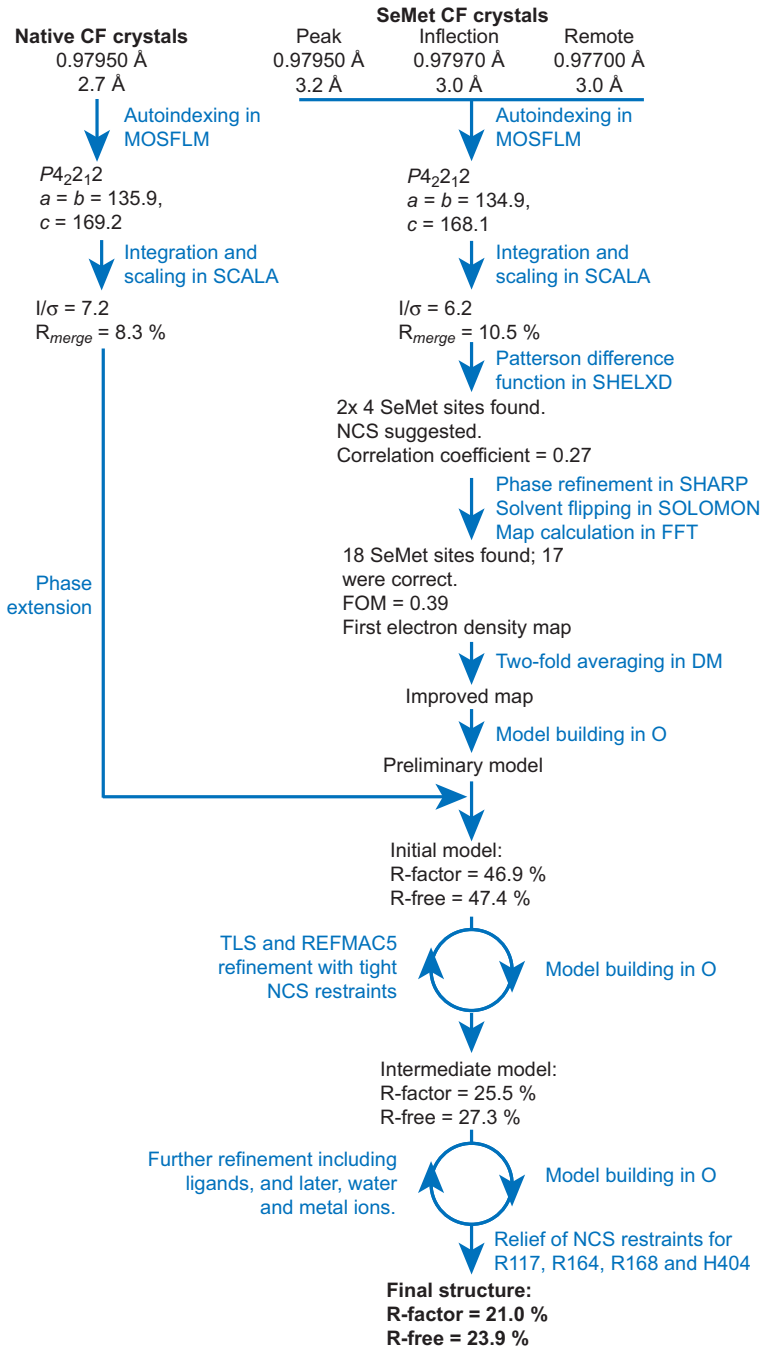


Figure 3.18: Scheme detailing the procedures in structure determination of PleD.

Figure 3.19: Crystallographic data of CF crystal, and statistics of phasing and refinement.

Experimental data				
Data collection	**Native**	**SeMet**		
		Peak	**Inflection**	**Remote**
Wavelength (Å)	0.97950	0.97950	0.97970	0.97700
Resolution (Å)	2.7	3.2	3.0	3.0
Highest shell	2.85 – 2.70	3.27 – 3.20	3.16 – 3.00	3.16 – 3.00
Space group	$P4_22_21$			
$a = b$ (Å)	135.9	134.9		
c (Å)	169.2	168.1		
Unique reflections	42707	26249	32080	32140
Average redundancy[a]	5.7 (5.5)	4.1 (3.9)	4.8 (4.9)	4.6 (3.6)
I / σ [a]	7.2 (1.9)	6.2 (2.1)	7.2 (1.8)	7.3 (1.1)
Completeness[a] (%)	99.3 (98.7)	99.9 (100)	99.9 (100)	99.6 (97.6)
Anom complete (%)	-	99.7	99.9	97.5
R_{merge} [a,b] (%)	8.3 (36)	10.5 (33)	9.4 (39)	11.0 (63)

Phasing								
Resolution shell	13.1	7.69	5.98	5.06	4.47	3.72	3.46	3.16
	2							
Figure of merit	0.74	0.67	0.62	0.55	0.50	0.34	0.27	0.13
Mean figure of merit	0.39							

Refinement			
No of molecules in AU	2	Protein atoms	6883
Resolution (Å)	50.0 – 2.7	Water molecules	15
R^c / R_{free}[d] (%)	21.0 / 23.9	Ligand atoms	233
R.m.s.d.[e]		Residues in	92.8
bond lengths[e] (Å)	0.008	Ramachandran core (%)	
bond angles[e] (°)	1.2		
R.m.s.d.[e] ΔB bonded atoms (Å^2)		Residues in	0.0
Main chain	1.13	disallowed region (%)	
Side chain	2.42	Average B (Å^2)	25.6

[a]The numbers in parentheses refer to the highest shell.
[b]$R_{merge} = \Sigma| I - \langle I \rangle | / \Sigma I$, where I is the scaled intensity of a given reflection.
[c]$R = \Sigma| F_o - \langle F_c \rangle | / \Sigma F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes for reflection.
[d]R_{free} refers to the R value of 5 % of randomly selected reflections excluded from the refinement.
[e]r.m.s.d. from ideal stereochemistry

Phasing

Considering 2 molecules per asymmetric unit ($V_m = 3.88$) and 12 out of 13 methionines labelled per monomer, there are 24 SeMet sites in the asymmetric unit. Using the programme SHELXD [55], 25 Patterson peaks were searched from the peak dataset at 4.5 Å. The best run with correlation coefficients at 28.2 (all) and 11.2 (weak) and a figure of merit (FOM) at 0.64 gave 18 Patterson peaks at > 29 % intensity. Among these peaks, two clusters, each consisting of 4 peaks, were observed to be related by NCS two-fold symmetry (Figure 3.20).

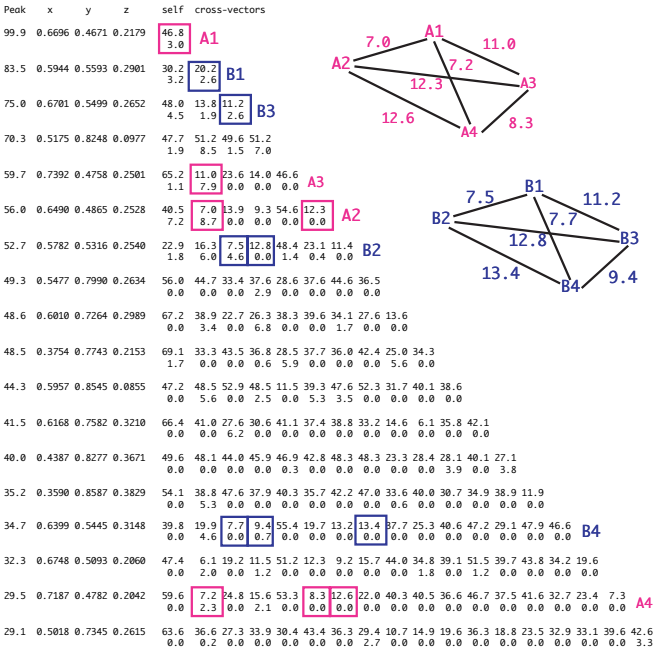


Figure 3.20: Crossword table output from a SHELXD run and assignment of sites to NCS-2-fold related clusters from the peak dataset. The Patterson peaks are sorted in decreasing order of intensity (first column), and each peak is annotated with fractional coordinates x,y,z (second to fourth columns). The self-vectors and cross-vectors relating different Patterson peaks are annotated with distance (in Å) at the top and Patterson minimum function at the bottom. Cluster A, consisting of peaks A1-A4, was assigned to be related to cluster B, consisting of peaks B1-B4, by a local two-fold symmetry.

These eight sites were used as input in a MAD run at 3.2 Å in the programme SHARP [11] to search for other SeMet sites. The newly found

sites that coincided with the hits from the SHELXD run were incorporated as input into new rounds of searches. The position, occupancy and thermal parameters of the SeMet sites were then refined to give a final mean FOM at 0.39. 18 out of 24 SeMet sites in PleD dimer were found of which 17 were correct (Figure 3.21). Solvent flipping was performed at 60 % with programme SOLOMON and the first electron density map was calculated by fast Fourier Transform with the programme FFT. Figure 3.22 panel A shows the superimposition of this electron density map onto the Cα trace of the final structure. The central β-sheet was already clearly visible at this stage and there were even some indications of sidechains albeit serious ambiguities at the helix and loop region.

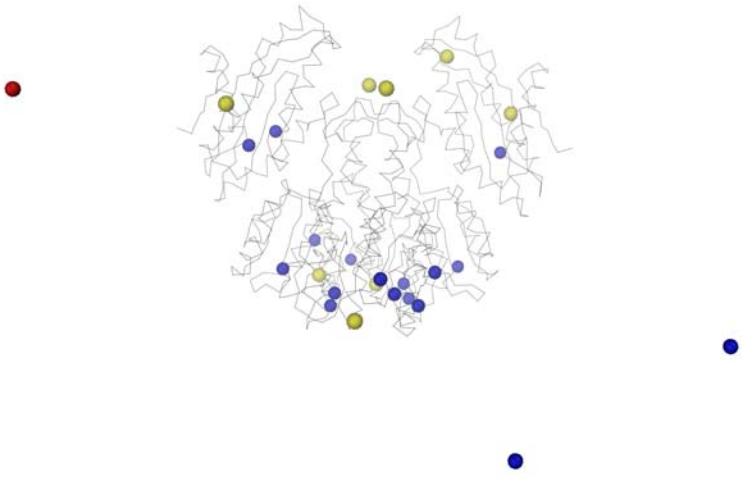


Figure 3.21: SeMet sites found in SHELXD and SHARP. The correct sites are highlighted in blue and the wrong site in red. The unfound sites, including the two unlabelled ones, are shown in yellow. The Cα trace of CF is also shown. The two sites in blue that are away from the Cα trace are in a different asymmetric unit but are crystallographic symmetry related to the asymmetric unit shown.

Density modification with two-fold averaging

The electron density map was improved by NCS averaging [66]. The operator describing the NCS symmetry that related the two clusters of heavy atoms mentioned above was calculated in the programme MODTRAFO. It obeyed a two-fold symmetry of polar angles (85.7, 274.4, 180.0) with a screw

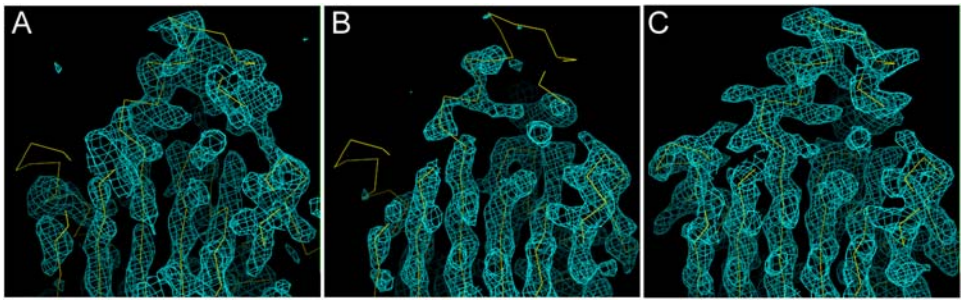


Figure 3.22: Comparison of the electron density maps calculated from different stages in structure elucidation showed improvements in terms of completeness and clarity. A, Map before two-fold NCS averaging in DM. B, Map after NCS averaging; C, the final map. Cα trace of the final structure is shown in yellow.

component of -0.03 $\mathring{A}$. Using the NCSMASK option in the programme DM [8], an NCS averaging mask comprising the two clusters was computed at a solvent content of 63 %. NCS operations were applied to this mask and the densities of the two molecules were averaged. The modified map was back-transformed to calculate a new set of phases that were used to generate a new map. These density modification procedure was carried out in cycles until convergence with phase extension to the diffraction limit of 2.7 Å by merging with the amplitudes of the observed structure factor from the native dataset. Further cycles of electron density modification yielded a refined NCS symmetry axis of polar angles (85.9, 273.7, 179.9) with a screw component of -0.16 $\mathring{A}$, and an intermediate map (Figure 3.22B), which showed an improvement in the signal-to-noise ratio especially in the helix and loop region where ambiguities were resolved. More sidechains were also identified.

Model building

Using the BONES programme in the programme MAPMAN [27, 30], the electron density was skeletonised and then visualised in the programme O. Due to the predicted topological homology of the D1 domain, and possibly of the D2 domain, to the CheY protein [35, 58], the Cα trace of CheY was used to fit the bones to trace the polypeptide chain. It was possible to fit two Cα traces of CheY to the bones proving that D2 domain has the same fold as CheY. Using the SeMet sites as sequence markers, the orientation of D1 and D2 domains was determined as a result. On the other hand, the building of the DGC domain was achieved by manual rectification. After the bones were

edited, the positions of the Cα atoms were assigned by placing consecutive Cα atoms at a distance of around 3.8 Å along the skeleton. The mainchain was then placed on the Cα trace followed by addition of sidechains using rotameric conformations that fitted the electron density. Subsequently, the built polypeptide was regularised to restore the proper bond lengths, bond angles and dihedral angles. The first manually built model gave an R-factor of 46.9 % and R_{free} of 47.4 %.

Refinement

Successive rounds of TLS / REFMAC5 refinement [44] with strict NCS restraints and model building lowered the R-factor to 25.5 % and R_{free} to 27.3 %. As the electron density map improved, we could find three clusters of electron density that did not belong to the protein. They were all in the shape of a horse-shoe (Figure 3.23), to which c-diGMP could be fitted. One of them was located on the DGC domain while the other two were intercalated and were situated between the D2 and DGC domains.

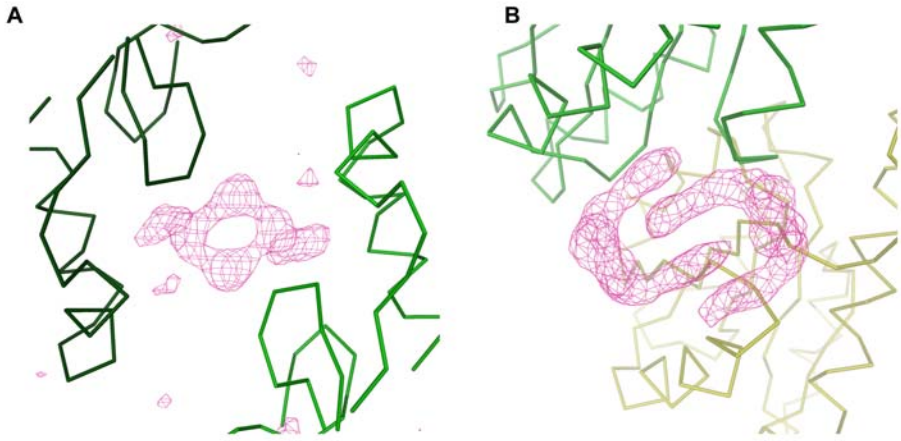


Figure 3.23: The F_o-F_c map, contoured at 4 σ and shown in magenta, shows electron density that does not belong to the protein in two regions of the asymmetric unit. A, Density in the shape of a horse-shoe is located between two crystallographic symmetry related molecules (green and dark green) at the border of an asymmetric unit. B, Two of these horse-shoe shaped electron density masses are located at the domain interface between D2 (yellow) and DGC (green).

In order to include a ligand molecule in structural refinement, its internal

geometry has to be specified for REFMAC5. This is often described in the monomer library in CCP4 [44] if the ligand occurs commonly. Due to the absence of a c-diGMP molecule entry in both the monomer library and the Protein Data Bank (PDB) [3], a library file for c-diGMP was created using the internal geometry of a GMP molecule described in a monomer library file. In addition to the specified geometry requirements, connectivity was built between the ribosyl O3* and the Pα of the second GMP molecule while restraining their interatomic distance to 1.59 Å, the usual bond length. The new bond angles at Pα between the ribosyl O3* and the non-ester oxygens O1A and O2A were set to the usual bond angle of 109°.

The two GMP molecules were then set in an anti-parallel fashion and fit into the electron density map. Care was taken to place the ribosyl 3' oxygen of one molecule to the 5' phosphorus of the other molecule close to the bond distance and bond angle to facilitate proper stereochemistry in the refinement process.

Another issue was to address was the packing in this protein crystal. Two c-diGMP molecules are situated in the strategic locations that crosslink two crystallographically related CF dimers. In other words, there is half a c-diGMP molecule per PleD monomer. However, refinement of the PleD monomers with two halves of c-diGMP in an asymmetric unit would lead to individual refinement of the two GMP molecules, with the daunting result that the proper stereochemistry of an intact c-diGMP molecule at this crystal contact might be lost.

This problem was circumvented by re-defining the asymmetric unit. Instead of assigning two halves of the c-diGMP molecule to the dimer, one intact c-diGMP molecule was assigned to a monomer chain. The consequence was that the two GMP molecules were subjected to the restraint measures required to bring them to the right stereochemistry of a c-diGMP molecule.

In the subsequent rounds of refinement, we searched for peaks in the F_o-F_c map that had a σ-value above 3. We assigned them to 15 water molecules and 3 metal ions including two Mg^{2+} ions and one Zn^{2+} ion. The two Mg^{2+} ions are located in the acidic clusters on chain A and B, whereas the Zn^{2+} ion is located at the end of the disordered His-tag on chain A (Figure 3.24).

As the R-factors finally converged, the electron density map revealed that four sidechains did not adopt the same conformation in the NCS dimer. They were R117, R164, R168 and H404. R117 of chain A might pack against F113 whereas its counterpart on chain B does not interact with any residues. R164 and R168 are both involved in non-isologous intra-layer crystal contacts and will be explained later. H404 is located very close to the disordered

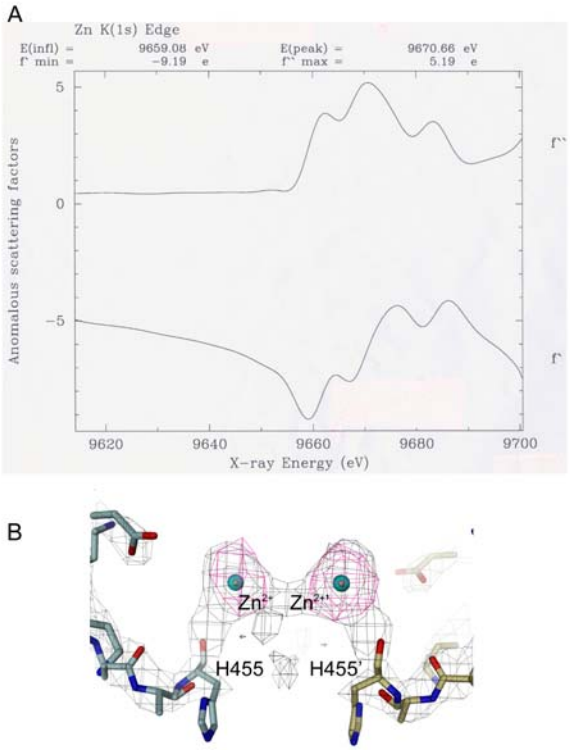


Figure 3.24: Evidence of Zn^{2+} ion in CF crystal. A, Absorption spectrum of CF crystal shows characteristic absorption edge of Zn^{2+} at 9.6586 keV. B, The map (pink) showing the anomalous difference of zinc is contoured at 2.5 σ and shows a Zn^{2+} ion (cyan) bound to the disordered C-terminus at H455 (carbons in grey). The electron density on this sidechain is absent in the $2F_o$-F_c map (black) of the native dataset. The last five histidines of the His-tag are disordered as reflected by the proximal arrangement of the C-terminal end of another molecule (carbons in khaki).

D1-D2 domain linker on the same chain of the crosslinked dimer. It might be due to the disorder in this region that caused the difference in the conformation of H404 on the two chains. NCS restraints were only relieved for these sidechains.

The final R-factor is 21.0 % and R_{free} is 23.9 %, and the electron density map is shown in panel C of Figure 3.22. In comparison to the initial and intermediate maps (Figure 3.22A, B), the final map is complete including a whole loop which did not show electron density before. It also shows details of most sidechains.

Evaluation of structure quality

The stereochemical quality of the crystal structure was continuously evaluated throughout the refinement steps using the programme PROCHECK [32]. In the final structure, 92.8 % residues are found in the most favourable region on the Ramachandran plot (Figure 3.25) and none in the disallowed region. The mainchains and sidechains obey good stereochemistry (Figure 3.26 and 3.27). The structure was deposited in PDB under the accession code 1W25.

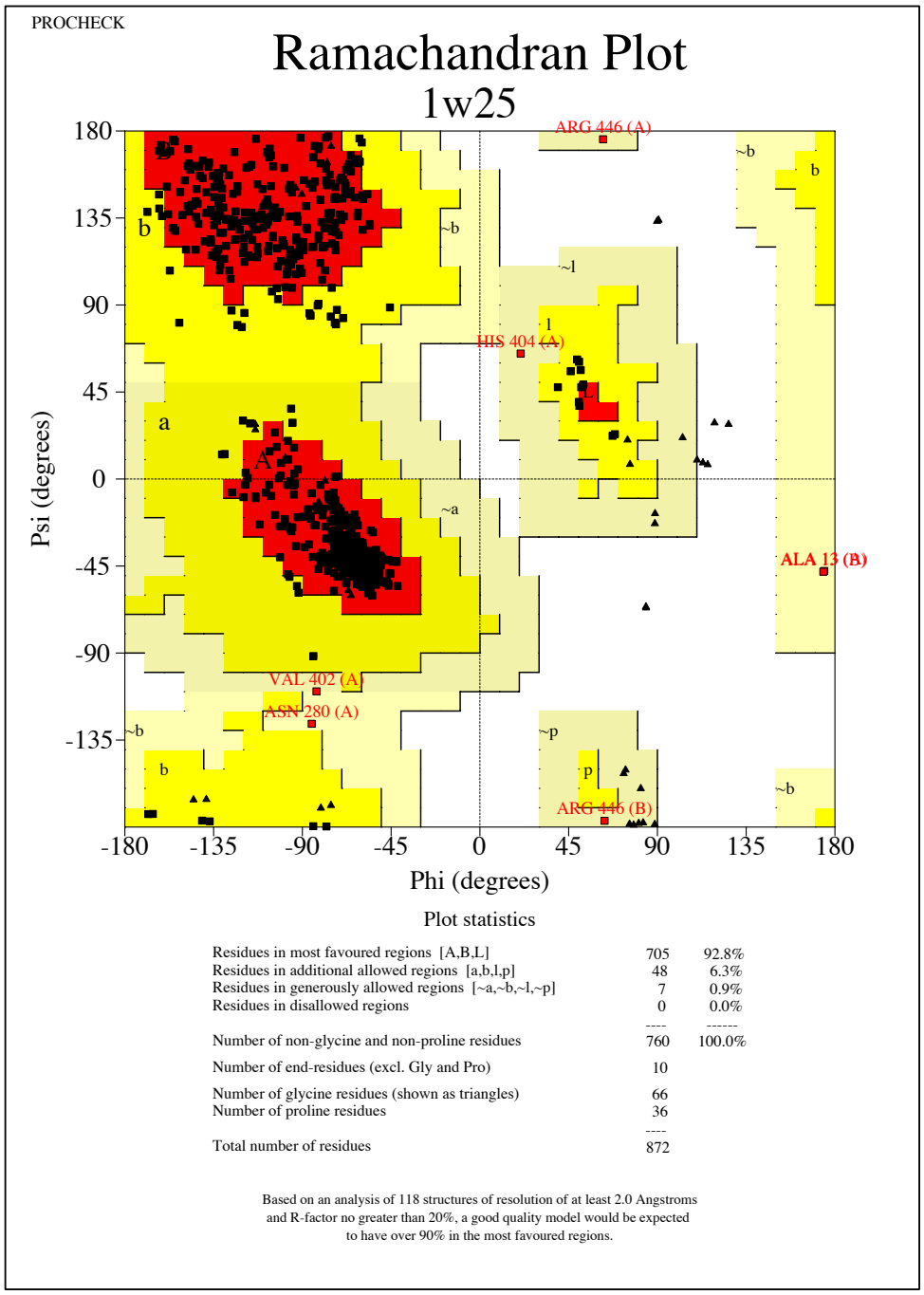


Figure 3.25: Ramachandran plot of the final CF structure.

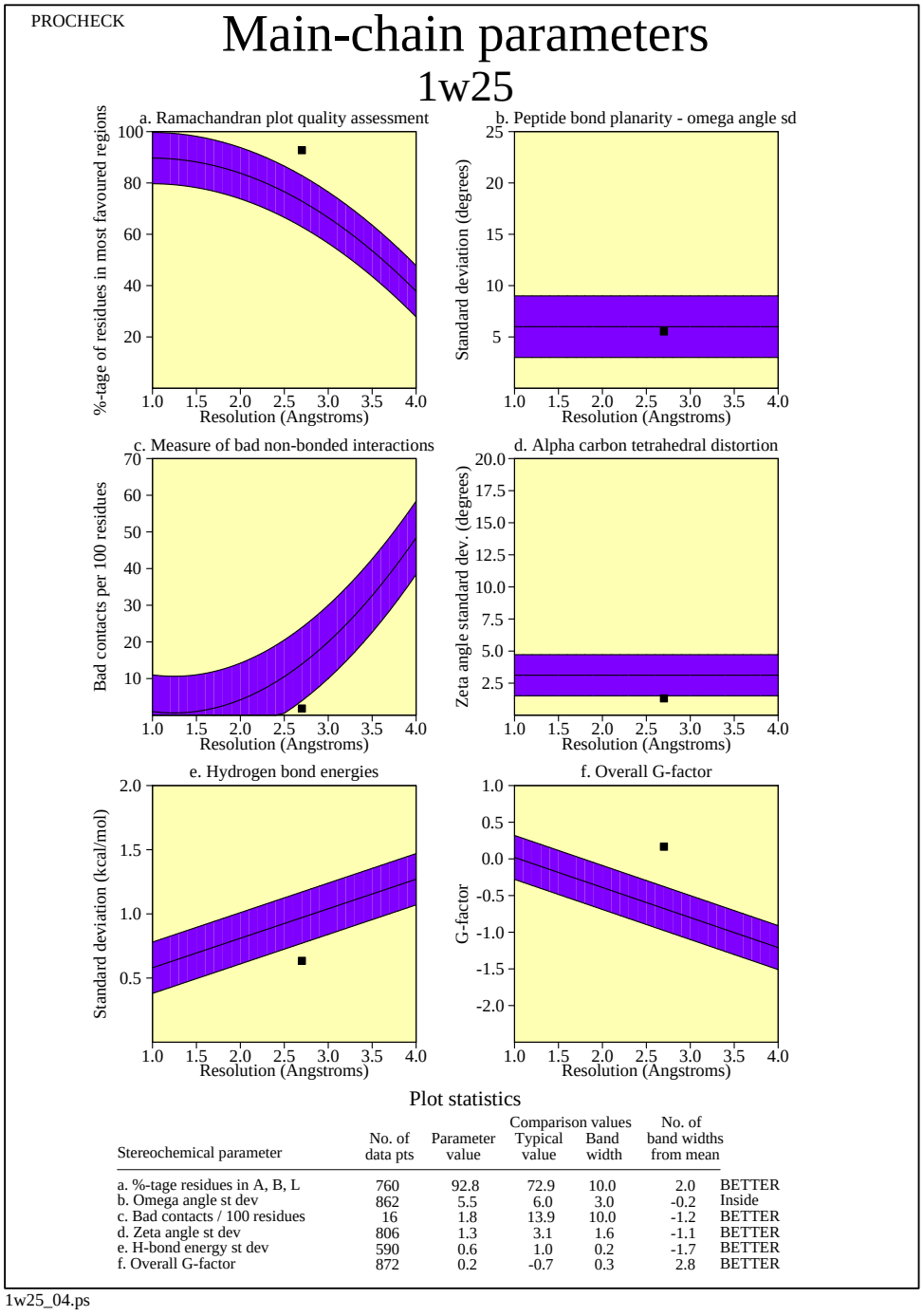


Figure 3.26: PROCHECK assessment of the overall geometry of the mainchains in CF.

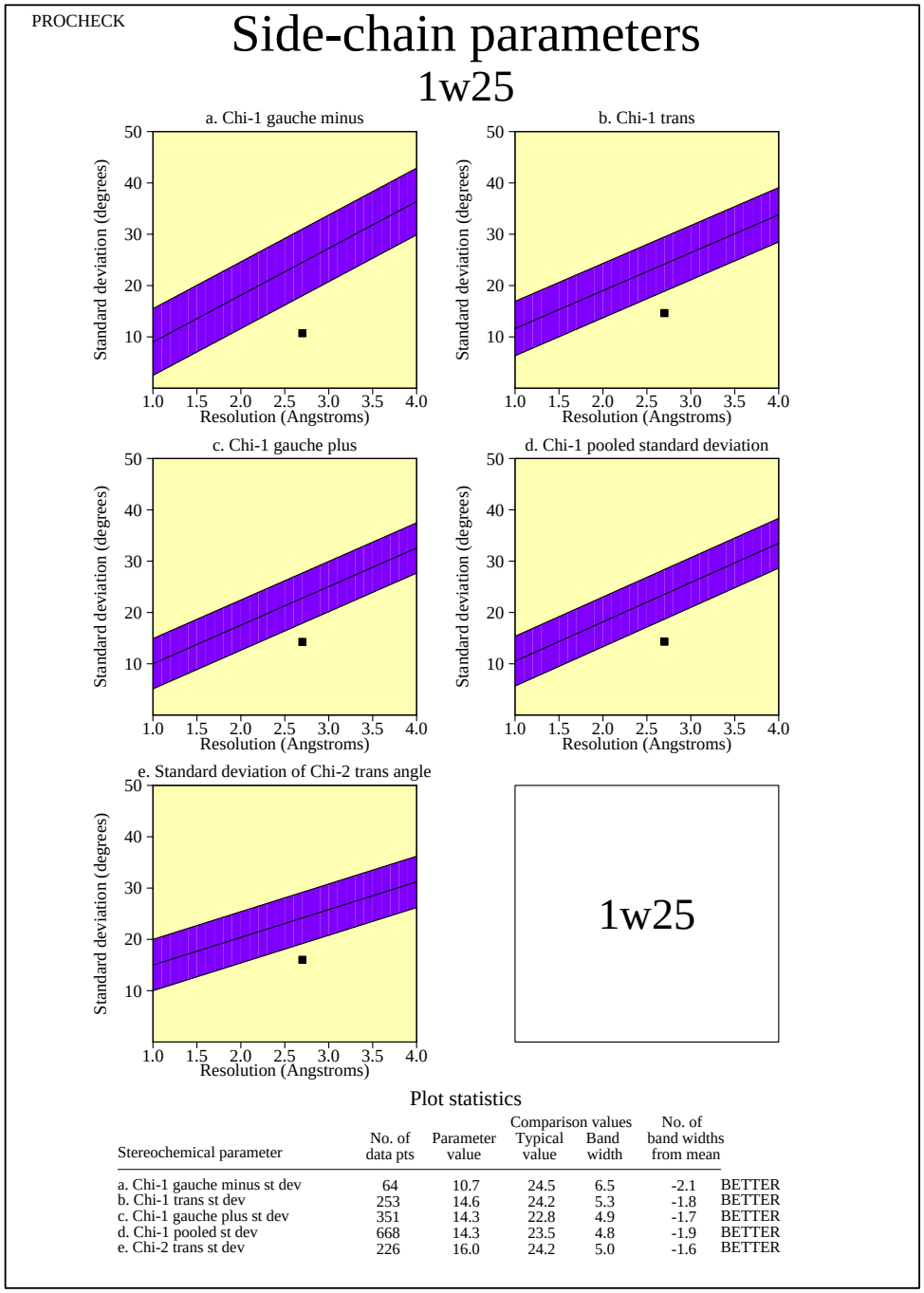


Figure 3.27: PROCHECK assessment of the overall geometry of the sidechains in CF.

Structural analysis of PleD

PleD architecture

The final electron density map was of good quality and allowed the building of most of the PleD molecule. PleD comprises three domains, D1 (aa 2-136), D2 (aa 147-281) and DGC (aa 289-455) (Figure 3.28). D1 and D2 domains have the same $(\beta/\alpha)_5$ topology as CheY, whereas DGC shares a similar fold as the AC catalytic domain that consists of a mixed topology $\beta\alpha\alpha\beta\beta\alpha\beta$ (Figure 3.29). The domain linkers (aa 137-146 and 288-454) and the C-terminal His-tag (aa 456-460) show poor electron density quality and could not be modelled.

Secondary structures were assigned using the programme DSSP [28, 32] with additional manual rectification. In particular, we have made the extra assignment of strands β0 and β0' in DGC domain after observing an extension of the hydrogen bond pattern of the central anti-parallel β-sheet (inset of Figure 3.29). The final assignment is shown in Figure 3.30. In each monomer, there are 15 α-helices, 8 β-strands, and 2 short 3_{10} helices comprising residues 221-223 and 403-405. In addition, two cis-prolines, P106 and P255, are observed.

Crystal packing

The unit cell of CF crystal consists of 8 asymmetric units. Each of the asymmetric units contains a dimer of NCS-related CF molecules, the interactions of which will be discussed in more detail on page 77. The asymmetric units are arranged in layers that are almost parallel to the xy-plane (Figure 3.31). These layers stack on top of each other along the z-direction to build up the 3D lattice. Crystal packing is therefore provided by contacts within the layer, i.e. intra-layer, and across the layers, i.e. inter-layer.

The intra-layer contacts involve only protein residues (Figure 3.32). They include contacts 1 and 2 which are responsible for lattice expansion in the y- and x-directions respectively. The contacts are held by non-isologous D1-D1 and D2-D2 interactions from chain A of one dimer, and chain B of the neighbouring dimer.

Contact 1 involves the α1-α1 helix-helix interactions of the D2 domains. Whereas Arg168 of chain B forms a hydrogen bond with the mainchain carbonyl of Ala171 on chain A, its counterpart in chain A forms hydrogen bonds to the mainchains of Gln167 and Ala170 on chain B. In addition,

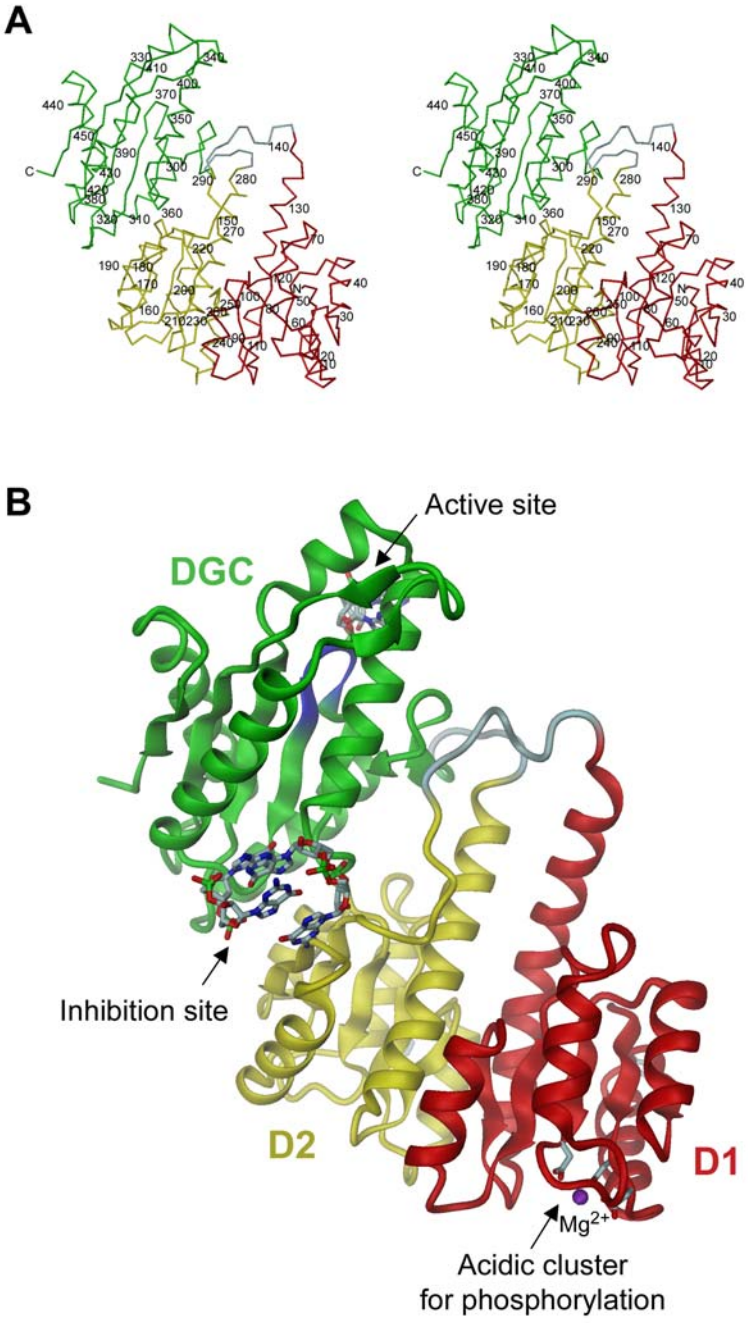


Figure 3.28: PleD monomer structure. A, The Cα-trace of PleD monomer, numbered at every ten residues, shows the modular arrangement of three domains, D1 (red), D2 (yellow), and DGC (green). The domain linkers (grey) are disordered and are modelled here. The last five residues of the His-tag, H456-H460, are disordered and not shown. B, Ribbon representation of the monomer. D1 domain carries the phosphoacceptor (in sticks), the acidic cluster (in sticks) and the Mg^{2+} ion (purple) required for phosphorylation. D2 domain does not have a phosphorylation site. DGC domain has a conserved β-hairpin (blue) that bears the conserved GGEEF sequence motif. PleD is bound with c-diGMP molecules at the catalytic active site, A-site, located on the conserved β-hairpin, and the product inhibition site, I-site, located at the domain interface of D2 and DGC.

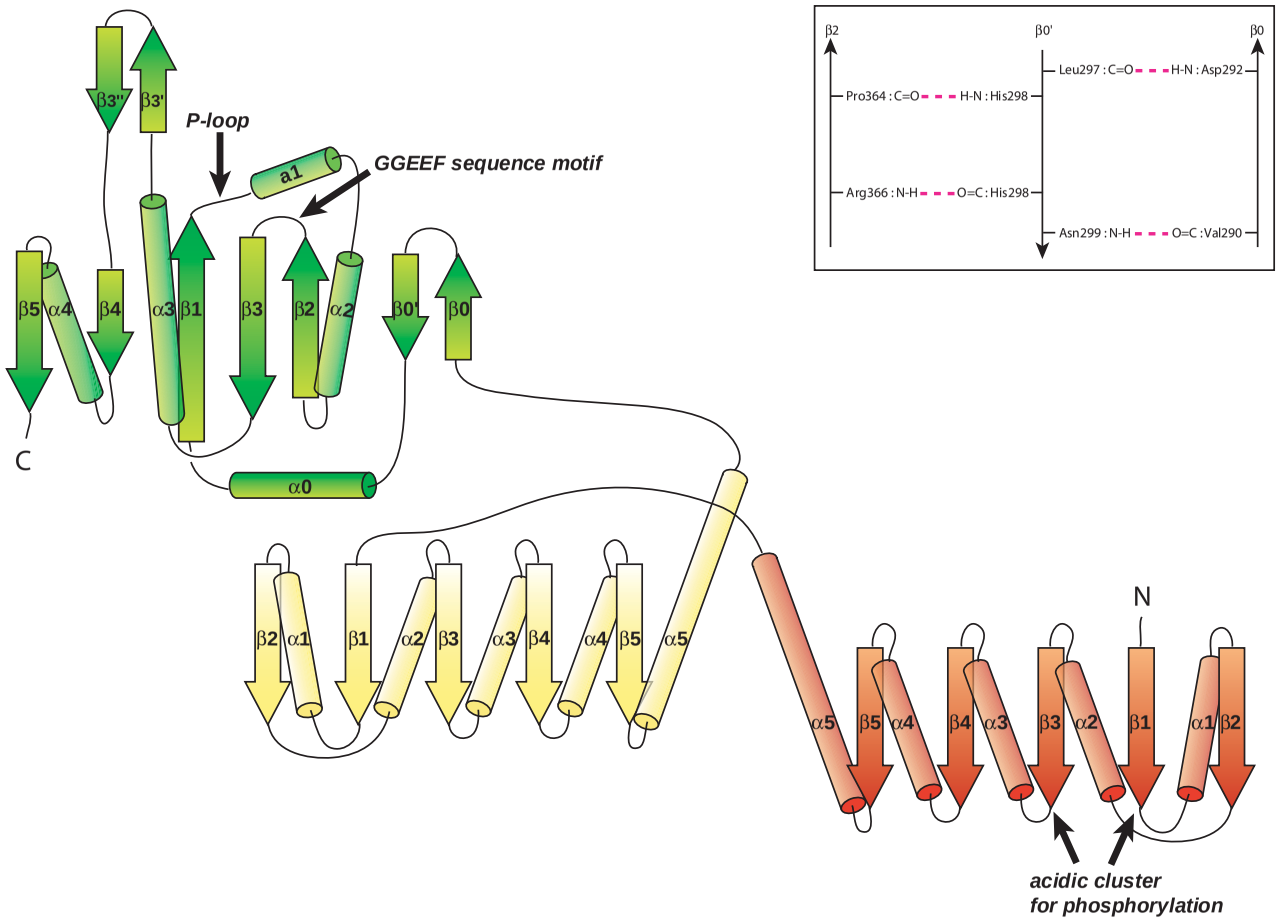
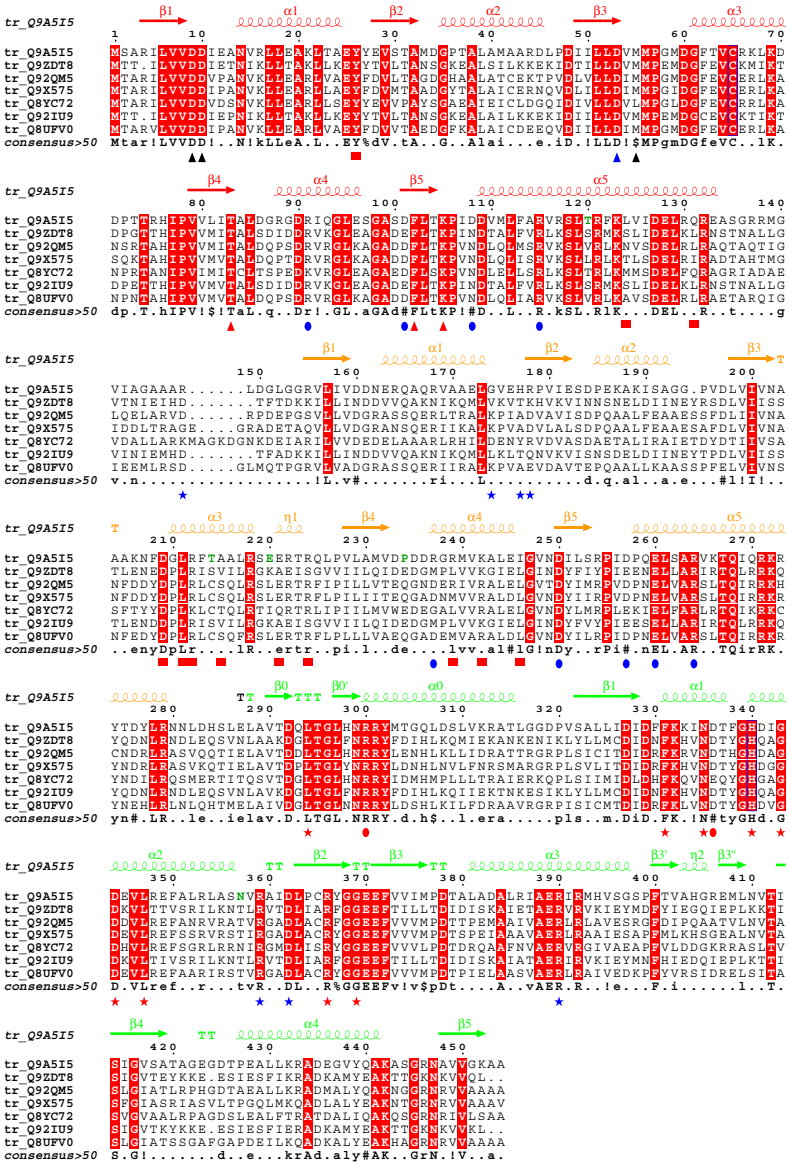


Figure 3.29: Topological diagram of PleD coloured according to individual domains. Arrows denote β-strands and cylinders denote α-helices. D1 (red) and D2 (yellow) domains share the same $(\beta/\alpha)_5$ fold. DGC (green) domain consists of a central mixed β-sheet of seven β-strands in the order 0,0',2,3,1,4,5 that is flanked by five α-helices on both sides. It also has a small sheet of the anti-parallel strands 3' and 3''. The secondary structures were determined using the programme DSSP [28, 32]. The extra assignment of strands β0 and β0' in DGC domain was due to the extension of the central anti-parallel β-sheet via β2 (inset). The two short 3_{10} helices comprising residues 221-223 and 403-405 are not shown here. Special sequential, structural and functional features are labelled in italics. N- and C-termini of the peptide are labelled accordingly.

Figure 3.30: Multiple sequence alignment of PleD homologues containing the modular arrangement of two CheY domains and a DGC domain. PleD sequence (top) is shown with six homologues found by the programme METAMOTIF from the EMBnet server (www.ch.embnet.org). The secondary structure of PleD is shown according to the domain colour scheme above the alignment. Arrows represent β-strands, coils represent α-helices and the letter 'T' represents turns. Strictly conserved residues are shown on a red background. Functional residues are indicated by symbols below the alignment: black triangle, Mg^2-binding in acidic pocket of phosphorylation; blue triangle, phosphoacceptor; red triangle, residues expected to undergo big conformational change led by phosphorylation; red star, nucleotide-binding at A-site; red circle, intra-domain salt bridge that might ensure formation of complete two-fold symmetric active site; blue star, nucleotide-binding at S-site; red square, dimer interface; blue circle, ionic pairs at D1-D2 domain interface; positions that are mutated in the constitutively active PleD are highlighted in green. The consensus sequence is calculated using the Blossum 62 algorithm and is shown below the alignment. The figure was produced by ESPript [22].



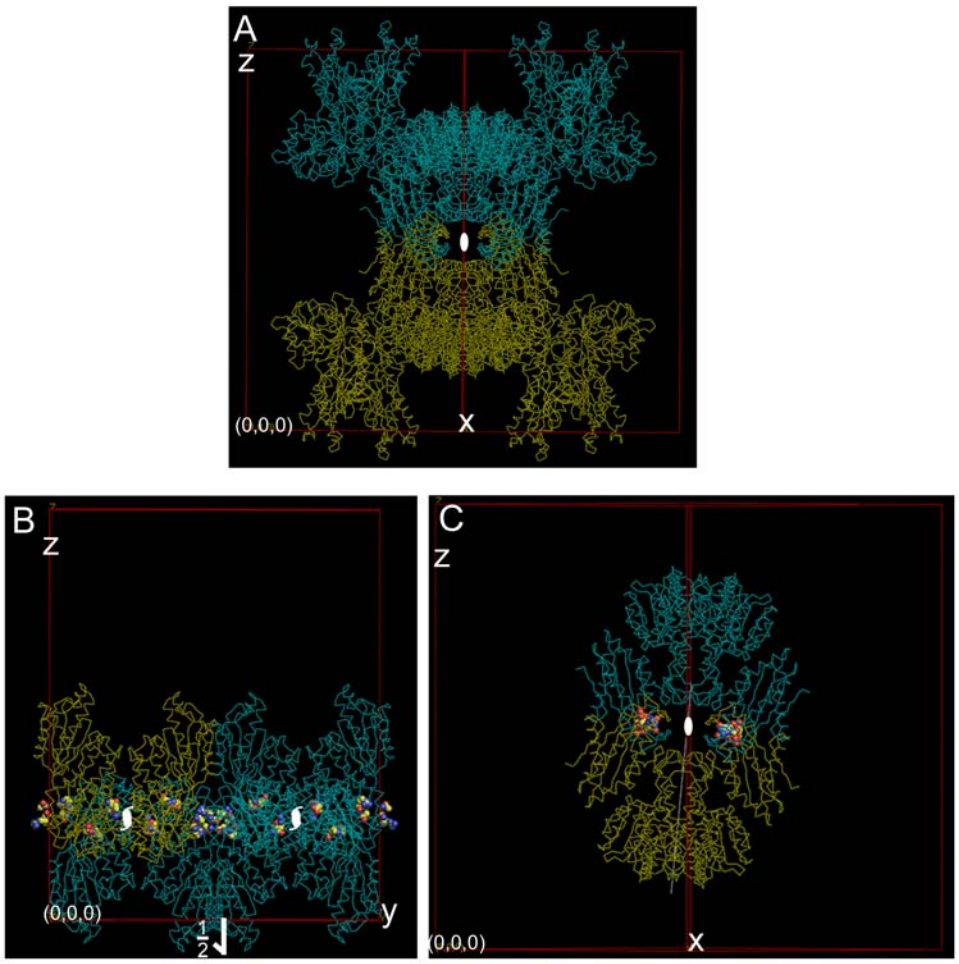


Figure 3.31: Packing of the CF crystal. A, The unit cell of CF crystal viewed along the xy-diagonal direction. It consists of 8 dimers which can be grouped into two layers (cyan and yellow) related by the crystallographic two-fold axis. B, Intra-layer packing as viewed along the x-direction. The crystal contacts are highlighted in cpk. Our reference molecule is in yellow. C, Inter-layer packing in CF crystal. Two molecules, each from a crystal layer in panel A, are shown here. The NCS axis that relates the two monomers in our reference molecule is shown as a white line. It intersects at right angle with the crystallographic two-fold axis that goes along the xy-direction. Crystal contacts are highlighted in cpk.

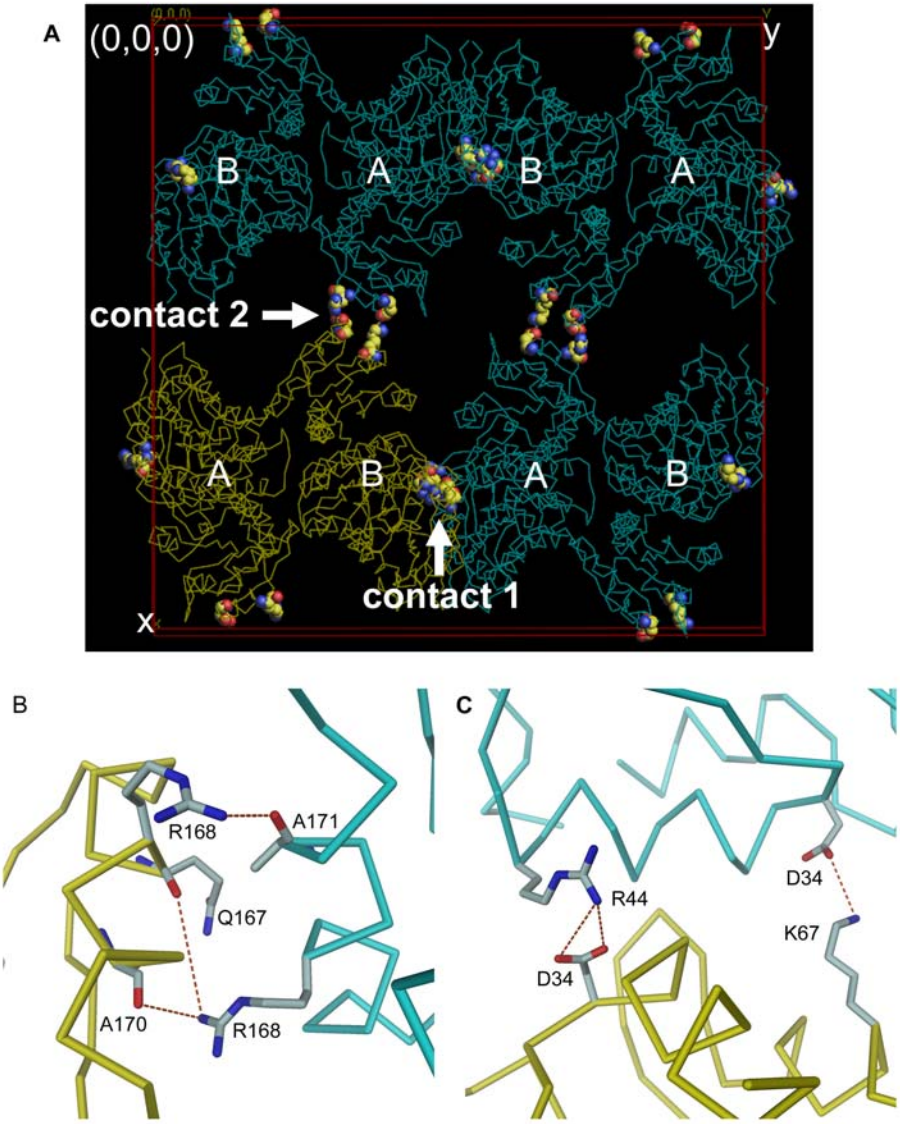


Figure 3.32: Intra-layer packing in CF crystal. A, Topview of the xy-plane shows the two types of crystal contacts within the crystal layer. Contact 1 is responsible for the expansion of the crystal in the y-direction and contact 2 in the x-direction. Residues involved in the contacts are highlighted in cpk as before. The reference dimer molecule is shown in yellow. Chains A and B of all four dimers on this plane are labelled. B, Enlarged view of crystal contact 1 shows non-isologous interactions between the two asymmetric units. Arg168 from chain B (yellow) makes a hydrogen bond with Ala171 from chain A (cyan) whereas its counterpart in chain A interacts with two other residues. C, Non-isologous interactions at crystal contact 2. Asp34 from each chain interacts with different residues from the other chain.

Arg164 of chain B forms a hydrogen bond with c-diGMP of chain A while its counterpart on chain A does not make contacts of any sorts. These non-isologous interactions are also found in Contact 2 that involves D1-D1 domain interactions. Asp34 of chain A makes a hydrogen bond to Lys67 of chain B and Asp34 of chain B makes two hydrogen bonds to Arg44 on chain A.

The inter-layer contacts do not involve direct protein-protein interactions but are provided instead by the ligand that was co-crystallised. C-diGMP crosslinks the crystal layers via the DGC domain of chain A and the DGC domain of chain B from the neighbouring dimer (Figure 3.33). Due to NCS symmetry, there are two such contacts per pair of symmetry-related dimers. The interactions are isologous, and both hydrogen bonding and hydrophobic interactions are involved. More details, including the biological relevance, will be discussed later on page 80.

Possibility of DGC domain swapping

Weak electron density between the end of D2 domain and the beginning of the DGC domain of the other subunit of the dimer has been observed (Figure 3.34). The gap between the D2 and DGC domains from different monomers is 10.2 Å, which can probably allow the five residues in the linker to adopt a coiled or a somewhat helical conformation, considering the pitch of an α-helix is 5.4 Å for 3.6 residues. Thus, the possibility of domain swapping exists but has to be proven.

Domain structures

D1 and D2 domains adopt the typical fold of response regulator receiver domain

D1 domain ranges from residues 2 to 136. It adopts the typical $(\beta/\alpha)_5$ fold found in CheY and many other response regulator receiver domains including FixJN. Superimposition of the D1 domain with CheY (PDB code 2CHE [59]) gave an rmsd of 1.3 Å (Table 3.2). The central β-sheet of D1 domain consists of five parallel β-strands in the order 2,1,3,4,5 and is flanked by five α-helices (Figure 3.29 and 3.35). This domain carries all the residues required for activation of a response regulator. In the acidic cluster for phosphorylation, the carboxyl groups of Asp9, Asp10, and phosphoacceptor Asp53, and the mainchain carbonyl of Met55 coordinate a Mg^{2+} ion (Figure 3.35B). The carboxyl group of Asp53 is additionally hydrogen bonded to Lys105. In addition, the D1 domain carries the residues Thr83, Phe102

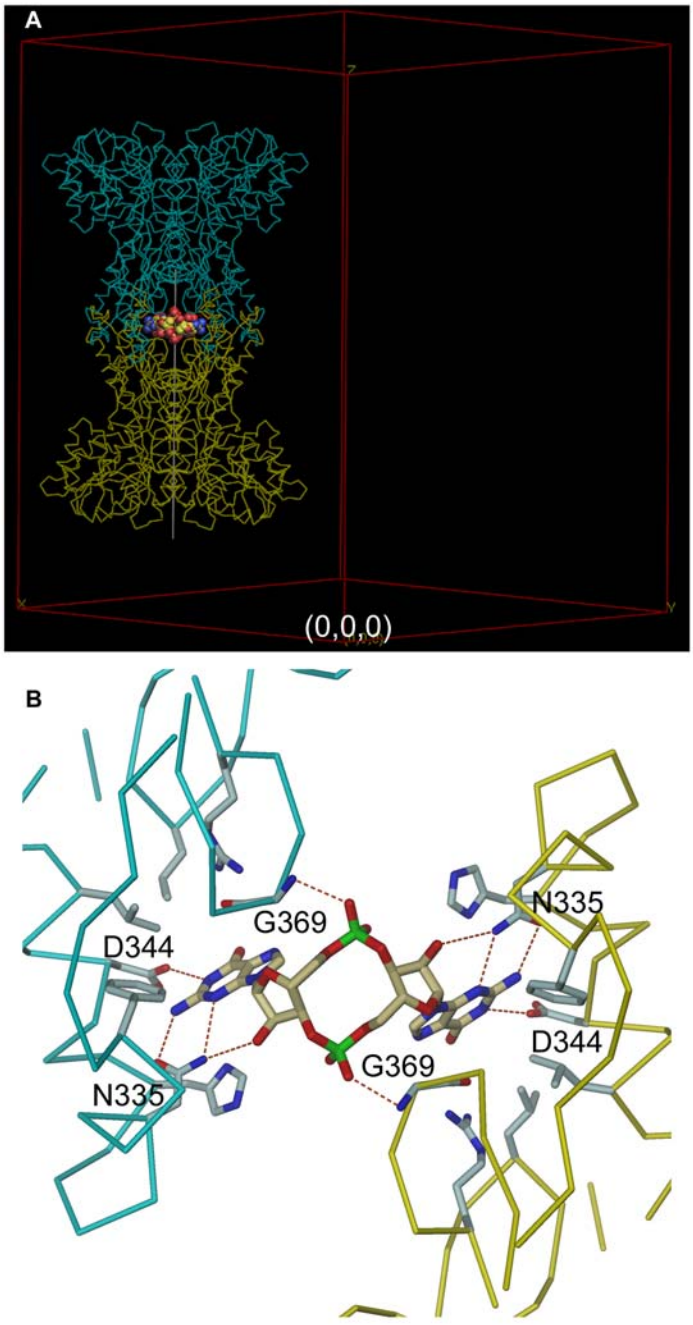


Figure 3.33: Inter-layer packing in CF crystal. A, Ligands (shown in cpk) mediate the packing of crystal layers in the unit cell, as viewed along the diagonal axis from the origin to xy. B, Closed-up view of the inter-layer crystal packing. The ligand, c-diGMP, mediates the contacts of DGC domains by hydrogen bonding and hydrophobic interactions.

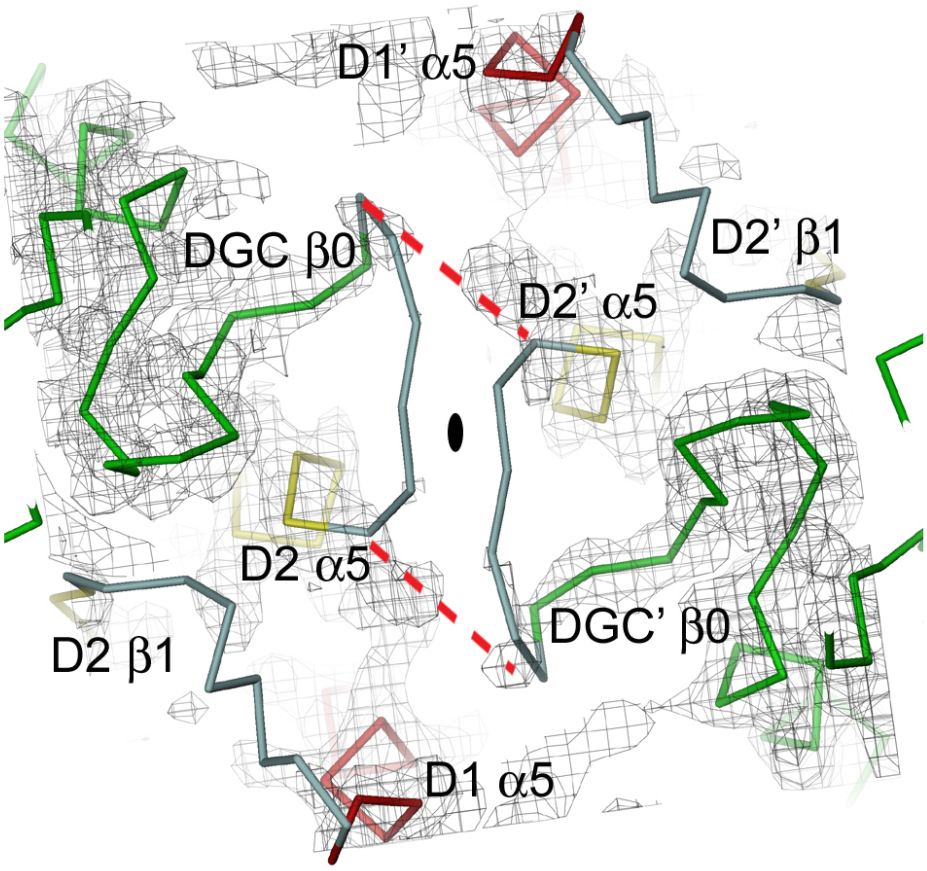


Figure 3.34: Possibility of DGC domain swapping. Topview down the NCS two-fold axis reveals that the ends of the domains are very close. Helix $\alpha 5$ of D1, helix $\alpha 5$ of D2, and strand $\beta 0$ of DGC line the dimer interface. The $2F_o$-F_c map contoured at 1 σ shows the poor electron density of the linkers but possible connectivities (red broken line) between helix $\alpha 5$ of D2 domain and strand $\beta 0$ of the DGC domain of the other monomer.

Table 3.2: Structural comparison of domains at a glance. Cα atoms of molecule 2 have been superimposed on that of molecule 1 using the programme TOPP [40].

Molecule 1	Molecule 2	Cα matched	rms (Å)	PDB file
D1	D2	123	1.2	-
D1	CheY	107	1.3	2CHE
D2	CheY	91	1.2	2CHE
DGC	AC catalytic	92	1.5	1CJK
DGC	DNAP I palm	63	1.7	1NK4

and Lys105 in the characteristic positions at which large structural changes induced by activation have been shown [7, 36, 35] (Figure 3.35C).

The D2 domain ranges from residues 147 to 280. It also shares the CheY fold (Figure 3.29) with an rmsd of 1.2 Å when superimposed onto the structure of CheY (Table 3.2). However, it lacks the phosphoacceptor and the conserved residues necessary for conformational changes induced by phosphorylation. D1 and D2 domains are related by a quasi-two-fold symmetry and shows an rmsd of 1.2 Å . In both domains, the helix α5 is significantly longer than that in CheY homologues and protrudes from the globular domain. For example, the D1 and D2 domains are 1.7 x and 1.5 x, respectively, longer than that in CheY.

D1 domain is in the inactive form

The characteristics of the inactive response regulator receiver domains include outward orientations of the β4-α4 loop and the relevant residues, like T87 and Y106 in CheY, with respect to the acidic pocket [50]. The β4-α4 loop and the sidechains of T83, F102 and K105, which typically bring about conformational changes driven by phosphorylation [7, 36, 35], also adopt similar conformations as that in the inactivated CheY, much in contrast to the conformations found in the activated form of CheY.

In the D1 domain, the sidechains of T83 and F102, as well as the β4-α4 loop, are in outward conformations as the corresponding sidechains and the β4-α4 loop in the inactivated CheY (Figure 3.35C-D). The sidechain of K105, however, forms a hydrogen bond with the phosphoacceptor D53, which is not found in the inactivated CheY. In CheY that is activated using the phosphate mimicry BeF_3^- (PDB code 1FQW; [36]), the T87 sidechain swings towards the acidic cluster to form a hydrogen bond with the fluoride

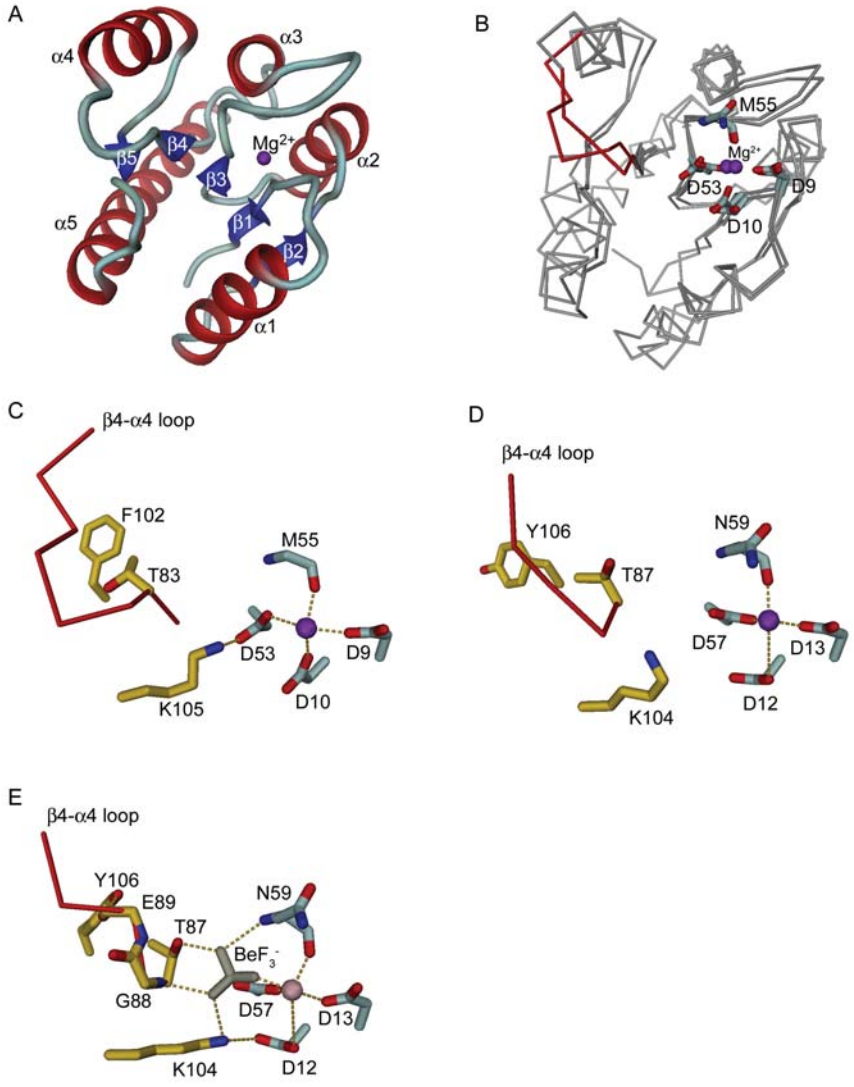


Figure 3.35: D1 domain structure. A, Ribbon representation of D1 shows that it has the typical CheY fold. Secondary structures are labelled accordingly. B, Superimposition of the Cα trace of D1 on that of inactivated CheY [59] reveals similar overall structures, including the acidic cluster for phosphorylation (in sticks; sidechain of M55 not shown). The β4-α4 loop that carries out the landmark conformational change upon activation is shown in red. The relevant residues are labelled according to that in D1 domain. C, Close-up view of the conformation of the residues in the acidic cluster, residues showing characteristic conformational changes upon domain activation (carbon atoms in yellow), and the β4-α4 loop in the D1 domain. D, The same in inactivated CheY. E, The same in BeF_3^--activated CheY [7, 36, 35].

atom of BeF_3^- which also coordinates the mainchain amide of N59 and G88, the sidechain of K104, and a Mn^{2+} ion in the acidic pocket via hydrogen bonds. This brings the β4-α4 loop inwards to open up a pocket between helix α4 and the β-sheet that allows Y106 to swing in to form a hydrogen bond with mainchain carbonyl of E89 (Figure 3.35E). Such open conformations of the sidechains and the β4-α4 loop, and such hydrogen network are absent in the D1 domain. We conclude that D1 is in the inactive form.

DGC resembles the AC catalytic domain

The DGC domain ranges from residues 288 to 455. Its structure comprises an open β-sheet composed of seven β-strands in the order 0,0',2,3,1,4,5, all of them anti-parallel except the last strand (Figure 3.36). The β-sheet is packed by five α-helices on both sides of the central sheet. There is also a small sheet of two anti-parallel β-strands separate from the main sheet.

The $\beta\alpha\alpha\beta\beta\alpha\beta\alpha$ fold of the core of the DGC domain resembles that of the AC catalytic domain [61] and of the DNAP I palm domain [13] (Figure 3.29 and 3.37). This common core structure comprises an anti-parallel β-sheet of four strands in the order 2,3,1,4. The DGC domain and the AC catalytic domain have an additional β5 strand parallel to the β4 strand to finish the sheet. The sheet is packed by four α-helices in all three structures.

The highlight of the fold is the β-hairpin connecting the strands β2 and β3, and it has been shown to form part of the catalytic site of AC and DNAP. In the DGC domain, the GGEEF sequence motif that is highly conserved in the DGC domain family but not in AC or DNAP is located on this β-hairpin (Figure 3.30). Unlike in the AC catalytic domain or the DNAP I palm domain, the β-hairpin in DGC domain is preceded by a bend, located between the residues R366 and G368, that apparently stabilises it using an elaborate hydrogen bonding network that involves mainchain atoms from the strands β0 ad β0' (Figure 3.38A). Specifically, the mainchain carbonyl group of G369 receives a hydrogen bond from the N1 amino group of R366 on the bend (Figure 3.38B), whose orientation is stabilised by a hydrogen bond between its Nζ amine group and the sidechain carboxyl group of D292 from the β0 strand. The orientation of the sidechain carboxyl of D292 is in turn stabilised by hydrogen bond interactions with the sidechain hydroxyl and mainchain amide of T295, and the mainchain amide of L297 from β0'. The bend also seems to be further stabilised by the hydrogen bond between the mainchain amide group of Y367 and the carbonyl oxygen of E371. The bend, together with the β0 ad β0' strands, are absent in AC and DNAP (Figure 3.38C,D).

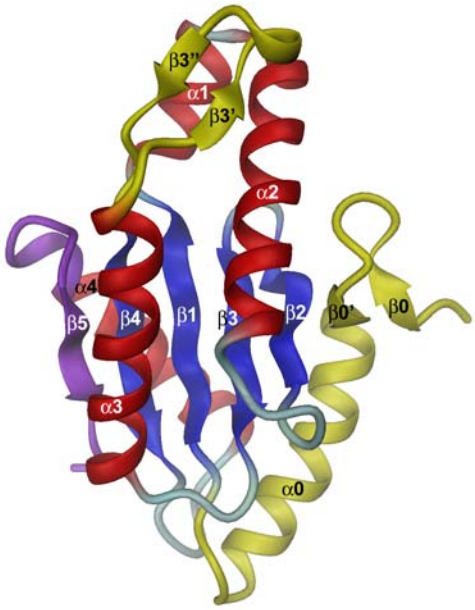


Figure 3.36: Ribbon representation of the DGC domain structure. The β-strands in blue and the α-helices in red refer to the core structure common in the DGC domain, AC catalytic domain and DNAP palm domain. The β-strand in purple is found in both the DGC domain and the AC catalytic domain. The secondary structures coloured in yellow are found only in the DGC domain. Secondary structures are labelled accordingly.

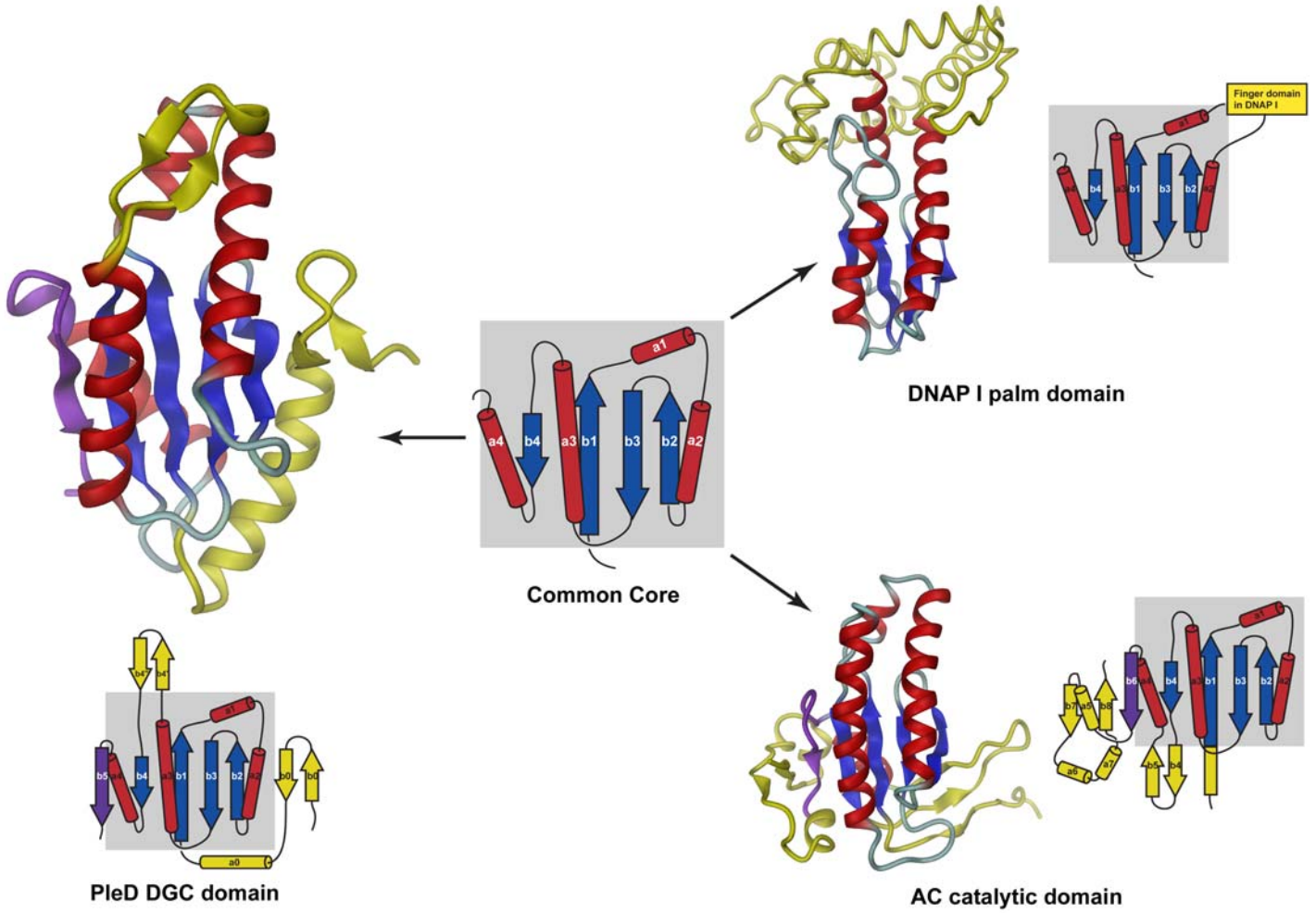


Figure 3.37: Comparison of the folds in the DGC domain (left), the AC catalytic domain (bottom, right) and the DNAP I palm domain (top, right). The common features in the three structures include an open β-sheet of four anti-parallel strands (blue arrows) in the order 2,3,1,4, and four flanking α-helices (red cylinders). They are represented by a topological diagram in the centre of the page. Topological diagrams referring to each structure are also shown. The colour scheme is the same as in Figure 3.36. The secondary structural elements unique to each of the structures are shown in yellow.

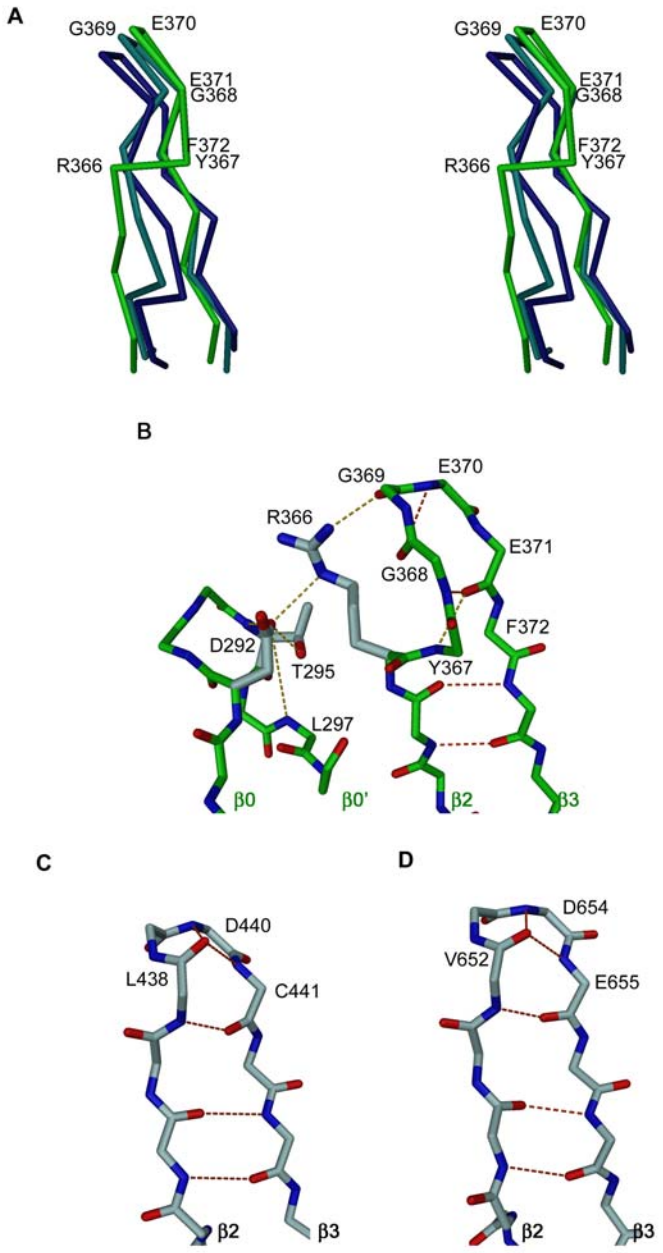


Figure 3.38: Differences in the conserved β-hairpins. A, Stereodiagram of the sideview of the superimposition of the β2 and β3 strands including the β-hairpin from DGC (green), AC catalytic domain (cyan) and DNAP palm domain (dark blue). Residues located at the bend preceding the hairpin, as well as the conserved sequence motif, in the DGC domain are labelled. B, Frontview of the hydrogen bonding network (yellow broken lines) that stabilises the hairpin. Only the mainchain backbone is shown except the sidechains of relevant residues. This hydrogen bonding network involves the mainchain of G369 from the β-hairpin, the sidechain of R366 from the bend, D292 from the β0 strand and other residues from the β0' strand, and the mainchain carbonyls of E371 and Y367. The usual hydrogen bonds found within the β-hairpins and the two β-strands are shown in red broken lines. Note that the carbonyl oxygen of G368 adopts an unusal conformation that prevents it from making a hydrogen bond with the amide group of E371. B, The same view of the hairpin for the AC catatytic domain. C, The same for the DNAP I palm domain.

Other structural features unique in DGC domain include the helix $\alpha 0$ that is N-terminal to the beginning of the core fold, and the small anti-parallel β-sheet of $\beta 3$' and $\beta 3$" that replaces the $\alpha 3$-$\beta 4$ loop (Figure 3.36).

Domain interfaces

D1-D2 domain interface is stabilised by quasi-isologous salt bridges

A surface of 1038.2 Å^2 is buried at the interface between D1 and D2 domains in the PleD monomer (Table 4.2). These two domains are related by a quasi-two-fold axis (Figure 3.39) and are stabilised by quasi-isologous salt bridges (Table 3.3) that involve helix $\alpha 4$ and strand $\beta 5$ from one domain and helix $\alpha 5$ from the other domain. Interestingly, these secondary structure elements have been shown in other response regulator receiver domains to carry out phosphorylation-induced conformational changes. It will be discussed in the next chapter how the D1-D2 domain interface might be affected by phosphorylation of the D1 domain through these conformational changes.

Table 3.3: Quasi-isologous salt bridges stabilising the D1-D2 domain interface connect residues from helix $\alpha 4$ on one domain to helix $\alpha 5$ on the other domain, and strand $\beta 5$ on one domain to helix $\alpha 5$ on the other domain. The domain that the residue belong to is provided in bracket.

$\alpha 4$-$\alpha 5$ salt bridges		$\beta 5$-$\alpha 5$ salt bridges	
$\alpha 4$	$\alpha 5$	$\beta 5$	$\alpha 5$
R91 (D1)	D257 (D2)	R115 (D1)	D250 (D2)
R237 (D2)	D108 (D1)	R264 (D2)	D101 (D1)

D2-DGC domain interface is smaller than D1-D2 domain interface

The domain interface of D2-DGC is smaller than that of D1-D2 and covers an area of 686.6 Å^2. The two domains interact mainly through two salt bridges between R155 of strand $\beta 1$ in D2 domain and D378 of strand $\beta 3$ in DGC domain, and E182 of strand $\beta 2$ of D2 domain and R313 of helix $\alpha 0$ of DGC domain. There is also a hydrogen bond between R222 in D2 domain and the mainchain amide of G305 of helix $\alpha 0$ in the DGC domain (Figure 3.40). In addition, the domain interface is further stabilised by hydrophobic interactions between the helix $\alpha 2$ and the $\alpha 2$-$\alpha 3$ loop of D2, and between the helix $\alpha 0$ and the $\alpha 2$-$\alpha 3$ loop of DGC.

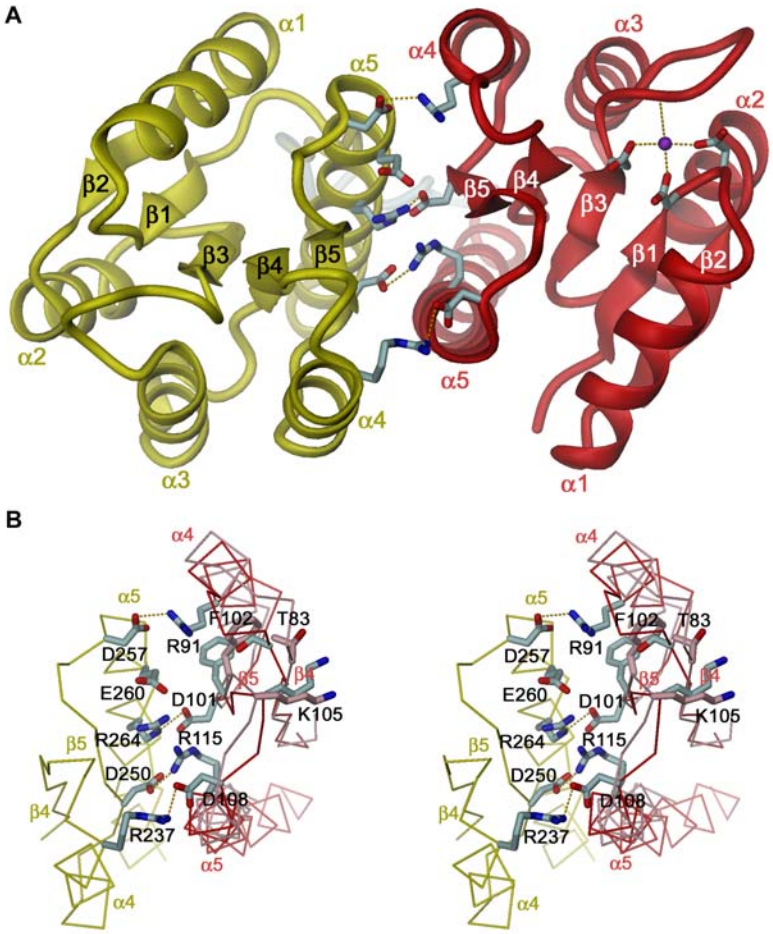


Figure 3.39: D1-D2 domain interface. A, D1 (red) and D2 (yellow) domains are related by a quasi-two-fold axis that goes into the paper. Their interface is stabilised by quasi-isologous salt bridges (in sticks) contributing by residues on the helices $\alpha 4$, $\alpha 5$ and the strand $\beta 5$. The sidechains of the residues involved in the acidic cluster for phosphorylation (in sticks) and the bound Mg^{2+} ion (purple) are also shown. B, Stereoview at the arrangement of the salt bridges at the D1-D2 domain interface. The salt bridges are in close vicinity to T83, F102 and K105 that are involved in conformational changes resulting from phosphorylation. The Cα trace of the phosphorylated FixJN (pink) and the relevant sidechains (carbons in pink) that are affected by phosphorylation are shown for reference but not labelled.

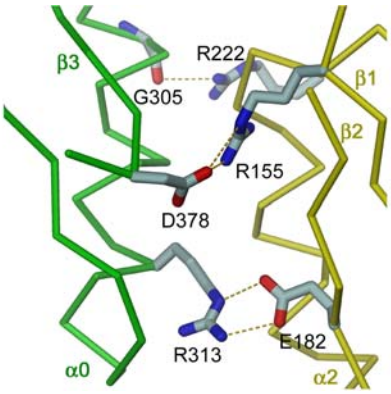


Figure 3.40: D2-DGC domain interface. Two ion pairs and a hydrogen bond that involves a mainchain atom are important in stabilising this interface.

PleD has a small dimer interface

PleD is a dimer in the asymmetric unit (Figure 3.41). The dimer arrangement is mainly stabilised by isologous interactions between residues from the D1 and D2 domains. The dimer interface covers an area of 815.7 Å^2 per monomer (Table 4.2). A central pore of the width of 9 Å and the length of 20 Å extends along the symmetry axis.

The dimeric contacts involve isologous D1-D2 interactions (Figure 3.41A-B). The first interaction consists of a hydrogen bond between Gln131 and Glu221 of which Glu221 is mildly conserved among the PleD homologues but only in terms of charge (Figure 3.41C). The second interaction consists of hydrophobic packing between Leu124 and the Cζ atom of Arg224, a residue that is also conserved. Both sets of interactions involve helix α5 of D1 and the 3_{10}-helix between α3 and β4 of D2. The third interaction seems to be the most significant. A strictly conserved residue Tyr26 among the PleD homologues from helix α1 forms a hydrogen bond with Arg239 and serves as a ridge that fits into the well-defined pocket formed by a number of residues, Arg212, Asp209, Ala215, Leu211, Lys242, Ile246 and Arg239, from the other monomer (Figure 3.41D). All these residues come from the helices α3 or α4 of D2, and hence, confer rigidity to the pocket. Asp209 and Leu211 are strictly conserved among PleD homologues whereas Arg212, Lys242 and Ile246 are charge- or type-conserved.

It is worth noting that the N-terminus of the peptide is in the close vicinity (Figure 3.42), suggesting that an N-terminal His-tag might cause steric clashes at the dimer interface.

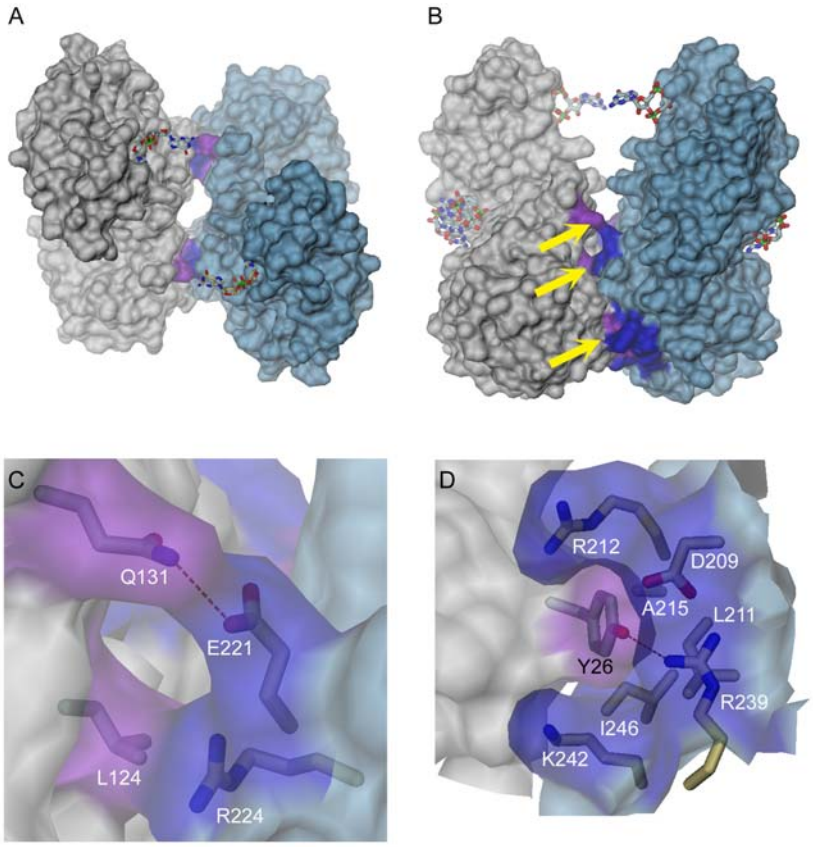


Figure 3.41: Packing of the NCS dimer. A, Surface representation of the CF dimer viewed along the NCS axis. DGC domains are at the top and D1 and D2 domains are at the bottom. Chain A is coloured in grey and chain B in light blue. The points of dimer contacts are coloured in purple on chain A and blue on chain B. The c-diGMP ligands are shown in sticks. B, Sideview of the NCS-related dimer reveals three points of contacts (arrows) on one side of the dimer, which are enlarged in panel C and D. C, Hydrogen bonding between Q131 and E221 and hydrophobic interaction between L124 and R224. Protein surface is set to semi-transparent for clarity. D, Hydrogen bonding between Y26 and R239 situated in a well-defined pocket.

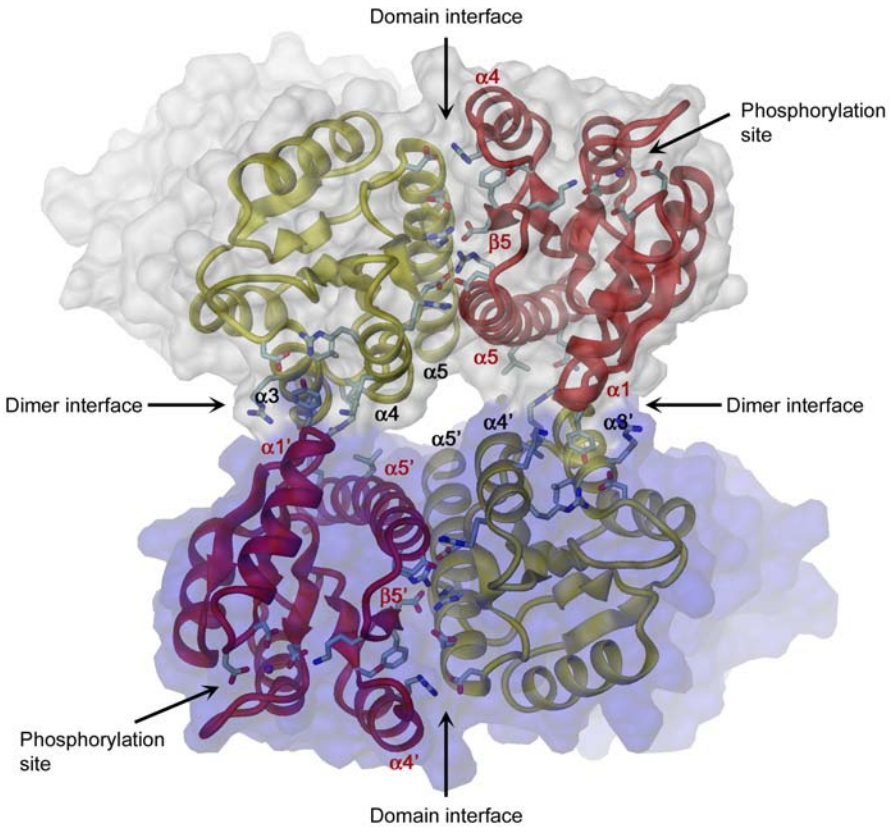


Figure 3.42: Overview of the arrangement of D1 (red) and D2 (yellow) domains in the PleD dimer viewed from the D1-D2 dimer stem. D1 and D2 domains interact using distinctive surfaces for domain packing within the PleD monomer and for PleD dimerisation. The overall shape of PleD is shown as a semi-transparent surface with the surface of one monomer coloured in grey and that of the other in blue. The DGC domains are located at the back of the dimer base but are omitted with the disordered linkers for clarity. Secondary structures involved in the two types of D1-D2 interfaces are lablled.

Ligand binding

PleD monomer has two c-diGMP binding sites (Figure 3.28B). The first one is annotated as the catalytic active site, A-site, which is located in the conserved β-hairpin region in the DGC domain. The second binding site is unexpected. It is named the product inhibition site, I-site, and is located at the D2 / DGC domain interface. The annotations of these nucleotide binding sites will be explained later.

Crystal packing suggests that the active site is two-fold symmetrical

The catalytic active site is suggested by the crosslinking of two DGC domains from adjacent symmetry-related PleD molecules by a c-diGMP molecule (Figure 3.43). The two DGC domains, as well as the product molecule, are related by local two-fold symmetry. This local axis is a direct result of the orthogonal intersection of the NCS and crystallographic axes, both of which being two-fold. Although this is an artefact of crystallisation specific to our case, we believe that this reflects the situation in solution in which DGC domains from the same PleD dimer can be bridged by the product molecule for the following reasons. Firstly, the active site is located on the β2-β3-hairpin not only found in the DGC domain, but also in AC and DNAP. Secondly, this β2-β3-hairpin carries the GGDEF sequence motif which is highly conserved in the DGC domain family. Thirdly, the presence of the highly specific guanine binding pocket suggests this nucleotide binding site is biologically relevant. In addition, the possible salt bridges between R300 and D336 that connect the two DGC domain lends support to the significance of such an arrangement of two DGC domains as shown in the crystal structure (Figure 3.44). All these will be discussed in more detail in the following sections.

GGEEF motif containing β-hairpin binds the nucleotide

As in the AC catalytic domain and the DNAP palm domain, the β2-β3-hairpin in DGC domain forms part of the nucleotide binding site (Figure 3.44). In particular, it carries the conserved GGEEF sequence motif. The ribosyl and the α-phosphate of the c-diGMP molecule extend over the β2-β3-hairpin and completely cover G369 which probably explains the conservation of a glycine at this position. The mainchain amide of G369 donates a hydrogen bond to a non-ester oxygen of the α-phosphate. The carboxyl

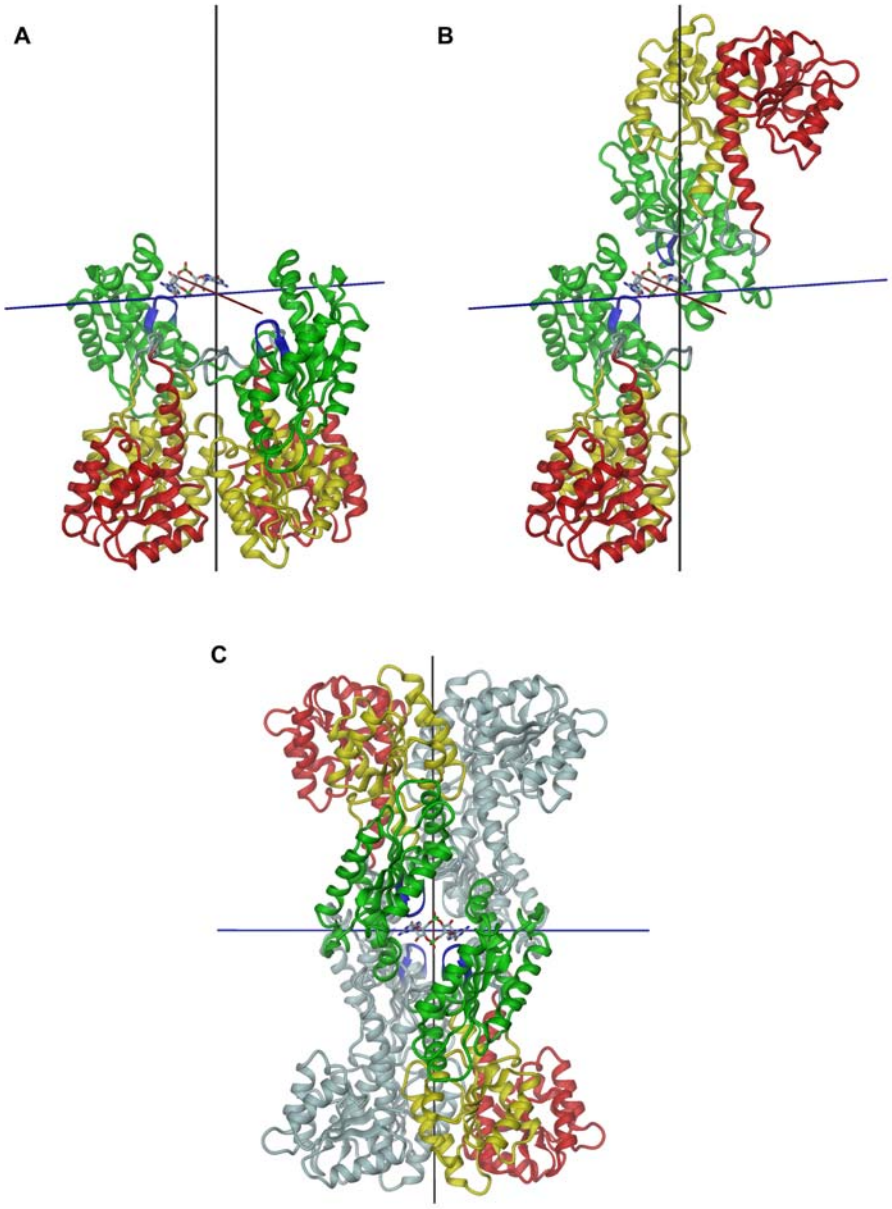


Figure 3.43: Two-fold related active site as implied by crystal packing. A, PleD dimer in the asymmetric unit. Colour scheme is the same as in Figure 3.28. NCS axis is coloured in black, crystallography axis in blue, and the local two-fold axis in red. B, The same view showing the monomers that are related by the local two-fold axis (red) interacting across the crystal layers. The top monomer is crystallographic symmetry related to the NCS partner (shown in panel A but not here) of the monomer below. C, Head-on view of this inter-layer monomer-monomer interaction along the local two-fold axis. The product molecule also obeys this two-fold symmetry. Another inter-layer interaction is given by the NCS partners (grey) of these monomers towards the back of the molecules. Their crosslinking product molecule is removed for clarity.

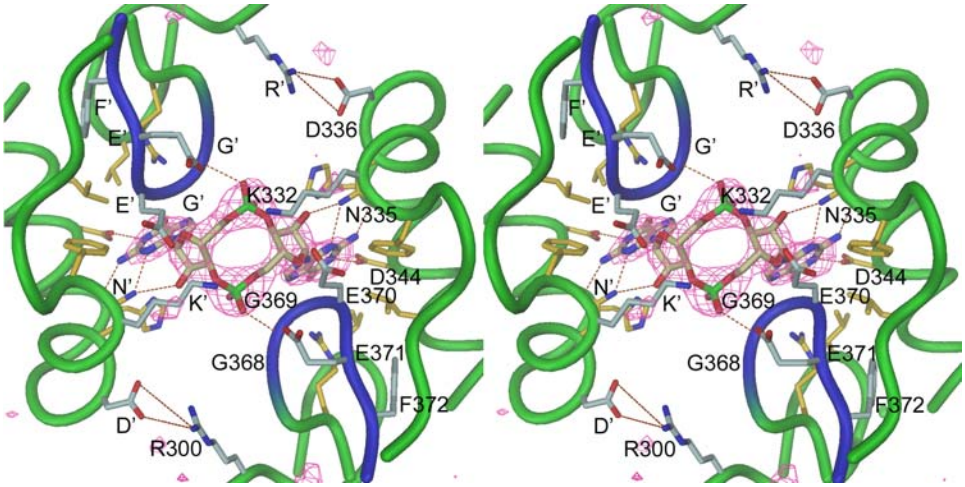


Figure 3.44: Stereodiagram of product binding at the A-site viewed along the local two-fold axis. Crystal packing suggests that the DGC active site requires two DGC domains arranged in a two-fold related fashion. The β-hairpin carrying the conserved GGEEF sequence motif is coloured in blue. The relevant residues are labelled in full on one DGC and in abbreviation with a 'prime' sign on the other DGC domain. Residues lining the specific guanine pocket are coloured in yellow. The F_o-F_c map is contoured at 4 σ. The sidechains of R300, K332, D336, E370 and E371 are partially disordered.

group of E370 is partially disordered but the position of its β-carbon indicates that the sidechain may point towards the ribosyl 3' oxygen direction in an enzyme-substrate complex. Together with K332 which is conserved in the DGC domain family, E370 probably serves a catalytic function that will be discussed later. The carboxyl group of E371 is also disordered but may point towards the α-phosphate as indicated by the β-carbon position. The possible role of E371 in nucleotide binding will be discussed in the next chapter. The final residue in the sequence motif, F372, is situated in a hydrophobic core in which the guanine is located.

c-diGMP binding at the active site involves specific guanine recognition

The guanine base is inserted in a pocket formed mainly by helices α1 and α2 of DGC (Figure 3.45). The strength and specificity of the binding is provided by hydrogen bonding interactions. The carboxyl group of Asp344 receives a hydrogen bond from the NH group at position N1 of the guanyl group. Moreover, Asn335 forms a bidentate hydrogen bond with the guanyl group. Whereas its carbonyl Oδ receives a hydrogen bond from the exocyclic amine group at position N2 of the guanyl base, its amine group donates a hydrogen bond to the N3 atom of the guanine. The hydrogen bond between the exocyclic N2 amine and Asn355 particularly explains the specificity of DGC domain towards guanine over adenine [45] because asparagine cannot interact with the carbon atom C2 in the adenine base. The location of the G343, which is invariant in the DGC domain family, also seems to make room for this exocyclic amino group in the guanine. Only the oxygen atom O6 of guanine, which is replaced by an amino group in adenine, does not make interactions with any protein residues.

In addition to the hydrogen bonding interactions, guanine also makes hydrophobic contacts with the apolar sidechains of L294, F331, and L347 (Figure 3.45A,C), which, together with F372 from the signature motif, are part of the hydrophobic core.

Apart from guanine recognition, the correct orientation of the ribosyl group might be ensured by a hydrogen bond between the ribosyl 2' oxygen and the amine group of Asn335. The hydrogen bond between an α-phosphate non-ester oxygen and the mainchain amide of G369 might also help in keeping the position of the phosphate group close to the β-hairpin.

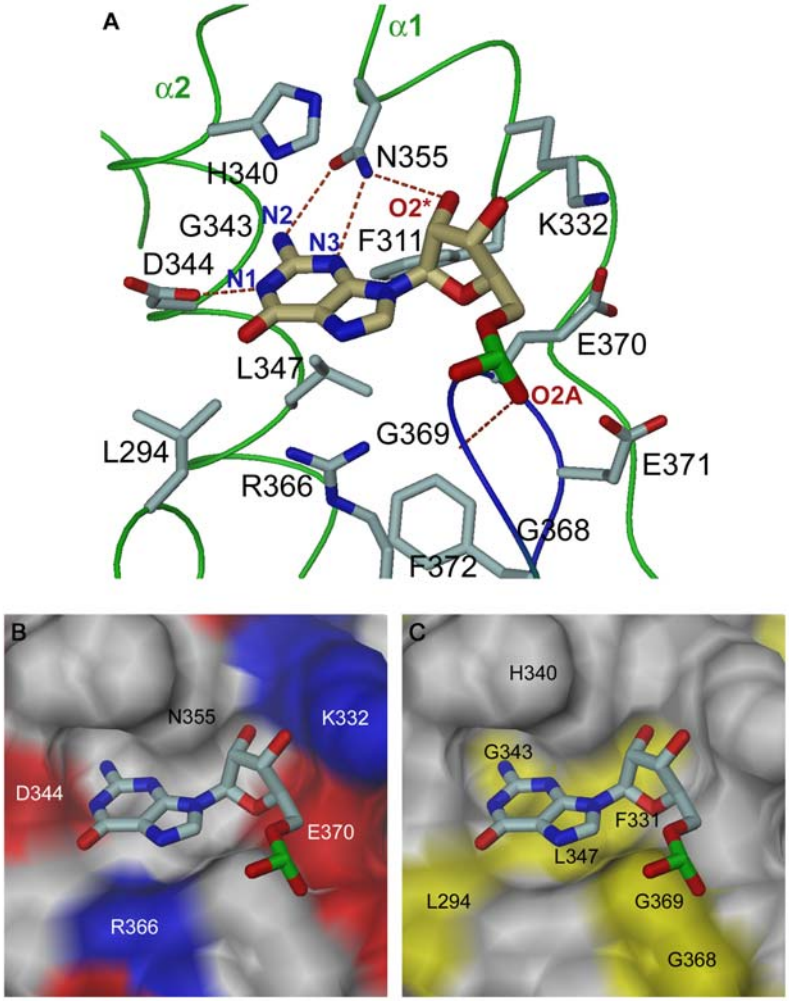


Figure 3.45: c-diGMP binding to the A-site viewed directly into the guanine binding site. A, Only half of the c-diGMP molecule is shown for clarity. All relevant sidechains and the atoms of the nucleotide involved in the protein-ligand interactions are shown and labelled. The β-hairpin is coloured in blue as before. B and C, Chemical properties of the c-diGMP binding surface. Acidic surface is coloured in red, basic in blue, hydrophobic in yellow, and the rest in grey.

Two intercalated c-diGMP molecules bind to the inhibition site at D2-DGC domain interface

The PleD structure shows an unexpected second c-diGMP binding site at the D2-DGC domain interface, known as the inhibition site. At this I-site, two c-diGMP molecules with mutually intercalated purine bases are bound (Figure 3.46A,B). Intercalation of this di-nucleotide has been reported previously [16, 39] and might also exist in solution. In our structure, the intercalation is provided by the stacking of the guanine bases and the intermolecular hydrogen bonds between the central guanines and the phosphate groups (Figure 3.46D, E). Unlike in the previous structure (Figure 1.2), however, no Mg^{2+} ion is found in PleD to coordinate the two central guanines.

The product ligand is bound to both D2 and DGC domains through a strong hydrogen bonding network (Figure 3.46C-F). The predominant guanyl base-arginine pairing occurs three times involving both intercalated molecules. Two of these pairing interactions are bidentate, in particular, and involve the N7 and O6 atoms of the guanyl base (Figure 3.46C,E). The binding of one of the product molecule to PleD seems to be tighter than the other molecule. In this tightly bound molecule, the binding of one of the guanyl groups is stabilised by a bidentate base-arginine pairing interaction with R390 and another bidentate interaction with D362 via its N1 and N2 amine groups (Figure 3.46C). The other guanyl group in this molecule is hydrogen bonded to R178 via its N7 atom (Figure 3.46D). As for the other product molecule, only one guanyl base-arginine pairing is involved (Figure 3.46E). The second guanyl group in this molecule is hydrogen bonded via its N2 amine group to the mainchain amide groups of G174 and H177 (Figure 3.46F). The tight binding of the c-diGMP molecules observed here is consistent with the observed co-purification of c-diGMP during protein preparation (Figure 3.6 and 3.7).

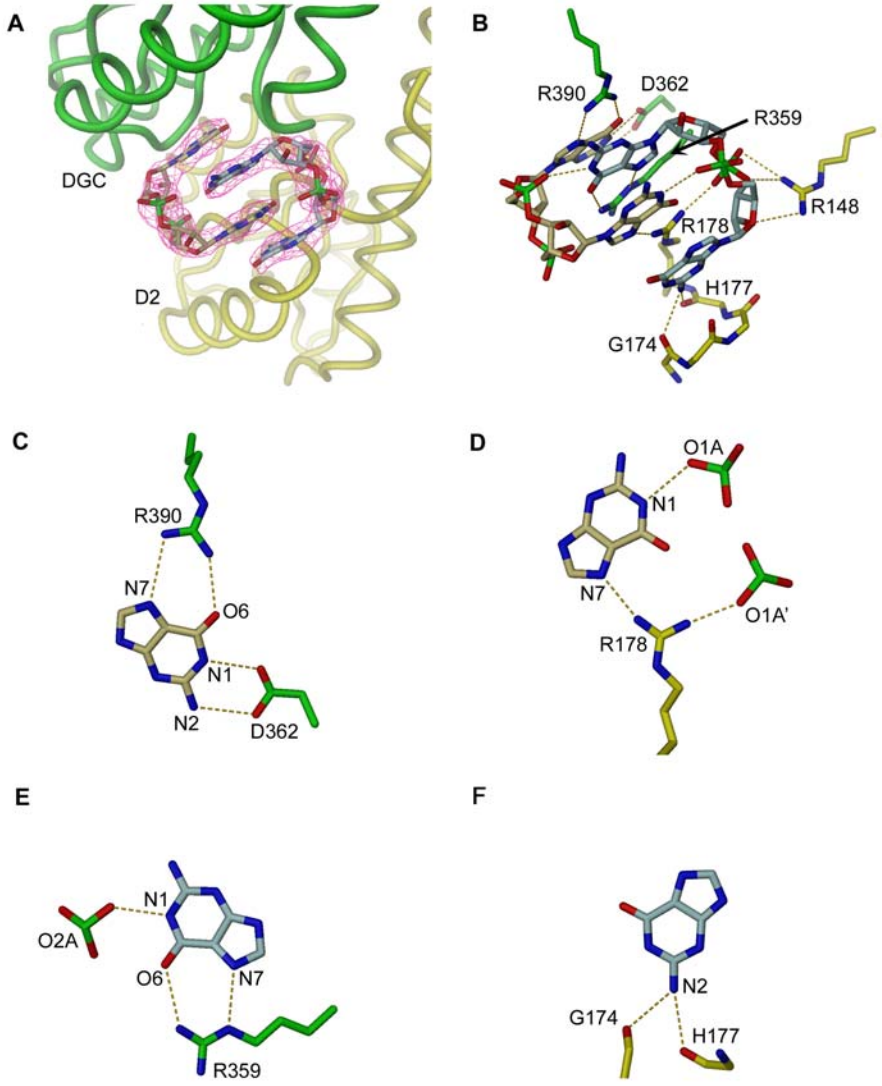


Figure 3.46: c-diGMP binding to the allosteric I-site. A, Two intercalcated c-diGMP molecules bind at the D2 (yellow)-DGC (green) domain interface. Guanyl groups from the two product molecules (carbons in grey and khaki) alternately stack on top of each other. The F_o-F_c map is contoured at 4 σ. B, Strong hydrogen bonding network connecting the ligand and D2, the ligand and DGC, and the two product molecules. C-D, Interactions of individual guanyl groups of one c-diGMP molecule (khaki) with PleD. The oxygen atoms of the α-phosphate groups that are hydrogen bonded to the guanyl groups are shown together with the entire α-phosphate group. E-F, Interactions of individual guanyl groups of the other c-diGMP molecule (grey) with PleD. The sidechain of H177 is not shown.

Chapter 4

Discussion

The signalling PleD protein is important for an asymmetric cell differentiation step in *C. crescentus* [1, 23]. It is a multidomain response regulator that belongs to the two-component signal transduction pathway prevalently used in bacteria [50]. Unlike the typical response regulators, PleD contains three domains. The N-terminal D1 domain serves as the phosphorylatable receiver domain like the CheY protein [35, 58], while the middle D2 domain only shares the CheY topology. The C-terminal DGC domain shows the proven diguanylate cyclase activity to synthesise the putative secondary messenger molecule, c-diGMP [45]. The high abundance of the DGC domain in diverse bacterial genomes and its role in producing c-diGMP has pointed to a novel signalling pathway [18]. No downstream components of this pathway have yet been characterised. The PleD protein, thus, provides the best starting point to understand c-diGMP signalling.

Evaluation of PleD crystallisation

Based on the PleD structure, it is now evident why the selected construct and ligand were useful for crystallisation. The selected construct makes room for crystal contact and the selected ligand provides a crucial crystal contacts.

The present crystal form of the CF protein crystal shows that the N-terminus of the peptide is located at the dimer interface, suggesting an N-terminal His-tag might prevent the formation of a PleD dimer. In fact, the activity assay conducted in Prof Urs Jenal's Group has shown that the N-terminal tagged NF is inactive (personal communication). These might explain why the NF protein only gave crystals of inferior quality (Figure

3.17). The C-terminal tagged CF protein, on the other hand, has the His-tag away from the dimer interface, thus allowing PleD dimer formation.

The crystal form of the CF protein also presents a special scenario in which the two-fold symmetrical ligand, c-diGMP, crosslinks adjacent crystal layers by inserting into a substrate binding site on each layer. Attempts in co-crystallisating CF protein with substrate analogues like GMP-PNP and GTPγS failed to give crystals diffracting better than 3 Å (Figure **??**). It is very likely that they fullfil the stereochemical requirements of the nucleotide binding site due to their structural similarities to the c-diGMP molecule. However, they are unable to physically bridge the two nucleotide binding sites together to maintain the active site, simply due to the lack of covalent bridges between the two bound molecules. This implies that we have to screen for a new crystallisation condition if we were to obtain a substrate-enzyme complex structure.

Evaluation of domain delineation results

The knowledge of the full-length structure of PleD also allows the evaluation of the three methods used to determine the boundaries of different domains prior to structure determination (Table 4.1).

Construct	Predicted domain(s)	Approach	Correct domain definition	Correct boundary prediction
Full-length	All	N/A	N/A	N/A
137	D2, DGC	Limited Proteolysis	-	+
150	D2, DGC	Limited Proteolysis	-	+
153	D2, DGC	Limited Proteolysis	-	+
287	DGC	Sequence alignment	+	+
319	DGC	Threading	+	-

Table 4.1: Performance of the three approaches on domain prediction and determination. '+' symbolises positive, '-' negative, and 'N/A' not applicable.

Published information was used in the first instance. The entry of PleD (Q9A5I5) in the SwissProt database [6] automatically defined a 'GGDEF' domain in the PleD sequence. I searched for 'GGDEF' domain homologues in the SwissProt database, carried out multiple sequence alignments and secondary structure prediction to map the domain boundary (Figure 3.1). The result was the constructs, N287 and C287, both starting at residue 287. As seen in the structure, this fragment is indeed identical to the do-

main which we have coined the term 'DGC' after the biochemical evidence of diguanylate cyclase activity. The predicted boundary starts in a small turn that runs before the first strand of $\beta 0$ starting at residue 290 in the DGC domain (Figure 3.30). Despite the correction prediction of the domain boundary, these constructs were, however, highly insoluble in solution.

In the second approach, the results of the domain analysis by Pei and Grishin were directly used [46]. Pei and Grishin extensively threaded GGDEF domain homologues and detected a mild homology of this domain with the AC catalytic domain. They aligned all these sequences and predicted the domain boundary. According to their alignment, the DGC domain should start at residue 319 in PleD. As in the first approach, they successfully predicted the GGDEF domain as a single domain. However, they failed to determine the correct domain boundary by missing the N-terminal strands of $\beta 0$, $\beta 0$' and helix of $\alpha 0$ (Figure 3.30 and Table 4.1). The reason is that the GGDEF domain of PleD has structural features that are additional to those of the AC catalytic domain (Figure 3.37). When this happens, any bioinformatic analysis based on structural homologues is prone to fail.

The domain architecture was also analysed by experimental methods. Limited proteolysis of the full-length PleD protein was carried out in search for protease-resistant fragments (Figure 3.9). Two stable protease-resistant fragments were identified. One of them was the D1 domain, the domain homologue of the CheY protein, while the other comprised the entire length of the D2 and DGC domains. Three sets of constructs, N/C137, N/C150 and N/C153, were then designed with different starting residues in the linker between the D1 and D2 domains. All constructs show the right boundary (Figure 3.30 and Table 4.1), and all performed well in terms of expression, purity and solubility in the first screen (Figure 3.11). However, our structure reveals that D2 and DGC domains exist as two individual domains.

What makes the D1-D2 domain linker more accessible to the protease than the D2-DGC domain linker? The reason is that there is actually no trypsin cleavage site on the domain linker connecting the D2 and DGC domains (Figure 3.30). A possible cleavage site is located before the last residue of the helix $\alpha 5$ of D2 domain but was protected from cleavage.

As a summary, only the combined approach of sequence alignment of homologues and secondary structure prediction gave the right domain boundary. Nevertheless, the resulting constructs did not overexpress and were not stable in solution. It seems that nature has selected the optimal domain architecture present in PleD, as the full-length constructs performed the best in all experimental aspects (Figure 3.2 and 3.4).

Model of substrate binding to DGC

Based on the observed product-enzyme complex structure (Figure 3.44) and our knowledge of the DGC reaction (Figure 1.3), we propose a model for the binding of the GTP substrate to the A-site (Figure 4.1).

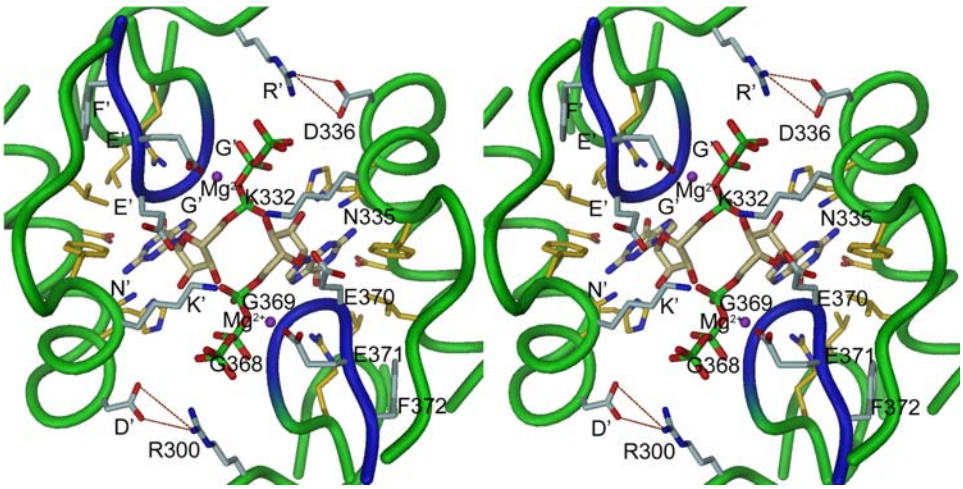


Figure 4.1: Modelled GTP substrate binding to A-site of DGC. The two GTP molecules are arranged in an anti-parallel fashion to allow β- and γ-phosphates to extend along the glycine side of the β-hairpin. The ribosyl and guanyl groups bind in a very similar way as they do in the enzyme-product complex structure.

Considering the specific interactions with the protein, the guanyl group of the GTP substrate probably binds in the same way as that of the product molecule, assuming that it remains stationary during catalysis. The ribosyl group and the α-phosphate probably also bind very similarly. The β- and γ-phosphate groups would extend from the α-phosphate alongside the highly conserved G368 of the β-hairpin, and bind via a Mg^{2+} ion to E371 from the conserved sequence motif. R300 would probably confer stabilisation of the negative charges of the phosphate groups. Applying the same rules to the GTP molecule on the other DGC domain, the result is two GTP molecules arranged in an anti-parallel fashion and well juxtaposed to allow for mutual intermolecular nucleophilic attack of the ribosyl 3' oxygen onto the α-phosphate.

Interestingly, a comparison of substrate binding in DGC domain, the AC catalytic domain and the DNAP palm domain reveals substantial differences, in spite of their common core structure (Figure 4.2). In both AC and DNAP,

the phosphate-binding site is located on the short P-loop between strand β1 and helix α1. The binding is stabilised by two basic residues such as R484 and K1065 in AC and H506 and R518 in DNAP, which apart from R484, come from the other subunit of the heterodimer in AC and from another domain in DNAP (Figure 4.2B,C). Although DGC domain contains this P-loop and the basic K442 and R446 that are at similar positions, juxtaposition of the phosphate groups in this location would forfeit the specific guanine-protein interactions and create a distance between the α-phosphate and the ribosyl 3' oxygen that is too long for an intermolecular attack (Figure 4.2A). On the other hand, an attempt to simultaneously fix the guanyl group to the guanine binding pocket and the phosphate groups to the P-loop would be sterically unfeasible.

In the DGC domain, the guanyl group in GTP is modelled to bind specifically as that in the product molecule to a guanine binding pocket formed between the helices α1 and α2 (Figure 3.45). For AC and DNAP, the base group is not bound at all to the catalytic domain or to the palm domain. Instead, it is recognised by the other subunit of the heterodimer in AC and by another domain in DNAP (Figure 4.2B,C). Specifically, the adenine base of the ATPγS substrate analogue in the AC catalytic domain is recognised by R938 and D1018 from the other subunit in the heterodimer (Figure 4.2B) [62]. As for DNAP I palm domain, the base of the incoming nucleotide for polymerisation with the primer DNA strand is stabilised by the stacking of Y526 from another domain in the protein that promotes the base-pairing of the nucleotide base with the template DNA strand (Figure 4.2C) [13].

Only the basic residue K332 in DGC (K1029 in AC and K522 in DNAP) is modelled to stabilise the α-phosphate during its pentavalent intermediate state and seems to play a similar role in the catalysis of the three enzymes.

Proposed catalytic mechanism

Based on the suggested mode of substrate binding in DGC domain, the following catalytic mechanism of DGC domain is proposed (Figure 4.3). For catalysis to occur, two substrate-loaded DGC domains have to arrange themselves very similarly as observed in the crystal. This might be facilitated by charge complementarity between D336 and R300 (Figure 3.44). Subsequently, the 3'-OH group of the GTP substrate has to be deprotonated to allow for an intermolecular nucleophilic attack onto the α-phosphate (Figure 4.3A, B). E370 from the sequence motif appears to be well positioned

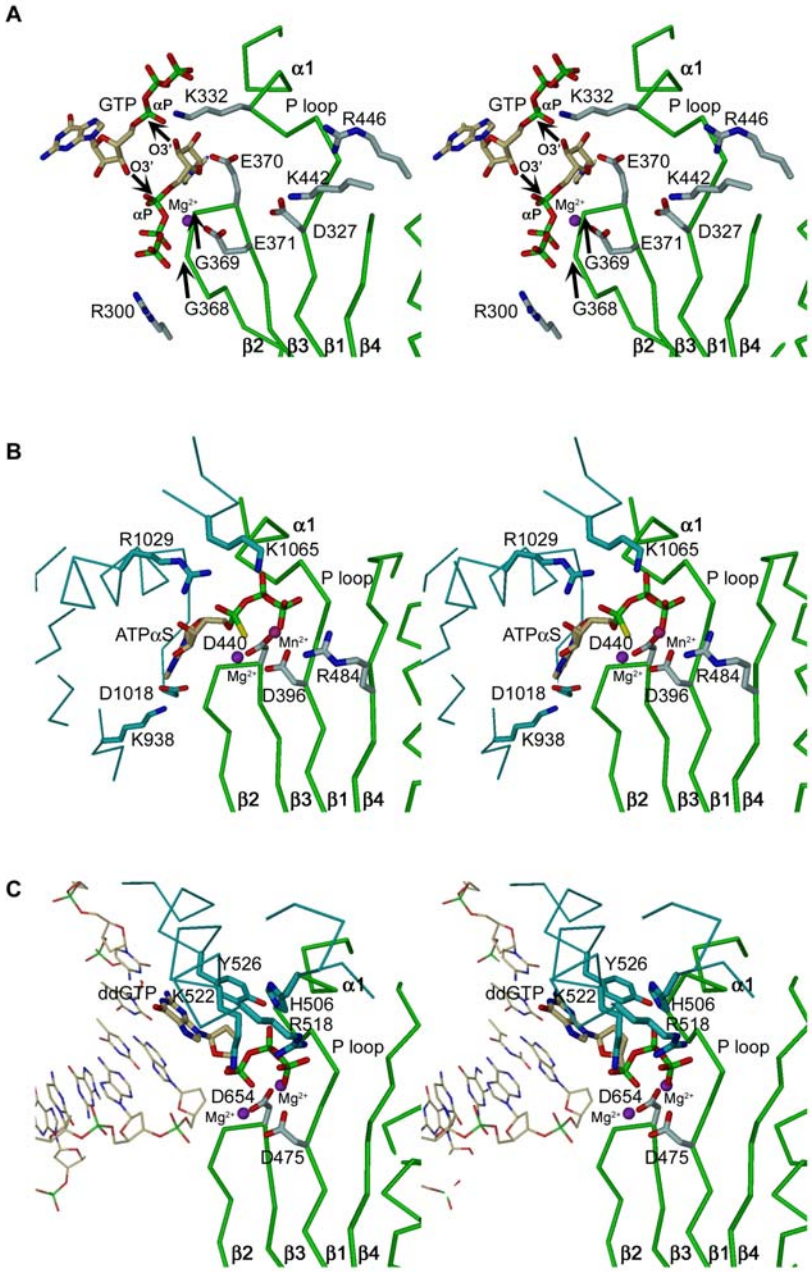


Figure 4.2: Stereoview of the different substrate binding modes in DGC domain, AC catalytic domain and DNAP I palm domain. The carbons of the substrates are shown in khaki. The core fold common in all three structures, except α2 helix that is removed for clarity, is shown in green. The backbone and the sidechain of the second subunit that are also required for nucleotide binding are shown in cyan. Parts of the peptides are clipped off for clarity. A, In DGC domain, the nucleotide is modelled to bind with the phosphate groups alongside the β-hairpin via a Mg^{2+} ion to E371. The sidechains of E370 and E371 are also modelled to adopt a plausible conformation that favours catalysis and nucleotide phosphate binding respectively. The arrows from the ribosyl 3' oxygen to the α-phosphate symbolise the intermolecular attacks. B and C, In AC and DNAP, the nucleotide binds across the β-hairpin with the phosphate groups resting on the P-loop. In DNAP, the template and the replicating DNA strands are shown in thin sticks.

for acting directly as a general base, whereas K332, a conserved residue in the DGC domain family, would stabilise the charge of the developing penta-coordinated transition state and the leaving pyrophosphate (Figure 4.3C, D). Water molecule(s) might be involved in the coordination with Mg^{2+} ion throughout the process, and might assist in deprotonating the catalytic E370 in the final step to reset the whole process.

The proposed catalytic mechanism in DGC domain makes use of the single catalytic acidic residue of E370 (Figure 4.2A). The suggested role of the modelled Mg^{2+} ion is to mediate the binding of the nucleotide phosphate groups to the other acidic residue of E371, which is conserved in the DGC domain family. This mechanism is significantly different from the two-metal ion catalysis in AC and DNAP, in which two invariant aspartates are used with the assistance of a bound Mg^{2+} ion in the deprotonation of the ribosyl 3' hydroxyl group [61]. In AC, the two conserved aspartates are D396 from the β-hairpin and D440 from the strand β1 (Figure 4.2B). Although PleD also exhibits two invariant acidic residues in very similar positions, namely E370 from the β-hairpin and D327 from the strand β1, there is no indication of a bound metal ion in the close vicinity. It is also unlikely that a Mg^{2+} ion coordinated by these residues would be able to reach the ribosyl 3' hydroxyl group.

The proposed mode of substrate binding in the DGC domain and the accompanying proposed catalytic mechanism favour the PleD reaction, in which two identical substrates are arranged in an anti-parallel manner to form a symmetrical cyclic di-nucleotide. The mechanism involved to complete this bimolecular reaction could be significantly different from that to complete a reaction such as the AC reaction, that uses only one substrate molecule to form a cyclic mono-nucleotide.

Allosteric product inhibition of DGC domain in PleD

The tight binding of c-diGMP at the I-site (Figure 3.46) offers the explanation for the observed co-purification of c-diGMP during protein preparation (Figure 3.6 and 3.7). In fact, kinetic experiments performed in the group of Urs Jenal revealed that PleD was significantly inhibited by the product c-diGMP, with an inhibition constant Ki of 0.5 μM (Figure 4.4). This value is lower than the estimated cellular concentration of c-diGMP [52] by about an order of magnitude. The product inhibition was also shown to be independent on the substrate concentration, suggesting that the inhibition was

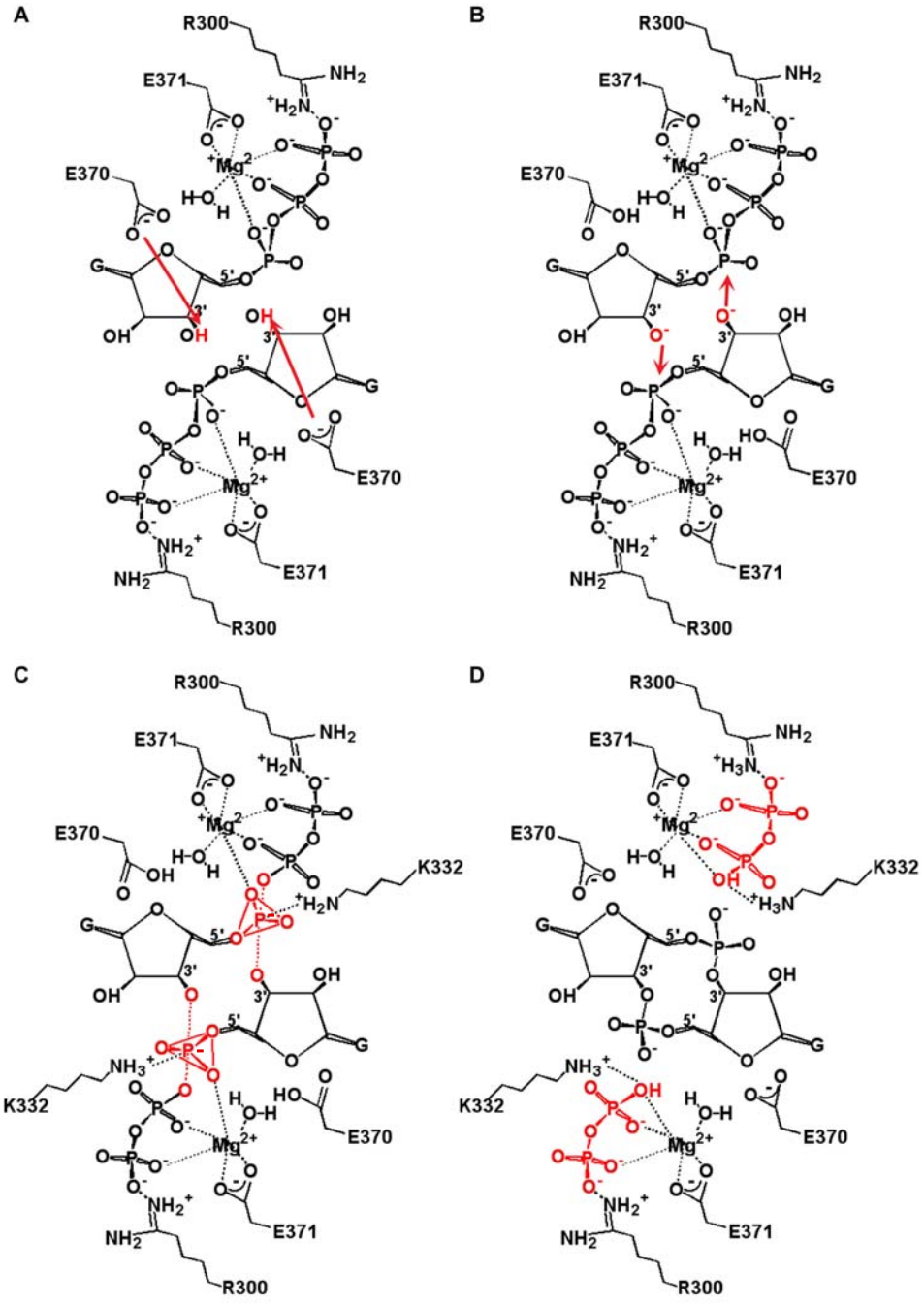


Figure 4.3: Proposed DGC catalytic mechanism. A Mg^{2+} ion and a water molecule are included in the model. A, The two GTP substrates align themselves in a symmetrical arrangement at the two-fold symmetrical active site formed by two DGC domains. The carboxyl group of E371 deprotonates the ribosyl 3' hydroxyl group. B, The ribosyl 3' oxygen initiates an intermolecular nucleophilic attack onto the α-phosphate of the other GTP molecule. C, α-phosphate is in the penta-coordinated state which is stabilised by K332. D, Pyrophosphates leave. c-diGMP is formed. A water molecule might deprotonate E370 to reset the whole process.

non-competitive. This finding suggests an allosteric regulation of the DGC activity in the PleD protein.

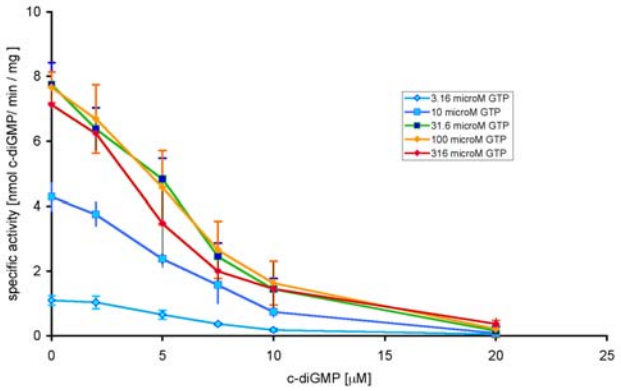


Figure 4.4: Product inhibition of PleD at varying GTP concentrations. Roughly 5 μM of c-diGMP concentration is required to attain 50 % inhibition at all tested GTP concentrations. Considering the enzyme concentration is 5 μM and the molar ratio of PleD to I-site bound c-diGMP is 1:2, the Ki is estimated to be 0.5 μM, assuming c-diGMP is a dimer in solution. The kinetic data were measured by Ralf Paul, Uni. Basel, in an assay using thin layer chromatography [45].

Mechanism of PleD regulation

PleD catalyses a condensation reaction of two identical substrates to yield a symmetrical product molecule. The crystal structure of PleD suggests that each PleD monomer forms a binding site for one substrate molecule. Although the non-activated PleD is monomeric in solution as determined by analytical ultracentrifugation and gel filtration (Figure 3.3) dimerisation of PleD appears necessary to promote effective active site formation by raising the local concentration of the DGC domains. Based on this reasoning, a simple mechanistic model of activation and product inhibition is proposed (Figure 4.5).

Phosphorylation at Asp53 of D1 would induce rotameric switches in Thr83 and Phe102 that would affect the C-terminal region of D1 domain, especially helix α4 and the α4-β5 loop, as in the studied examples of phosphorylated response regulator receiver domains [4, 35]. Since the residues in D1 domain that contribute to the salt bridges at the D1-D2 domain interface are also located in this region, a conformational change here might cause a

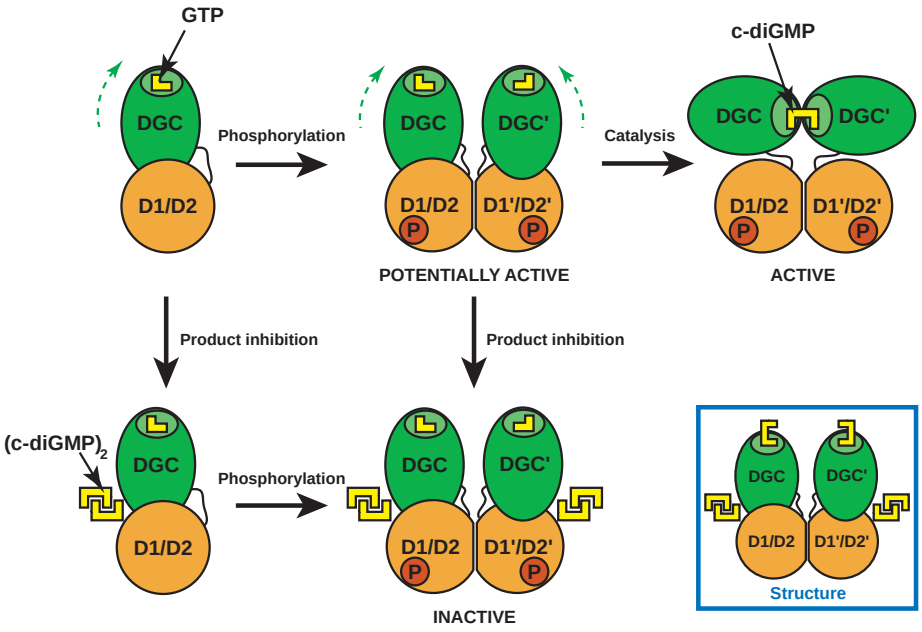


Figure 4.5: Mechanistic model of PleD regulation. The catalytic DGC domain (green) is tethered via a flexible linker to the D1-D2 stem (orange). It is postulated to be mobile with respect to the stem as indicated by the curved broken arrow. Top row: PleD is activated by phosphorylation at the D1 domain, which induces dimerisation mediated by the D1-D2 stem and allows the two substrate binding sites, bound with GTP substrates (yellow), to approach each other to undergo c-diGMP synthesis. Bottom row: Allosteric product inhibition occurs by the binding of the intercalated c-diGMP molecules to the I-site at the stem-DGC domain interface. DGC domain is immobilised with respect to the stem and barred from approaching its counterpart in the dimer. The schematic diagram of the crystal structure is shown in a box at the bottom, right hand corner for reference.

reshuffling of the salt bridges and a reorientation of the D1 and D2 domains (Figure 3.39,3.42). In effect, a reorientation of the domains might create an optimal interface that favours PleD dimerisation through isologous D1-D2 domain interactions.

The dimeric structure of non-activated PleD in the crystal might resemble that of an active dimer, though with a suboptimal interface. This view is supported by the lack of activity of the N-terminal His-tagged PleD protein (data not shown) that was probably due to the interference of the His-tag with the dimer interface (Figure 3.42). The basal activity observed for PleD protein *in vitro* [45] might actually be due to the presence of a small amount of non-optimal dimers.

Considering the small D2-DGC domain interface (Table 4.2), and assuming flexibility in the D2-DGC domain linker, a 'closed' conformation of the PleD dimer is conceivable (Figure 4.6) in which the two DGC domains within the dimer can associate to form the two-fold symmetrical active site as observed in the crystal packing.

Table 4.2: Solvent accessible area per monomer buried as a result of molecular assembly at different interfaces. Solvent accessible area of the molecules is calculated using a 1.4-Å probe rolling around the van der Waals surface of the molecule in the programme NACCESS [25, 34]. All disordered linkers were excluded from the structures for surface calculation.

Molecule / Domain 1	Molecule / Domain 2	Buried surface area($Å^2$)
D1	D2	1038.2
D2	DGC	686.6
Chain A	Chain B	815.7
A-site ligand	DGC	248.4
s S-site ligands	Monomer	521.7
S-site ligands	D2	192.8
S-site ligands	DGC	383.9

However, product binding to the I-site would stabilise the D2-DGC domain interface and immobilise the DGC domain relative to the D2 domain, thereby preventing the formation of the DGC active site.

The full-length structure of PleD shows the unusual domain architecture in this multidomain response regulator. The presence of a non-receiver D2 domain serves to assist in both the activation and the allosteric regulation of the enzyme.

The tight feedback regulation of PleD by product inhibition may indicate

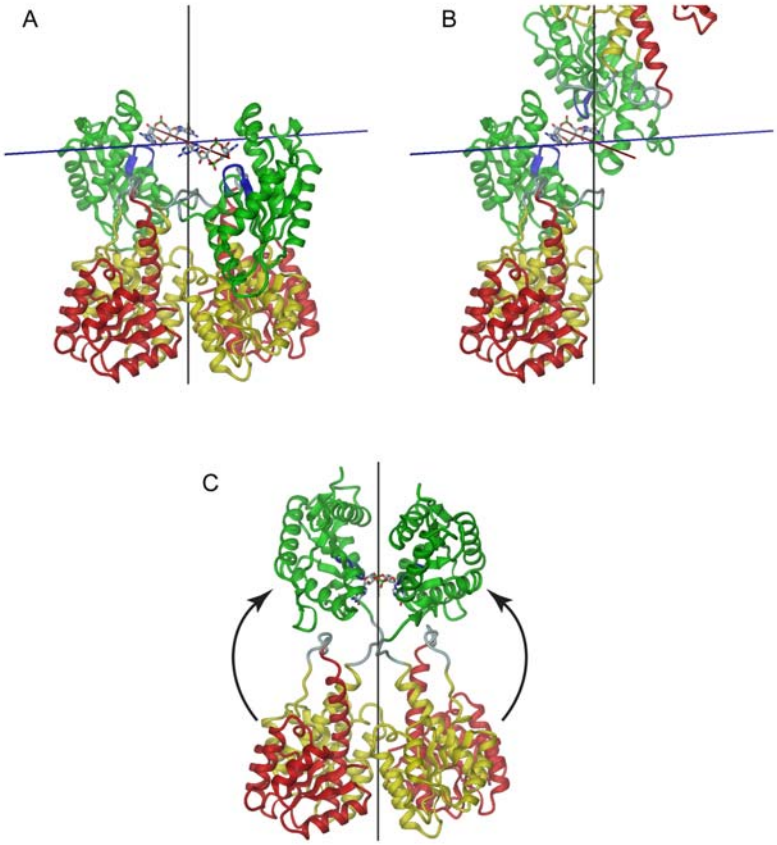


Figure 4.6: Proposed open and closed conformations of PleD. NCS axis is coloured in black, crystallographic axis in blue, and the local axis holding the A-site in two-fold symmetry is in red. All pictures here are shown from the same perspective. A, The crystal structure shows the PleD dimer in the 'open' conformation. The two c-diGMP molecules 'half-bound' to each PleD monomer are shown. B, Crystal packing shows that two DGC domains can arrange themselves to form a product binding site. Part of the upper PleD monomer is clipped off to maintain the same view and same molecule size as in panel A. C, To model the observed DGC-product-DGC complex in the context of a dimer requires bringing the observed complex to an orientation such that the local axis it obeys will coincide with the vertical NCS axis. The result is the modelled 'closed' conformation of PleD. Here we assume that the DGC domain is free to move with respect to the D2 domain.

the importance of imposing an upper limit on the cellular concentration of the secondary messenger c-diGMP. It remains to be shown that inhibition indeed constitutes, as predicted by the model, an overriding principle, i.e. that inhibition is independent on the phosphorylation state of the enzyme.

Chapter 5

Conclusions

Our crystal structure of the multidomain response regulator PleD protein in complex with its product at 2.7 Å has shed light onto the catalytic mechanism of the enzyme and its possible regulation. PleD contains three domains, D1, D2 and DGC. While the N-terminal D1 is a CheY-like receiver domain, the D2 domain is only structurally similar to CheY. The DGC domain is shown to share the fold of the AC catalytic domain. The GGEEF sequence motif is found at the β2-β3 hairpin that is conserved among all DGC homologues and is involved in nucleotide binding. It is also predicted that it participates in the DGC activity. However, the modelled substrate binding mode in DGC is significantly different from that of AC, suggesting a different catalyic mechanism.

Useful clues regarding the activation and inhibition of PleD have been obtained from our crystal structure. Two DGC domains are required to arrange symmetrically in order to form the catalytic active site. Thus, PleD is activated by dimerisation to release the putative secondary messenger molecule. The unexpected binding of the product molecules at the D2-DGC domain interface has enabled us to postulate an allosteric inhibitory mechanism, which is manifested by the product molecules in the immobilisation of DGC domain to the dimer base. This allosteric control provides a second layer of regulation of the enzymatic activity in this multidomain response regulator.

Chapter 6

Perspectives

Since the A-sites are formed at the crystal contacts by two crosslinked, symmetry related PleD dimers, the present crystal structure of PleD does not represent any of the states we have postulated in the mechanistic model (Figure 4.5). In order to prove this model, further studies are required to show:

- the active dimer with the optimal dimer interface
- the binding of the GTP substrate to DGC domain
- the active site within the PleD dimer
- the flexibility of DGC domain

The first goal can be achieved by obtaining the structure of activated PleD. Activation can be achieved using BeF_3^- modification or phosphorylation with acetophosphate as it has been achieved in other response regulator receiver domains [4, 29, 35, 38]. While a crystal structure of the activated PleD showing an optimal dimer interface would provide direct proof to the activation model of PleD, a similar structure but with product molecules bound at the I-site would lend support to an overruling role of allosteric control over phosphorylation.

Some may argue against our model of DGC catalytic mechanism, especially regarding the invariant D327 that is required for catalysis in AC and DNAP but not in DGC. Although we have shown that nucleotide binding to the usual location at the conserved P-loop is unlikely, our enzyme-product structure, however, does not give direct evidence as to how the phosphate

groups of the GTP substrate bind to the DGC domain. A complex structure of the protein with substrate or substrate analogue would clarify this ambiguity. Alternatively, a complex structure using a D327 mutant complemented with the activity assay would serve the same goal.

Our model of product inhibition through the DGC domain immobilisation is based on the assumption that the DGC domain is flexible with respect to the D1 / D2 dimer base. To test the influence of c-diGMP binding on DGC domain flexibility, the techniques of small angle X-ray scattering (SAXS) and NMR could be performed on the protein with and without allosterically bound product. We expect DGC domain attains a well-defined position, as in the crystal, when the product molecules are bound and carries out large movement when it is free of product molecules. The conformational change is estimated to be on the scale of tens of Å, given the maximum length of the product-bound protein of 80 Å and the length of the D1-D2 domains of 50 Å. It is known that conformational changes in the range of 10 Å are detectable by the SAXS technique [43]. On the other hand, when the conformational changes are studied using NMR, the large movement of the DGC domain would lead to slower relaxation compared to faster relaxation of the rest of the protein. We expect a broadening of the peaks that refer to the DGC domain. In both the SAXS and NMR experiments, we hope to determine the inhibition constant by titrating the protein with product molecules until the DGC domain becomes immobilised.

Among the few multidomain response regulators whose full-length structures are known [2, 12, 51], PleD is the only one that contains an additional CheY-like [35, 58] receiver domain, that is placed between the receiver domain D1 and the effector domain DGC. Such an unusual domain arrangement can now be explained by the involvement of the D2 domain in both the activation and the allosteric control of the protein. PleD employs an extra level of regulation, compared to many response regulators which negatively regulate the activity of the effector domain only by autodephosphorylation. Owing to the lack of structures of both active and inactive multidomain response regulator, the mechanism of the relay of phosphorylation signal from the receiver domain to an output signal from the effector domain still remains unclear. Biochemical data may provide clues to the problem but may not provide explanations at the molecular level. To unravel the mystery of the domain communication in the family of response regulators and to test our mechanistic model of activation and allosteric control, it is imminent to obtain the crystal structure of the activated PleD.

Appendix A

Data deposition

The coordinates of the PleD crystal structure are deposited in the PDB under the accession code 1W25.

Appendix B

Publications

1. **C. Chan**, R. Paul, D. Samoray, N.Amiot, B. Biese, U. Jenal and T. Schirmer
   Structural basis of activity and allosteric control of diguanylate cyclase. Accepted for publication in *Proc Natl Acad Sci USA*

2. R.Paul, S. Weiser, N. C. Amiot, **C. Chan**, T. Schirmer, B. Giese and U. Jenal
   Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel diguanylate cyclase output domain. *Genes Dev*, 18(6):715-27, 2004

Structural basis of activity and allosteric control of diguanylate cyclase

Carmen Chan[*], Ralf Paul[†], Dietrich Samoray[*], Nicolas C. Amiot[⊥], Bernd Giese[⊥], Urs Jenal[†] & Tilman Schirmer[*,§]

[*]Division of Structural Biology and [†]Division of Molecular Microbiology, Biozentrum, University of Basel, Klingelbergstr. 70, 4056 Basel, Switzerland

[⊥]Department of Chemistry, University of Basel, St. Johanns-Ring 19, 4056 Basel, Switzerland

Abbreviations: c-diGMP, bis-(3'→5')-cyclic di-guanosine monophosphate; DGC, diguanylate cyclase; RMSD, rms deviation; RRR, response regulator receiver; AC, adenylate cyclase

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 1W25.

[§] To whom correspondence should be addressed. E-mail: tilman.schirmer@unibas.ch.

Recent discoveries suggest that a novel second messenger, bis-(3'→5')-cyclic di-guanosine monophosphate (c-diGMP), is extensively used by bacteria to control multicellular behavior. Condensation of two GTP to the dinucleotide is catalyzed by the widely distributed diguanylate cyclase (DGC or GGDEF) domain that occurs in various combinations with sensory and/or regulatory modules.

The crystal structure of the unorthodox response regulator PleD from *Caulobacter crescentus*, which consists of two CheY-like receiver domains and a DGC domain has been solved in complex with the product c-diGMP. PleD forms a dimer with the CheY-like domains (the stem) mediating weak monomer - monomer interactions.

The fold of the DGC domain is similar to adenylate cyclase, but the nucleotide binding mode is substantially different. The guanine base is H-bonded to Asn335 and Asp344, whereas the ribosyl and α-phosphate moieties are extending over the β2-β3 - hairpin that carries the GGEEF signature motif. In the crystal, c-diGMP molecules are cross-linking active sites of adjacent dimers. It is inferred that, in solution, the two DGC domains of a dimer align in a 2-fold symmetric way to catalyze c-diGMP synthesis.

Two mutually intercalated c-diGMP molecules are found tightly bound at the stem - DGC interface. This allosteric site explains the observed non-competitive product inhibition. We propose that product inhibition is due to domain immobilization and sets an upper limit for the concentration of the second messenger in the cell.

Cyclic nucleotides like cAMP or cGMP have been recognized as important low-molecular-weight signaling molecules. While bacterial pathogens can interfere with cGMP signaling of their eukaryotic host cells (1), prokaryotes in general do not seem to use cGMP for signaling. In contrast, the cyclic bis-3'→5'-dinucleotide c-diGMP has been shown to regulate cell surface associated traits and community behavior like biofilm formation in a number of bacterial species (2-5). The general importance of c-diGMP is underscored by the omnipresence of the DGCdomain (hitherto named GGDEF or DUF1) in bacterial genomes, where it occurs in various combinations with other sensory and/or regulatory modules (6, 7). Despite the wide distribution and obvious regulatory relevance of DGC proteins, structural and functional information about this class of regulators is largely missing.

In *Caulobacter crescentus*, pole remodeling during development is regulated by several polar sensor histidine kinases, which control the diguanylate cyclase activity of the response regulator PleD (8, 9). PleD is composed of a CheY-like receiver domain (D1), a CheY-like adapter domain (D2) and a DGC domain. Upon phosphorylation of D1, activated PleD sequesters to the differentiating pole, where it catalyzes the conversion of two molecules of GTP to c-diGMP (10). Thus, by coupling of activity and subcellular localization, the PleD readout is controlled in time and space. The conformational changes invoked by phosphorylation of response regulator receiver (RRR) domains have been studied thoroughly on CheY (11) and other single domain receiver proteins (for a review, see (12)). Little, however, is known about how these alterations are transmitted to the effector protein or domain and only few intact multidomain response regulators have been analyzed structurally. Distinct mechanisms of activation, such as relief of active site obstruction (CheB (11, 13)), dimerization (DrrB, DrrD (14)), or both (NarL (15)), have been proposed.

Here, we report the structure of full-length PleD that allows drawing conclusions regarding its mechanism of activation and product inhibition. Also, by providing a detailed view on the enzyme - product interactions, our study offers insight into the catalytic mechanism of this widespread family of regulatory proteins.

Material and Methods

Overexpression and purification. Full-length PleD with a C-terminal His_6 tag was overexpressed in *E. coli* strain BL21 (DE3) pLysS. Purification was performed on a $NiSO_4$-charged HiTrap chelating HP column (Amersham Biosciences, Germany) with elution at approximately 200 mM imidazole. The pooled fractions were dialyzed overnight in 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 1 mM DTT at 4 ºC. The protein solution was concentrated and after clarification loaded on a Superdex 200 HR 10/30 column(Amersham Biosciences) pre-equilibrated with the same buffer. PleD appeared as monomer and fractions contributing to the peak were pooled. Selenomethionine-substituted PleD was expressed using the metabolic inhibition pathway. The purification procedure was the same as for native PleD.

Enzymatic assays. The assay for the measurement of the initial velocity of c-diGMP has been described previously (10). In short, GTP containing [α-^{32}P]GTP (Amersham Biosciences; 5 nCi/μL) was added to 50 µL of a solution containing 12.5 µg PleD in 50 mM Tris-HCl pH 7.8, 100 mM NaCl, 10 mM $MgCl_2$, 5 mM β-mercaptoethanol. Aliquots were taken at regular time intervals and the product was assayed by thin layer chromatography. For the product inhibition measurements, PleD was pre-incubated with c-diGMP at 25ºC for 5 min. Care was taken to remove any bound c-diGMP from PleD carried over from the purification by extensive dialysis.

Diguanylate cyclase activity was also assayed indirectly by monitoring the production of PPi using an enzyme coupled spectrophotometric assay (16). The reaction mixture contained PleD in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, 100 μM GTP and pyrophosphatase (500 mU/ml). The reaction was stopped by transferring 100 μL reaction mixture into the phosphate assay reagent (pH ≈ 0; 1 mL final volume) containing molybdate and malachite-green. The phosphate content in commercial GTP products had to be reduced from ca. 5 % to less than 0.5 % by anion exchange chromatography to reduce background absoprtion (17).

Crystallography. Crystals were obtained at room temperature by hanging drop vapor diffusion. For this, PleD at a nominal concentration of 200 μM = 10 mg/mL (assuming an ε_{280} of 9200 $M^{-1}cm^{-1}$) in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM DTT, 2 mM $MgCl_2$ and 0.8 mM

synthetic(10) c-diGMP was mixed with the reservoir (1.0 M glycine pH 9.2, 2 % dioxane, 14.5 % polyethylene glycol 20 000) at a ratio of 1:1. Selenomethione-substituted crystals were obtained in the same manner, but using a reservoir solution containing 1.0 M TAPS pH 9.0, 2 % dioxane, 11 % polyethylene glycol 20000. Diffraction data were collected from a single native (about 15 μm in diameter) and a single Se-Met substituted crystal (60 μm) at the Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland, and were processed with MOSFLM/SCALA (18). 18 selenium positions were identified using SHELXD (19). Phase refinement was performed by SHARP/SOLOMON (20) and was followed by 2-fold averaging and phase extension using DM (21). The model was built at the interactive graphics and refined by using REFMAC (18) imposing strict NCS constraints for the two copies in the asymmetric unit except for residues 117, 164, 168, and 404. The entire main-chain is defined by electron density except for residues 137 - 146, 282 - 288, and the C-terminal His-tag.

Results and Discussion

Crystal structure. The crystal structure of non-phosphorylated PleD has been solved by MAD phasing on Se-Met substituted protein to 2.7 Å (Table 1). The structure shows a linear arrangement of three structural domains (D1, D2, DGC; Fig. 1a) that are connected by single, disordered loops. As anticipated, domains D1 and D2 show the typical RRR fold with $(\beta\alpha)_5$ topology. However, in both domains, the C-terminal helix α5 is considerably extended and protrudes from the globular domain, apparently enhancing the D1 - D2 contact. Both domains closely resemble each other (RMSD = 1.2 Å for 119 Cα). Quasi-isologous contacts are formed between helix α5 of one domain and α4, β5, α5 of the other domain. The C-terminal extensions of the α5 helices form mutual apolar contacts, while their N-terminal parts are involved in several ionic interactions with α4 and β5 (Fig. 1b).

Only D1 carries all the residues required for RRR activation (11). These are primarily aspartates D9, D10, and the phosphoacceptor D53 of the 'acidic pocket', but also residues K105, T83, and F102 that have been shown to undergo large structural changes upon activation (11). Strong density is bridging the three carboxylates of the ‘acidic pocket’ and main-chain carbonyl 55. In line with its coordination geometry, this has been modeled as a Mg^{++} ion. Comparison of D1 with

other known RRR structures confirms that D1 is in the non-activated conformation with the β4-α4 loop in an "outward" orientation. The structural changes accompanying activation in RRRs are known (22). In Fig. 1b, (non-activated) PleD and activated FixJ (23) are superimposed. From this, one would predict that the largest conformational change upon PleD activation would occur in the β4-α4 loop with a concomitant shift (1-2 Å) in α4 that participates in the D1 - D2 interaction. Another prominent change would be a rotameric switch of F102. Aromatic residues in this position have been shown to swing from a half-buried location into a pocket between α4 and the β-sheet that opens up upon activation (11, 23). Notably and similar to CheB (24), F102 is located at a central position of the interface (Fig. 1b). This argues for a repacking of the D1 - D2 interface upon phosphorylation. How this may tie in with a global model of PleD activation is discussed below.

The DGC (GGDEF) domain consists of a 5-stranded central β-sheet surrounded by helices (Fig. 1a). Discounting two additional short β-hairpins (β0, β0' and β3', β3'') and the N-terminal helix α0, the succession of secondary structure elements is (βααββαβαβ), for which the topology is identical and the arrangement closely similar to that of the catalytic core of adenylate cyclase (AC) (25) as well as to the 'palm' domain of DNA polymerases (26) (Fig. S1). Structural relatedness with AC has been proposed earlier by sequence threading (27). Obviously, the DGC domain is functionally related to AC and DNA polymerases in that it also catalyzes 3' - 5' phosphodiester formation. The GG(D/E)EF signature motif locates to the central β-hairpin (Fig. 1a) which, as in AC and DNA polymerases, constitutes part of the active site.

In the crystal, PleD monomers are arranged to local dimers (Fig. 1c). Isologous contacts are formed, with Y26 of D1 being tucked between helices α3 and α4 of D2 and the N-terminus contacting α3. The small contact area (2 x 425 $Å^2$) is consistent with the monomeric state of PleD in solution. Along the direction of their symmetry axis, dimers are associated head-to-head to form tetramers of 222 symmetry. Intriguingly, this interaction does not involve protein - protein contacts. Instead, as shown in Fig. 2a, c-diGMP is bound with its two guanine bases in a 2-fold symmetric fashion to the active site formed by the adjacent monomers. This cross-linking of adjacent dimers represents a serendipitous crystallization artifact, since it provides comprehensive information about the product - enzyme interactions that appears relevant also for

the situation in solution where the product would bridge the catalytic domains within a dimer (see below).

Active site and catalytic mechanism. Figure 2b shows that the guanine base is inserted in a pocket formed mainly by helices α1 and α2 of DGC. All of its polar atoms except O6 make specific interactions explaining the observed GTP *versus* ATP discrimination (10). The N335 side-chain forms H-bonds with N3 and N2, whereas N1 is recognized by the D344 carboxylate. Both residues are highly conserved amongst DGC domains. Furthermore, the base abuts to the apolar side-chains of L294, F331, and L347. The ribosyl and the α-phosphate are extending over the turn of the β2-β3 - hairpin with G369 being completely covered explaining glycine conservation at this position. The 2'-hydroxyl is H-bonded to N335 and a phosphate non-ester oxygen is in the vicinity of the main-chain amide 369.

On the basis of the observed complex structure, extrapolations regarding the binding mode of the GTP substrate to an individual DGC domain as well as the catalytic mechanism can be proposed (Fig. 2c). Considering the specific interactions with the protein, it is most likely that the guanyl base does not change its position during catalysis. Furthermore, we assume that the ribosyl and α–phosphate moieties largely keep their position In this case, the β- and γ-phosphates of the substrate would be easily accommodated close to G368 of the β-hairpin and probably be bound *via* a Mg^{2+} ion to E371 with R300 conferring further stabilization. For catalysis to occur two substrate-loaded DGC domains have to arrange themselves in a similar way like the inter-dimer cross-link in the crystal (Figs. 2b,c). This might be facilitated by charge complementarity between D336 and R300 (Fig. 2b). Subsequently, the 3'-OH group of the GTP substrate has to be deprotonated to allow an intermolecular nucleophilic attack onto the α-phosphate. E370 appears well poised for acting as a general base, whereas K332 would stabilize the charge of the developing pentacoordinated phosphoryl transition state and the pyrophosphate leaving group. All the residues proposed to be of functional importance are highly conserved amongst DGC sequences.

Intriguingly, the proposed nucleotide-binding mode is different from that observed in AC (Fig. 2d) and DNA polymerases. In both enzymes, the phosphate-binding site is provided by the short

P-loop between β1 and α1. In AC, the base is lying on the β-hairpin and is recognized by residues of the second subunit of the heterodimer, whereas in DNA polymerases the base adopts even another position, which allows base pairing with the template strand (28). Is it conceivable that DGC can bind the substrate in a similar way as AC? This appears unlikely, since there would be no interactions with the base conferring specificity and contributing to binding affinity. An alternative model with the base bound to the DGC guanine binding pocket and the terminal phosphates bound to the P-loop appears sterically unfeasible. Thus, there seems to be no reasonable alternative to the proposed model of the substrate - DGC complex which, nevertheless, has to be confirmed experimentally. In AC, a magnesium ion coordinated by two aspartates (D396, D440; Fig. 2d) has been proposed to assist in the nucleophilic attack of the 3'-OH group onto the α-phosphate (25). Conspicuously, the DGC domain exhibits two invariant acidic residues (D327, E370) at very similar positions, but there is no indication for a bound metal, which is in line with the disorder of E370. In fact, a magnesium ion coordinated by these residues would be too distant from the substrate to be effective. E370 is proposed to act as a general base (see above), but the role of D327, if there is any, remains to be elucidated.

Allosteric site and product inhibition. Unexpectedly, the crystal structure shows two product molecules bound to the D2/DGC interface (I-site; Figs. 1 and 3). The dimeric ligand is of compact, roundish shape with mutually intercalated purine bases. Each of the two central guanyl moieties forms an intermolecular H-bond with a phosphate group. A similar structure has been observed for several crystal forms of c-diGMP (29, 30) and may be present also in solution. The ligand is bound to both D2 and DGC by a multitude of specific interactions. Hereby the common base - arginine pairing motif involving O6 and N7 of the guanyl (31) occurs three times (with R390, R359 and R178; Fig. 3b). Tight binding is consistent with the observed co-purification of c-diGMP during protein preparation. Intriguingly, kinetic data reveal strong product inhibition, with a K_i (0.5 μM; Fig. S2) about an order of magnitude lower than estimated for the cellular concentration of c-diGMP (32). Clearly, inhibition is non-competitive (Fig. S2) , i.e. independent of substrate concentration, and can thus be attributed to an allosteric effect of I-site binding.

Mechanistic model of regulation. PleD catalyzes a condensation reaction of two identical substrates (that are bound to two identical DGC domains) to yield a C2-symmetric product

molecule. In solution, non-activated PleD appears to be monomeric as determined by analytical ultracentrifugation and gel filtration (not shown). For the enzyme to be efficient, dimerization appears necessary, since a reaction catalyzed by monomeric PleD would be severely limited by the macromolecular diffusion rate. Based on this reasoning, a simple mechanistic model of activation and product inhibition can be proposed (Fig. 4). Phosphorylation at D53 of D1 would induce repacking of the D1/D2 interface and reorientation of D1 with respect to D2. This in turn will enhance dimer formation mediated by isologous D1 - D2 contacts across the interface. In fact, the dimeric structure of non-activated PleD in the crystal may well resemble the active dimer, though with a suboptimal interface. This view is supported by the lack of activity of N-terminally His-tagged PleD (data not shown) which would be due to interference of the tag with the interface (Fig. 1c). The low constitutive activity observed *in vitro* (10) would thus be due to a small fraction of non-dissociated dimers. Consistent with this, the specific activity shows a strong dependence on the enzyme concentration (unpublished data).

The large distance between the substrate binding sites of the crystal dimer (Fig. 1c) would not permit catalysis. However, allowing for flexibility of the DGC domains relative to the stem (a realistic assumption considering the small D2/DGC interface area of 690 $Å^2$) a 'closed' dimer conformation competent for catalysis can be readily modeled (schematically shown in Fig. 4) in which a complete 2-fold symmetric active site is formed as observed for the DGC/DGC'' pair (Fig. 2b). In the context of this 'activation by dimerization' model the observed allosteric product inhibition can easily be explained (Fig. 4). Stabilization of the D2/DGC interface by product binding to the I-site would simply prevent encounter of the substrate binding sites ('inhibition by domain immobilization').

The tight feedback regulation of PleD by product inhibition probably demonstrates the importance of imposing an upper limit on the concentration of the second messenger c-diGMP. It remains to be shown that inhibition indeed constitutes, as predicted by the model, an overriding principle, i.e. that inhibition is independent of the phosphorylation state of the enzyme. The structure of DGC provide the first insight into the molecular interactions of c-diGMP with its possible partners. The quest for downstream targets of this messenger that appears to be ubiquitous in prokaryotes (6) has started.

Final note. We thank T. Bickle, Z. Markovic-Housley, O. Mayans, and C. Peneff for critical reading of the manuscript, J. Saito for crystallization optimization, and D.V. Palanivelu and the staff of beamline X06SA of the SLS synchrotron (Villigen, Switzerland) for assistance with data acquisition. The work was supported by Swiss National Science Foundation grant 31-59050.99 to U.J.

Supporting online information.

Figs. S1 and S2.

References

---References---

1. Uzzau, S. & Fasano, A. (2000) *Cell Microbiol* **2,** 83-9.
2. Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., Braun, S., De Vroom, E., Van der Marel, G. A., Van Boom, J. H. & Benziman, M. (1987) *Nature* **325,** 279-281.
3. Jenal, U. (2004) *Curr Opin Microbiol* **7,** 185-91.
4. Simm, R., Morr, M., Kader, A., Nimtz, M. & Romling, U. (2004) *Mol Microbiol* **53,** 1123-34.
5. Tischler, A. D. & Camilli, A. (2004) *Mol Microbiol* **53,** 857-869.
6. Galperin, M. Y. (2004) *Environ Microbiol* **6,** 552-67.
7. Galperin, M. Y., Nikolskaya, A. N. & Koonin, E. V. (2001) *FEMS Microbiol Lett* **203,** 11-21.
8. Hecht, G. B. & Newton, A. (1995) *J Bacteriol* **177,** 6223-9.
9. Aldridge, P., Paul, R., Goymer, P., Rainey, P. & Jenal, U. (2003) *Mol Microbiol* **47,** 1695-708.
10. Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B. & Jenal, U. (2004) *Genes Dev* **18,** 715-27.
11. Lee, S. Y., Cho, H. S., Pelton, J. G., Yan, D., Berry, E. A. & Wemmer, D. E. (2001) *J Biol Chem* **276,** 16425-31.
12. Stock, A. M., Robinson, V. L. & Goudreau, P. N. (2000) *Annu Rev Biochem* **69,** 183-215.
13. Djordjevic, S. & Stock, A. M. (1998) *J Struct Biol* **124,** 189-200.
14. Robinson, V. L., Wu, T. & Stock, A. M. (2003) *J Bacteriol* **185,** 4186-94.
15. Maris, A. E., Sawaya, M. R., Kaczor-Grzeskowiak, M., Jarvis, M. R., Bearson, S. M., Kopka, M. L., Schroder, I., Gunsalus, R. P. & Dickerson, R. E. (2002) *Nat Struct Biol* **9,** 771-8.
16. Baykov, A. A., Evtushenko, O. A. & Avaeva, S. M. (1988) *Anal Biochem* **171,** 266-70.

17. Birnbaumer, L., Torres, H. N., Flawia, M. M. & Fricke, R. F. (1979) *Anal Biochem* **93,** 124-33.
18. CCP4 (1994) *Acta. Cryst.* **D50,** 760-763.
19. Schneider, T. R. & Sheldrick, G. M. (2002) *Acta Crystallogr D Biol Crystallogr* **58,** 1772-9.
20. De La Fortelle, E. & Bricogne, G. (1997) *Methods Enzymol.* **276,** 472-493.
21. Cowtan, K. D. & Zhang, K. Y. (1999) *Prog Biophys Mol Biol* **72,** 245-70.
22. Robinson, V. L., Buckler, D. R. & Stock, A. M. (2000) *Nat Struct Biol* **7,** 626-33.
23. Birck, C., Mourey, L., Gouet, P., Fabry, B., Schumacher, J., Rousseau, P., Kahn, D. & Samama, J. P. (1999) *Structure Fold Des* **7,** 1505-15.
24. Djordjevic, S., Goudreau, P. N., Xu, Q., Stock, A. M. & West, A. H. (1998) *Proc Natl Acad Sci U S A* **95,** 1381-6.
25. Tesmer, J. J., Sunahara, R. K., Johnson, R. A., Gosselin, G., Gilman, A. G. & Sprang, S. R. (1999) *Science* **285,** 756-60.
26. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. (1998) *Nature* **391,** 251-8.
27. Pei, J. & Grishin, N. V. (2001) *Proteins* **42,** 210-6.
28. Tesmer, J. J. & Sprang, S. R. (1998) *Curr Opin Struct Biol* **8,** 713-9.
29. Liaw, Y. C., Gao, Y. G., Robinson, H., Sheldrick, G. M., Sliedregt, L. A., van der Marel, G. A., van Boom, J. H. & Wang, A. H. (1990) *FEBS Lett* **264,** 223-7.
30. Egli, M., Gessner, R. V., Williams, L. D., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Rich, A. & Frederick, C. A. (1990) *Proc Natl Acad Sci U S A* **87,** 3235-9.
31. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. (2001) *Nucleic Acids Res* **29,** 2860-74.
32. Ross, P., Mayer, R. & Benziman, M. (1991) *Microbiol Rev* **55,** 35-58.
33. Philippsen, A. (1998) *http://www.dino3d.org*.

Table 1. Crystallographic data

Data collection	**Native**	**SeMet**		
		Peak	**Inflection**	**Remote**
Wavelength (Å)	0.97950	0.97950	0.97970	0.97700
Resolution (Å)	2.7	3.2	3.0	3.0
Highest shell (Å)	2.85 – 2.70	3.27 – 3.20	3.16 – 3.00	3.16 – 3.00
Space group	$P4_22_12$			
a = b (Å)	135.9	134.9		
c (Å)	169.2	168.1		
Nr. of unique reflections	42707	26249	32080	32140
Redundancy	5.7 (5.5)	4.1 (3.9)	4.8 (4.9)	4.6 (3.6)
I / σ	7.2 (1.9)	6.2 (2.1)	7.2 (1.8)	7.3 (1.1)
Completeness (%)	99.3 (98.7)	99.9 (100)	99.9 (100)	99.6 (97.6)
Anom. completeness (%)	-	99.7	99.9	97.5
R_{merge} (%)	8.3 (36)	10.5 (33)	9.4 (39)	11.0 (63)

Phasing								
Resolution shell	13.12	7.69	5.98	5.06	4.47	3.72	3.46	3.16
Figure of merit	0.74	0.67	0.62	0.55	0.50	0.34	0.27	0.13
Mean figure of merit	0.39							

Refinement			
No of molecules in a.u.	2	Protein atoms	6883
Resolution (Å)	50.0 – 2.7	Water molecules	15
R / R_{free} (%)	21.0 / 23.9	Ligand atoms	233
RMSD		Residues in Ramachandran core (%)	92.8
bond lengths [a] (Å)	0.008		
bond angles [a] (°)	1.2		
Rms ΔB of bonded atoms (Å^2)		Residues in disallowed region (%)	0.0
main chain	1.13		
side chain	2.42	Average B (Å^2)	25.6

Values in parentheses are for the highest resolution shell.

[a] RMSD from ideal stereochemistry.

Legends to the Figures

Figure 1. Crystal structure of PleD. **a**, The monomer consists of three domains connected by disordered linker peptides (light grey). Domains D1 (residues 2-140, red) and D2 (141-285, yellow) show the CheY-like fold. D1 carries the phosphoacceptor D53. The catalytic DGC domain (286-454) is shown in green. The GGEEF signature motif is located on the β-hairpin (blue) and constitutes part of the active site (A-site) to which a c-diGMP molecule is bound. Two c-diGMP molecules are found at the D2/DGC interface (I-site). **b**, The D1(red)/D2(yellow) interface as viewed along the quasi-two-fold. Compared to the view in panel **a** the structure has been rotated by 90^0 approximately around the horizontal. Ionic residues in the interface and residues implicated in activation (phosphoacceptor D53, K105, T83, F102) are shown. The trace of the β4-α4 loop and F101 of phosphorylated FixJ (magenta; PDB code 1d5w (23)) is shown superimposed on D1. **c**, The two monomers of the asymmetric unit form a 2-fold dimer. The view is related by a -60° rotation about the dimer (vertical) axis with respect to the view in **a**. The c-diGMP molecules that are bound to sites A and A' cross-link to another dimer above (see Fig. 2a).

Figure 2. Ligand binding to the active site of PleD. **a**, Dimers are packed head-to-head to form 222 tetramers. The view is rotated by 135° about the vertical axis with respect to Fig. 1c. The dimers are held together by two c-diGMP ligands that are located on a local 2-fold axis of the tetramer (the viewing direction), the ligand molecule in the back has been omitted for clarity. **b**, Stereographic close-up view of **a**. The G_{368}GEEF signature motif comprises residues important for substrate binding (G368, G369 and E371) and catalysis (E370). The omit map of the ligand is contoured at 4 σ. The side-chains of E370, E371, R300, K332, and D336 are partly disordered. **c**, Complex of DGC with substrate GTP-Mg as modeled on the basis of the product complex shown in panel b. The positions of the guanine, ribose and α-phosphoryl moieties are the same as in the product complex structure. The upper substrate, which has been shifted arbitrarily by about 2 Å to the upper right, would be bound to another, 2-fold related DGC domain (not shown). The side-chain conformations of E370 and K332 have been adjusted to bring the functional groups into catalytically competent positions. The arrows indicate the nucleophilic attack of the 3'-oxygens on the α-phosphates. **d**, AC in complex with substrate analog ATP-α-S (light-blue, PDB code

1cjk (25)), view corresponding to that in panel c. The Cα-trace corresponds to that part of the α-chain which is structurally homologous to DGC. The β-chain, which provides specific interactions with the adenine base, has been omitted for clarity.

Figure 3. Product binding to the allosteric inhibitory I-site. **a**, A close-up view of the two mutually intercalated c-diGMP molecules (khaki and grey carbon atoms) bound at the D2 (yellow) - DGC (green) interface. The omit map of the ligand is contoured at 4 σ. **b**, The ligand is tightly bound to both domains (carbons are colored in magenta (D2) and cyan (DGC)) by a multitude of specific interactions including a recurrent arginine - guanine binding motif. Figures were generated by DINO (33).

Figure 4. Mechanistic model of PleD regulation. The catalytic DGC domain (green) is tethered *via* a flexible linker peptide to the D1/D2 stem. The DGC domain is postulated to be mobile with respect to the stem as indicated by the curved broken arrow. Top row: PleD is activated by phosphorylation at the D1 domain, which induces dimerization mediated by the stems and allows the two substrate binding sites (with bound GTP substrate in yellow) to approach each other and to the condensation reaction (2 GTP -> c-diGMP + 2 PPi) to occur. Bottom row: Allosteric product inhibition occurs by binding of $(c\text{-diGMP})_2$ to the I-site at the stem - DGC interface. Whereby, the DGC domain is immobilized with respect to the stem and barred from approaching its counterpart in the dimer.

Fig 1a

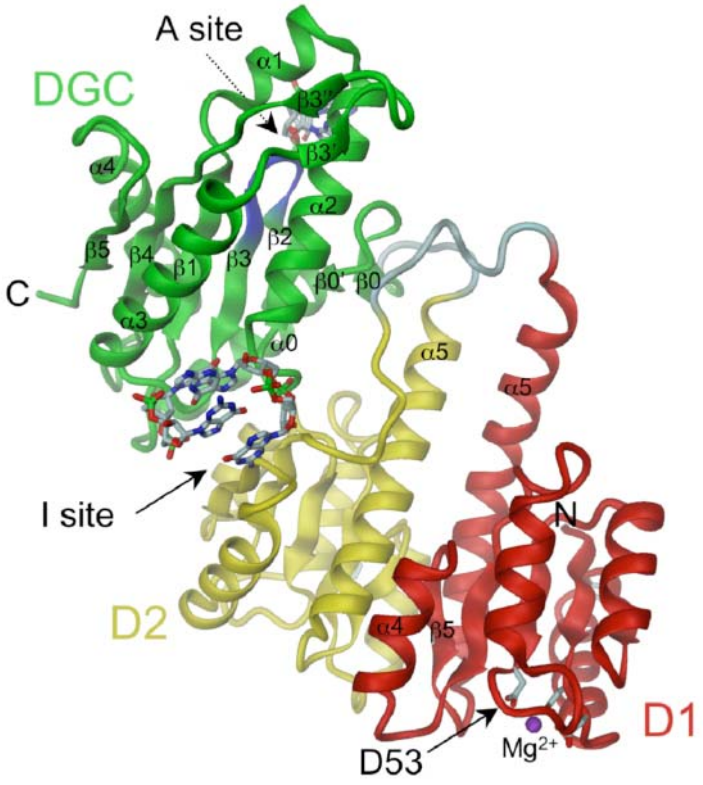

A site
DGC
α1
β3'
β3''
α4
α2
β5
β4
β1
β3
β2
β0'
β0
C
α3
α0
α5
α5
I site
N
D2
α4
β5
D1
D53
Mg2+

Fig 1b

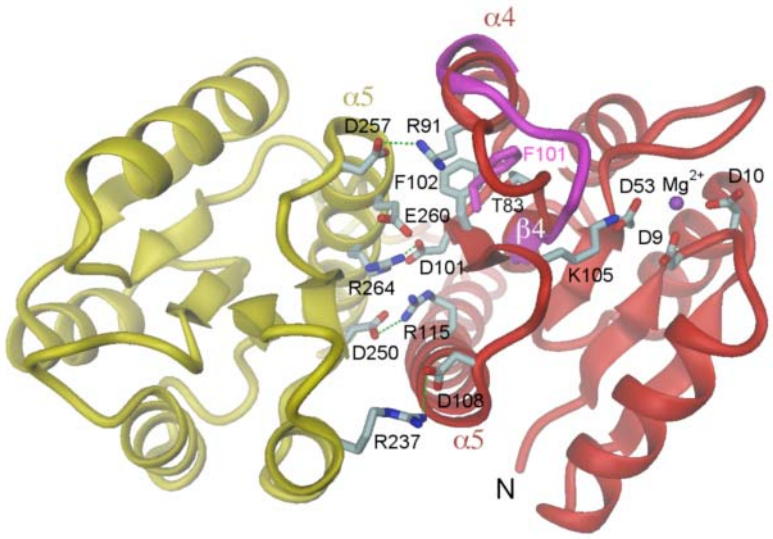

α4
α5
D257
R91
F101
F102
T83
E260
β4
D53
Mg2+
D10
D9
D101
K105
R264
R115
D250
D108
R237
α5
N

Fig 1c

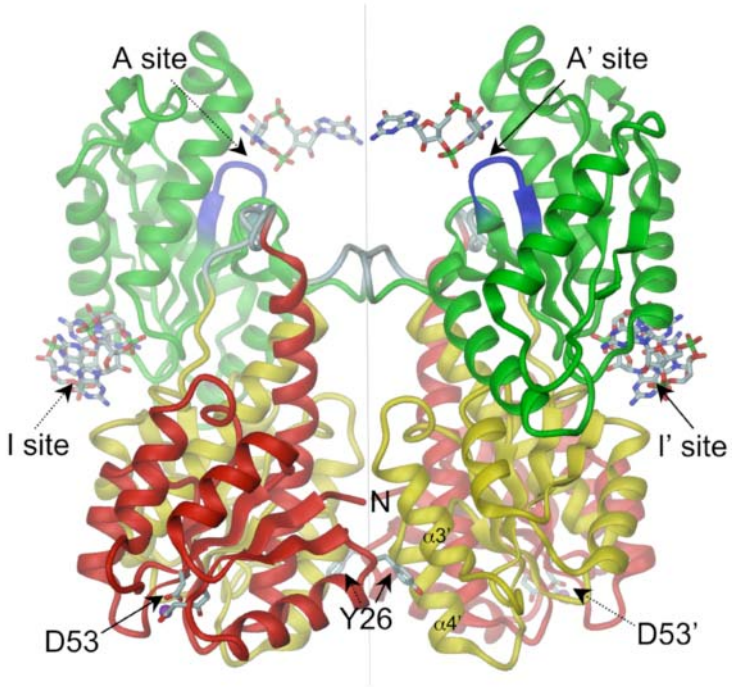

A site
A' site
I site
I' site
N
α3'
D53
Y26
α4'
D53'

Fig2a

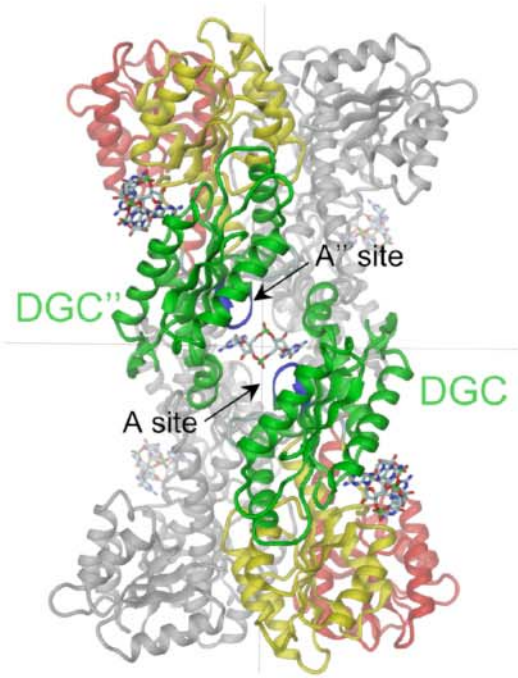


Fig2b

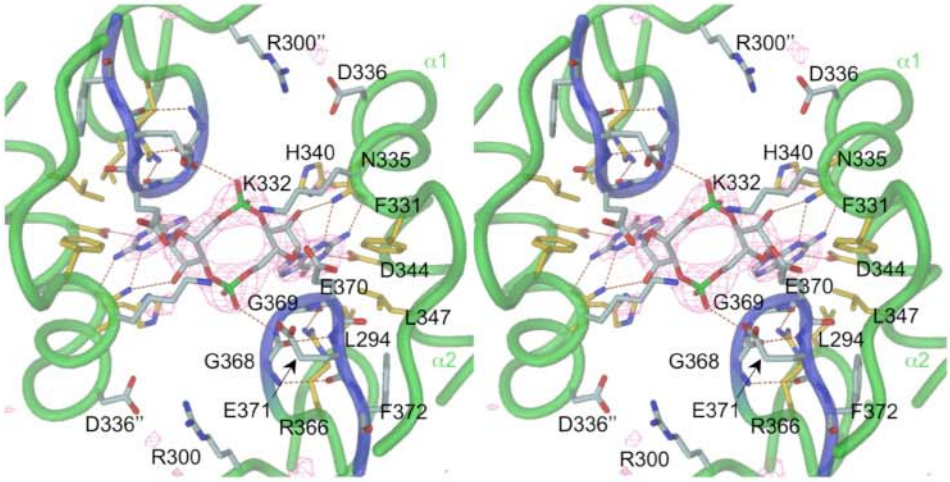


Fig2c

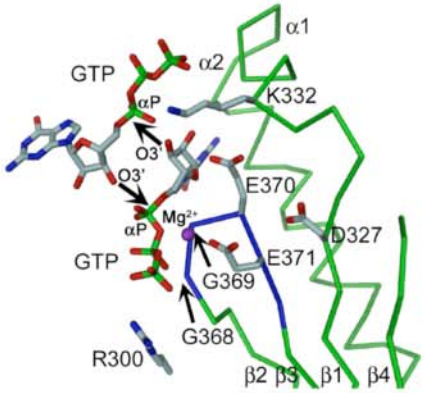


Fig2d

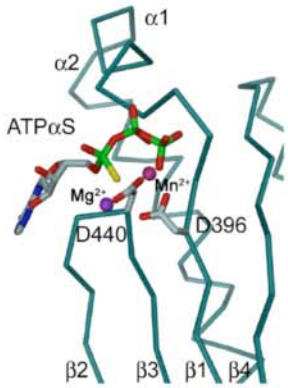


Fig3a

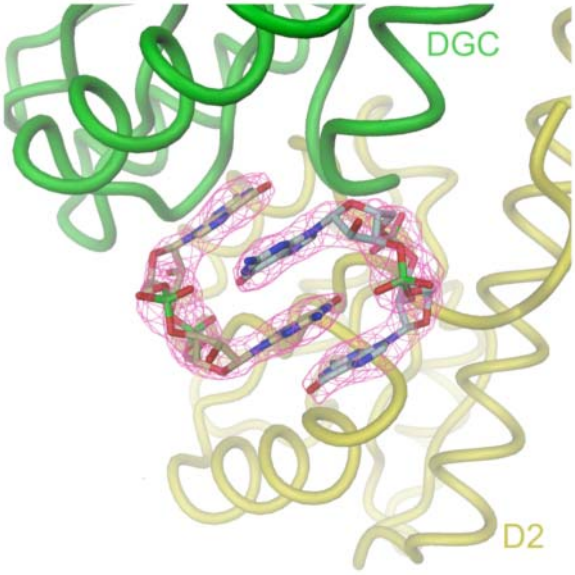


Fig3b

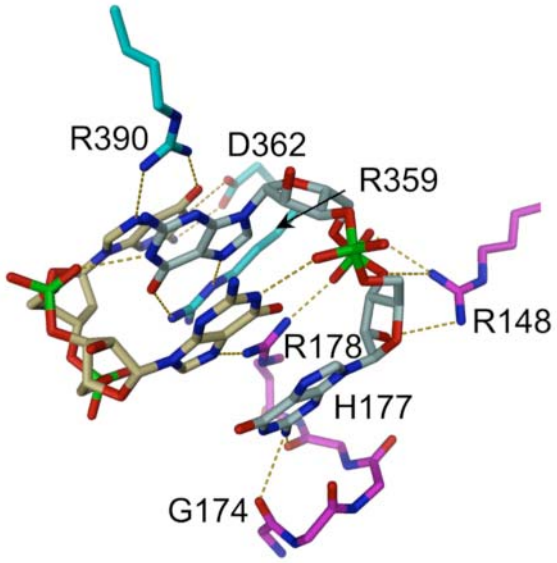
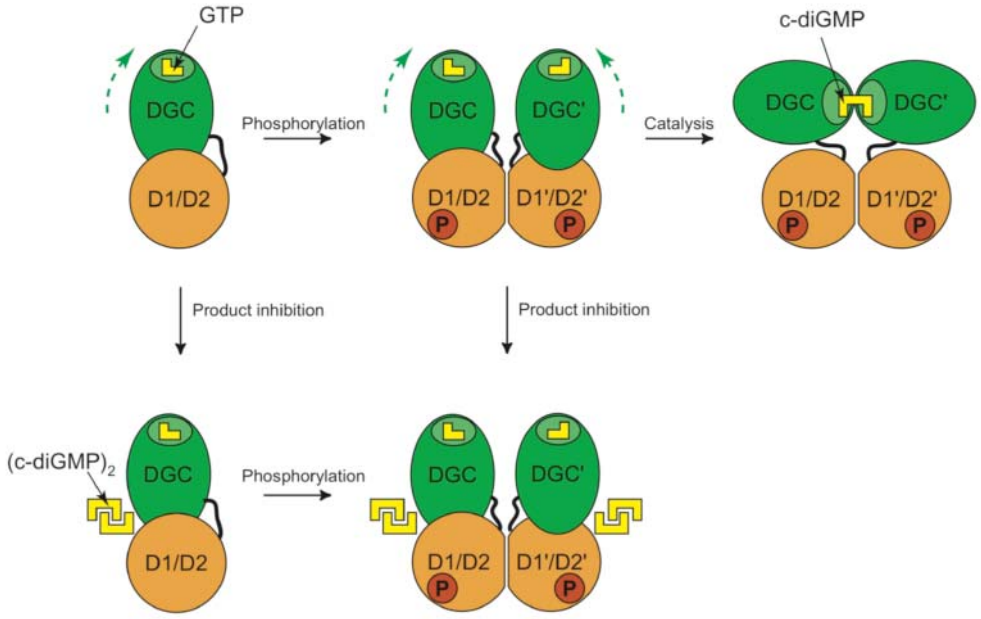


Fig. 4

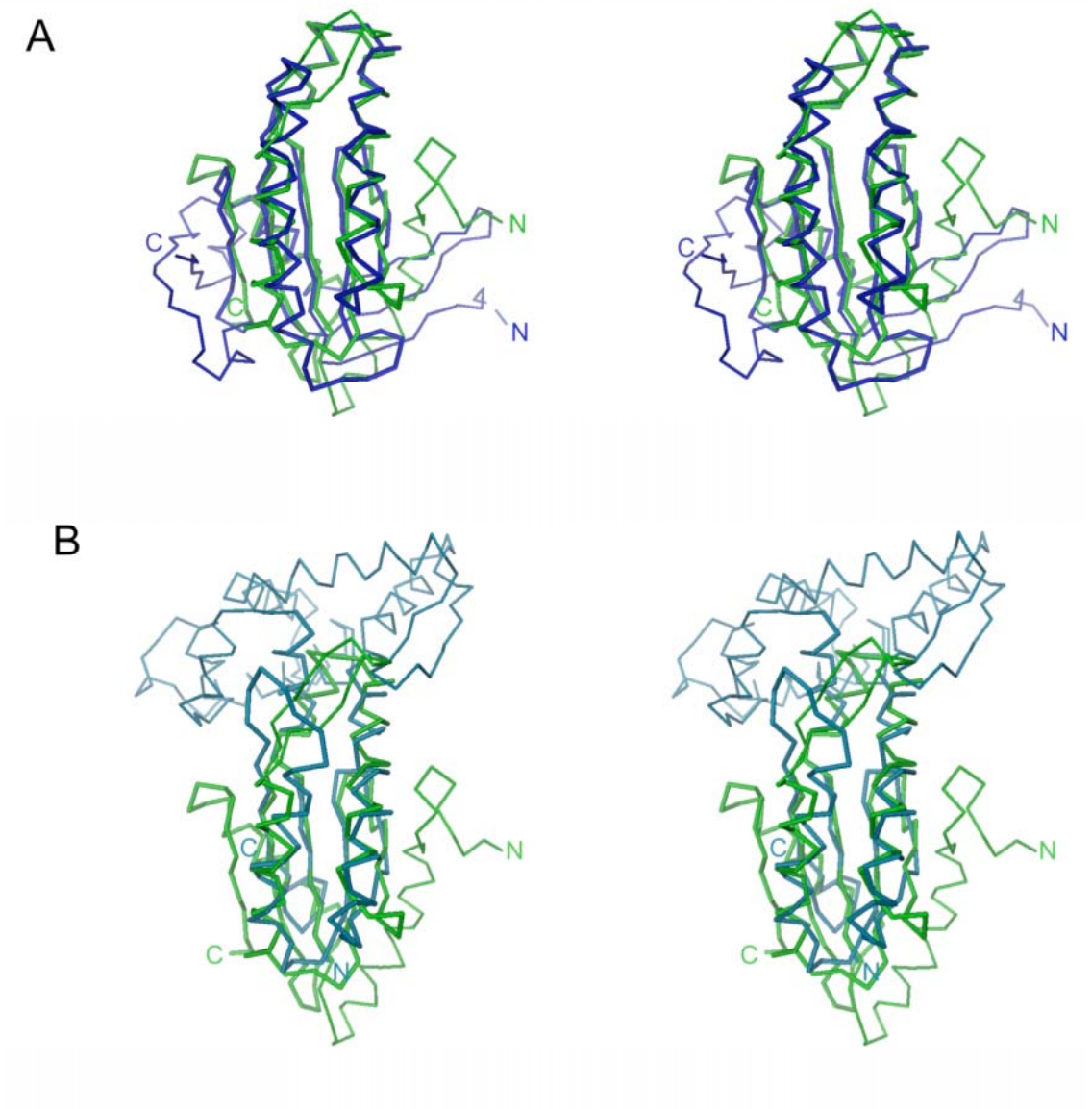

GTP
DGC
D1/D2
Phosphorylation
DGC
DGC'
D1/D2
D1'/D2'
P
P
Catalysis
c-diGMP
DGC
DGC'
D1/D2
D1'/D2'
P
P
Product inhibition
Product inhibition
(c-diGMP)2
DGC
D1/D2
Phosphorylation
DGC
DGC'
D1/D2
D1'/D2'
P
P

Supporting Online Material

Structural basis of activity and allosteric control of diguanylate cyclase

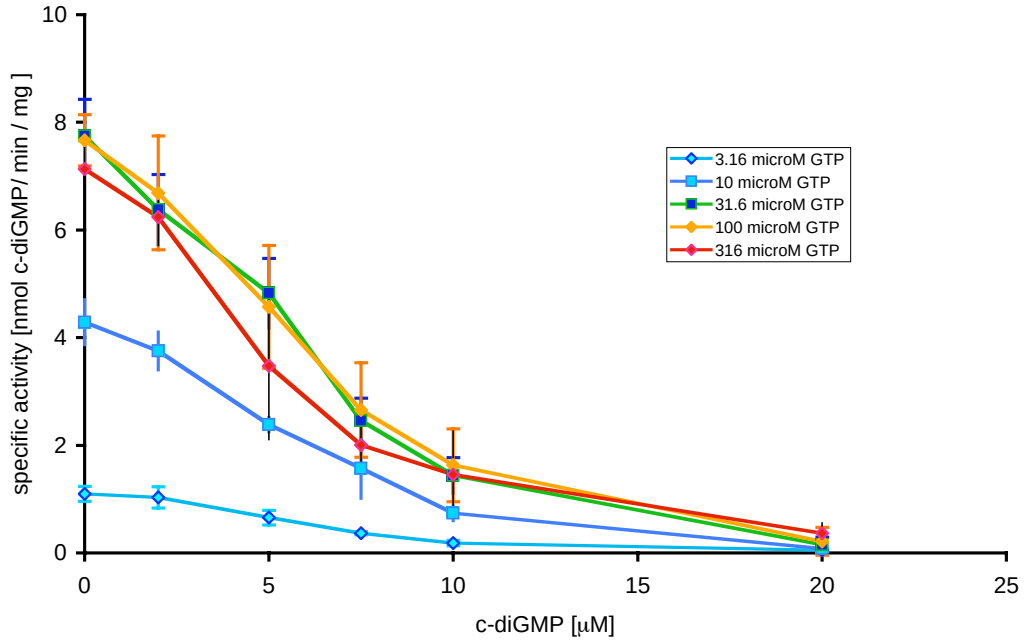
Carmen Chan, Ralf Paul, Dietrich Samoray, Nicolas C. Amiot, Bernd Giese, Urs Jenal & Tilman Schirmer

Figure S1.



Common core structure shared by the DGC domain, the catalytic domain of AC and the palm domain of DNA polymerase. **a**, Stereographic view of the superposition of DGC (green) on AC (blue, PDB code 1cjk). 92 Cα were superimposed with an rmsd of 1.5 Å. The common core is shown in a thicker line. **b**, Stereographic view of the superposition of DGC (green) on DNA polymerase I (light blue, PDB code 1nk4). 63 Cα were superimposed with an rmsd of 1.7 Å. The common core here is different from the that in panel **a** in that DNA polymerase I lacks a β5 strand.

Figure S2.



Product inhibition of PleD at varying GTP concentrations as determined by thin layer chromatography. Data refer to the initial velocity, i.e. they were acquired at time points before the reaction slowed down due to product inhibition. The values represent the averages of 4 independent experiments for 10, 31.6, 100 and 316 μM GTP, and 2 independent experiments for 3.16 μM GTP, respectively. For all substrate concentrations roughly the same c-diGMP concentration of 6 μM is required to attain 50% inhibition. Taking into account the enzyme concentration of 5 μM, a K_i of about 0.5 μM corresponding to the bimolecular association reaction: PleD + $(c\text{-diGMP})_2$ <-> PleD-$(c\text{-diGMP})_2$ can be obtained by numerical solution. This assumes that c-diGMP is present as a dimer in solution.

Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain

Ralf Paul,[1] Stefan Weiser,[1] Nicholas C. Amiot,[2] Carmen Chan,[3] Tilman Schirmer,[3] Bernd Giese,[2] and Urs Jenal[1,4]

[1]Division of Molecular Microbiology, [2]Department of Chemistry, and [3]Division of Structural Biology, Biozentrum, University of Basel, 4056 Basel, Switzerland

Pole development is coordinated with the *Caulobacter crescentus* cell cycle by two-component signaling proteins. We show that an unusual response regulator, PleD, is required for polar differentiation and is sequestered to the cell pole only when it is activated by phosphorylation. Dynamic localization of PleD to the cell pole provides a mechanism to temporally and spatially control the signaling output of PleD during development. Targeting of PleD to the cell pole is coupled to the activation of a C-terminal guanylate cyclase domain, which catalyzes the synthesis of cyclic di-guanosine monophosphate. We propose that the local action of this novel-type guanylate cyclase might constitute a general regulatory principle in bacterial growth and development.

[*Keywords*: Caulobacter development; GGDEF domain; protein localization; response regulator]

Received October 28, 2003; revised version accepted February 5, 2004.

During developmental transitions, localized changes of cellular morphology are mediated by adaptation in levels and arrangement of proteins. Temporal and spatial control often relies on the timed synthesis or activation of transcriptional regulators and on the establishment of gradients through the compartmentalization of signaling complexes. Although the regulatory mechanisms of gene expression are relatively well understood, it is often not clear how morphogenetic changes are controlled and coordinated locally. In prokaryotes, the major paradigm for signal transduction is the two-component regulatory system (Parkinson and Kofoid 1992). On signal input, the first component, a sensor kinase, autophosphorylates on a histidine residue. The second component, a soluble response regulator, often functions as a transcriptional regulator. Its phosphorylation by the cognate histidine kinase on a conserved aspartate residue in the N-terminal receiver domain usually results in increased DNA binding affinity (Parkinson and Kofoid 1992). Here we present evidence that a novel-type response regulator acts at a distinct subcellular site where it contributes to local changes in cell morphology through the production of a novel signaling molecule.

[4]Corresponding author.
E-MAIL urs.jenal@unibas.ch; FAX 41-61-267-2118.
Article and publication are at http://www.genesdev.org/cgi/doi/10.1101/gad.289504.

The unicellular bacterium *Caulobacter crescentus* goes through an obligate developmental transition that allows it to switch between a sessile, adhesive, and a motile, planktonic cell during its cell cycle. As a consequence, cell poles are continuously remodeled during cell differentiation to facilitate assembly and removal of motility and surface adherence organelles at the right time and in the correct order. Asymmetry is established in the predivisional cell with a single flagellum, a chemotaxis machinery, and pili being assembled at one pole, whereas the opposite pole consists of a stalk and an adhesive organelle, the holdfast (Fig. 1). As a result, division generates two cell types with distinct properties: a surface-attached stalked cell and a motile swarmer cell. The swarmer progeny first differentiates into a stalked cell before it initiates DNA replication and cell division. During this transition the pili retract, flagella are released, and the adhesive organelles are synthesized at the same pole. Here we investigate the function and regulation of the PleD response regulator in *C. crescentus* polar development. Cells that lack a functional PleD protein are hypermotile, are unable to eject the flagellum, and fail to synthesize a complete stalk structure (Hecht and Newton 1995; Aldridge and Jenal 1999). In contrast, the presence of a constitutively active mutant protein PleD* results in elongated stalks and has a dominant negative effect on motility (Aldridge et al. 2003). The PleD* protein contains four point mutations (Asn/Thr 120, Ala/

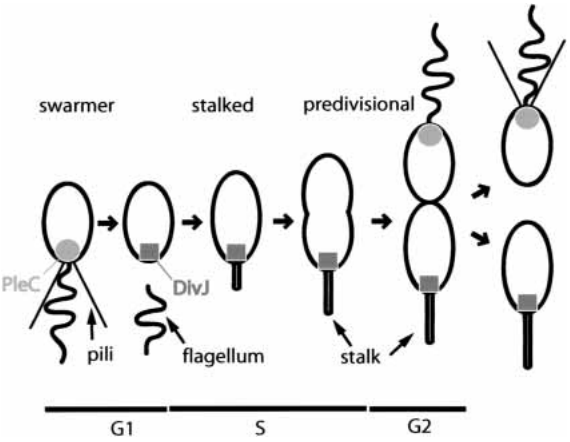


Figure 1. Dynamic localization of the PleC and DivJ sensor protein kinases during the *C. crescentus* cell cycle. The positioning of PleC (circle) and DivJ (rectangle; Wheeler and Shapiro 1999) during the cell cycle are indicated. Polar organelles and cell cycle stages are specified.

Thr 214, His/Pro 234, and Tyr/Asn 234) and retains its activity even when its phosphoryl acceptor site Asp 53 is modified, suggesting that this mutant form does not rely on phosphorylation input by a cognate kinase (Aldridge et al. 2003). In vivo phosphorylation experiments indicate that the polar kinases DivJ and PleC are involved in modulating the phosphorylation status of PleD. Whereas PleD~P is reduced to about 10% in cells lacking DivJ, it is undetectable in a mutant lacking both DivJ and PleC (Aldridge et al. 2003). DivJ and PleC are asymmetrically positioned at opposite cell poles, with DivJ localizing to the stalked pole coinciding with the requirement for active PleD during cell differentiation (Fig. 1; Wheeler and Shapiro 1999; Ohta and Newton 2003). Here we provide evidence that DivJ and PleC directly interact with PleD to modulate its phosphorylation state, suggesting that together they are responsible for PleD phosphorylation in vivo. We show that the PleD regulator dynamically localizes to the differentiating stalked pole during the cell cycle as a function of its phosphorylation state. Our results indicate that only activated PleD is sequestered to the stalked pole, providing a mechanism that spatially restricts PleD activity to the emerging stalked pole, where it coordinates polar morphogenesis.

PleD is a multidomain protein with two N-terminal receiver modules arranged in tandem and a C-terminal domain apparently serving as an output module (Hecht and Newton 1995). This putative output domain, termed "GGDEF" or "DUF1," is widespread and highly conserved in many bacterial species. Postulating a local activity of PleD at one cell pole calls for a molecular mechanism that converts the phosphorylation input into a readout that affects downstream targets. We propose that the PleD readout is the production of a cyclic nucleotide, which acts as secondary messenger. In vitro experiments with *C. crescentus* crude extracts and with purified PleD protein show that PleD contains an intrinsic nucleotide cyclase activity, which converts two molecules of GTP into cyclic diguanylic acid (c-di-GMP). Cyclase activity correlates with PleD activation by phosphorylation and requires an intact PleD C-terminal output domain. This suggests that the GGDEF domain constitutes a novel class of guanylate cyclases, which in PleD is specifically activated in response to phosphorylation of the N-terminal receiver domain. Our findings, together with the observation that more than 900 GGDEF proteins are reported in the nonredundant SMART database (Schultz et al. 1998), implies that diguanylate cyclases are abundant in the bacterial kingdom and that the diffusible molecule c-di-GMP might be a common secondary messenger in prokaryotes.

Results

DivJ and PleC directly control PleD phosphorylation

Genetic data (Sommer and Newton 1991; Aldridge et al. 2003) and in vivo phosphorylation experiments (Aldridge et al. 2003) established a role of the polar kinases DivJ and PleC in PleD control. To test whether DivJ and PleC directly modulate phosphorylation of PleD, in vitro phosphorylation assays were carried out using purified full-length PleD, fused to either a GST or a hexa-histidine tag, and purified soluble catalytic domains of DivJ (DivJ′) and PleC (PleC′). DivJ′ and PleC′ autophosphorylate in the presence of ATP and Mg^{2+} (Fig. 2). The addition of GST-PleD to autophosphorylated DivJ′ and PleC′ results in transfer of phosphate to PleD (Fig. 2A), whereas the purified PleD protein is not phosphorylated in the presence of ATP alone (data not shown). Autophosphorylation of purified PleC′ is relatively inefficient (Fig. 2A). Although this is in agreement with earlier findings (Hecht et al. 1995), we find that the addition of GST-PleD to the autophosphorylated soluble PleC′ kinase fragment results in a rapid loss of PleC~P, presumably by transfer of the phosphoryl group to the response regulator (Fig. 2A). The phosphotransfer from the kinases to GST-PleD is incomplete, possibly because of interference by the N-terminal GST tag. When using a PleD–His6 fusion protein instead, efficient phosphotransfer from DivJ′ and PleC′ is observed. The addition of PleD but not $PleD_{D53N}$ (lacking the phosphoryl acceptor site) to autophosphorylated DivJ′ or PleC′ results in an almost complete phosphotransfer to the response regulator (Fig. 2B). Because both PleD–His6~P and GST-PleD~P could barely be detected, we hypothesize that under these conditions the stability of the phosphorylated form of PleD is relatively low. This suggests that both DivJ and PleC directly interact with the PleD response regulator.

The PleD response regulator dynamically localizes to the stalked pole during the cell cycle

Whereas the sensor kinases DivJ and PleC are membrane-bound, the PleD response regulator is a soluble cytoplasmic protein (Fig. 3A). However, the fact that

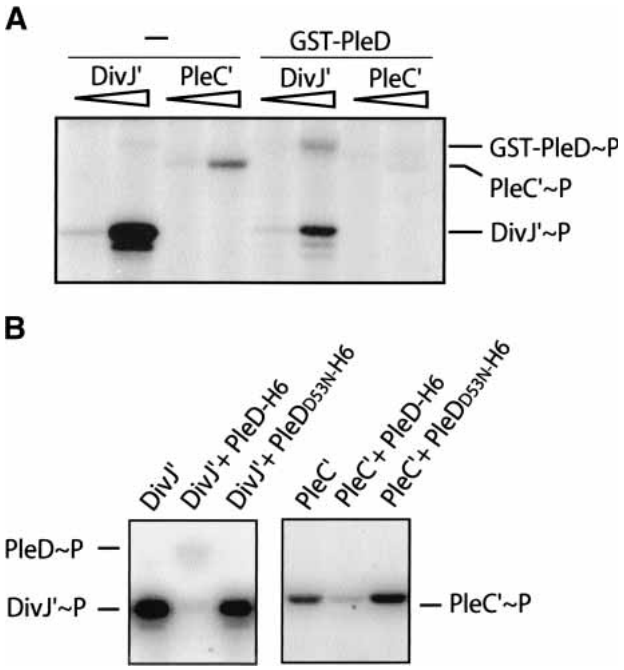


Figure 2. In vitro phosphotransfer between the protein kinases DivJ and PleC and the response regulator PleD. DivJ and PleC autophosphorylation in the presence of [γ-^{32}P]ATP and subsequent phosphotransfer to PleD are shown. The bands corresponding to the phosphorylated proteins are marked on the side. (*A*) Assays contained 0.5 and 5 µg of the soluble kinase fragments and 37.5 µg GST-PleD as indicated. (*B*) Assays contained 12.5 µg DivJ′, 20 µg PleC′, and 50 µg PleD-H6 and PleD$_{D53N}$H6, respectively, as indicated. In this experiment, DivJ′ and PleC′ were preincubated with [γ-^{32}P]ATP for 15 min before PleD or PleD$_{D53N}$ were added to the reaction mix for an additional 5 min (PleC′) or 10 min (DivJ′), respectively.

DivJ and PleC are specifically localized in the cell (Fig. 1) implies that information transfer from the sensor kinase to PleD requires the physical presence of the response regulator at the cell poles. In addition, the role of PleD in controlling assembly and function of polar organelles during development suggested that the PleD regulatory output might be restricted to the cell pole. To test the hypothesis that PleD might perform its regulatory function locally, we first analyzed the subcellular distribution of PleD during the *C. crescentus* cell cycle. A PleD–GFP fusion was introduced into the Δ*pleD* strain UJ284 on a low-copy number plasmid, and the analysis of motility and stalk formation of the resulting strain (UJ626) confirmed that the PleD–GFP fusion protein was functional (data not shown). The same fusion was also introduced into the wild-type strain CB15N (UJ627), and immunoblot analysis with anti-PleD and anti-GFP antibodies confirmed that in both strains the PleD–GFP fusion was produced at similar levels to PleD wild-type and excluded degradation of the fusion protein and the release of soluble GFP (data not shown). Analysis of strain UJ627 by fluorescence microscopy revealed that in a large fraction of stalked and predivisional cells, PleD–GFP is concentrated at the stalked pole (Fig. 3B, Table 1). From a total of 1000 cells counted, 36% had visible GFP foci at the cell pole, whereas only 4% of the cells had nonpolar foci. Localization of PleD–GFP in strains UJ626 and UJ627 was qualitatively and quantitatively indistinguishable (data not shown). Importantly, in all cases in which the identity of the cell poles could be determined

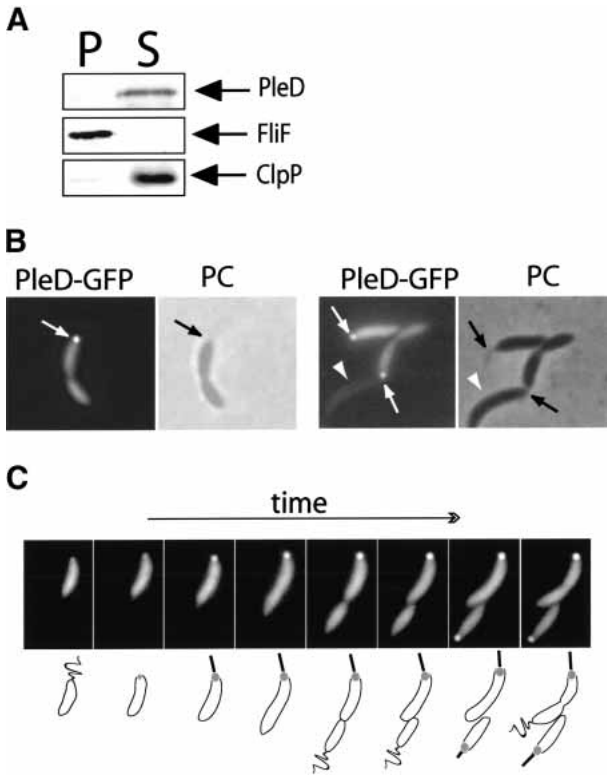


Figure 3. PleD is a soluble protein that localizes to the stalked pole of *C. crescentus* cells. (*A*) Immunoblot analysis of fractionated cell extracts of *C. crescentus* CB15N wild-type with anti-PleD, anti-FliF, and anti-ClpP antibodies. Cells of a logarithmically growing culture were lysed and soluble, and insoluble fractions were separated as indicated in Materials and Methods. Staining of the membrane-integral FliF and the soluble ClpP protein demonstrates the quality of the cell fractionation. (P) Insoluble pellet fraction; (S) soluble fraction. (*B*) PleD specifically localizes to the stalked pole. Phase contrast (PC) and fluorescent images of wild-type CB15N strains producing PleD–GFP from a low-copy number plasmid. The arrows indicate the polar foci of PleD–GFP in the fluorescent images and the stalk structures visible by phase contrast. (*C*) PleD dynamically localizes to the stalked pole during the *C. crescentus* cell cycle. Representative time-lapse experiment on *C. crescentus* wild-type cells producing PleD–GFP from a low-copy number plasmid. Fluorescent images (*top*) and a schematic representation of the cell cycle-dependent localization of PleD–GFP (*bottom*) are shown. In young swarmer cells, PleD–GFP is uniformly distributed in the cytoplasm. As the cells progress through the cell cycle and differentiate into stalked cells, PleD–GFP accumulates at the old flagellated and emerging stalked pole coinciding with flagellar ejection and stalk formation. PleD–GFP remains at the stalked pole throughout the cell cycle and is randomly dispersed in the newly formed swarmer cell. Only when the swarmer cell differentiates into a stalked cell again does PleD–GFP localize to the pole.

Table 1. *Quantitative analysis of PleD-GFP localization in* Caulobacter crescentus *wild-type and mutant strains*

Construct	Strain (genotype)	% cells with polar foci[a]	% cells with non-polar foci	Total cells counted
PleD-GFP	CB15N (wild type)	36	4	1000
$PleD_{D53N}$-GFP	CB15N (wild type)	0	0	500
PleD-GFP	UJ506 (Δ*pleC*)	22	12	396
PleD-GFP	UJ998 (*divJ*::Ω)	13	22	494
PleD-GFP	UJ1000 (Δ*pleC divJ*::Ω)	0	0	500
$PleD^{\star}_{D53N}$-GFP	CB15N (wild type)	37	10	348
$PleD^{\star}_{D53N}$-GFP	UJ1000 (Δ*pleC divJ*::Ω)	40	17	381
$PleD_{GG368DE}$-GFP	UJ284 (Δ*pleD*)	44	6	617

[a]Based on the dynamic behavior of PleD-GFP during the cell cycle (Fig. 3), about two-thirds of the cells of a mixed population are expected to show polar PleD-GFP foci. Whether this discrepancy is a biological property of the system or due to a technical property of the experiment is not clear.

unequivocally by the presence of a visible polar stalk, the GFP foci were associated with the stalked pole. This strongly implied that the PleD–GFP protein specifically localizes to the stalked pole and is absent from the flagellated swarmer pole. This, in turn, suggested dynamic behavior of the PleD protein during the *C. crescentus* cell cycle. To resolve the dynamic spatial distribution of PleD–GFP during the cell cycle, we performed time-lapse fluorescence microscopy with isolated swarmer cells of strain UJ627. Swarmer cells were grown directly on a microscope slide coated with a thin layer of agar, and progression though the cell cycle was visualized by phase contrast microscopy (Fig. 3C). The PleD protein is evenly distributed within *C. crescentus* swarmer cells, but then concentrates at the emerging stalked pole during the swarmer-to-stalked cell differentiation. With increasing time, the signal at the stalked pole increases in strength, whereas the pole opposite the stalk remains free of PleD–GFP throughout the entire cell cycle (Fig. 3C). This results in an asymmetric PleD–GFP distribution throughout most of the cell cycle and, on division, generates two different progeny cells: a swarmer cell with a uniform distribution of PleD–GFP and a stalked cell with an accumulation of PleD–GFP at the stalked pole. Only after the newborn swarmer cell has undergone the morphological transition into a stalked cell does PleD–GFP concentrate at this pole (Fig. 3C). The new poles generated by cytokinesis remain free of PleD–GFP protein.

Only activated PleD localizes to the pole

It is evident from the illustrations in Figure 3B and C that even in stalked and predivisional cells only a fraction of the PleD–GFP protein accumulates at the pole, whereas the rest seems to be evenly distributed in the cytoplasm. This is most evident from the observation that few cells, which do not seem to express the *pleD–gfp* copy, not only lack polar foci but also have a lower cytoplasmic fluorescence signal (Fig. 3B, short arrows). One possible explanation for this is that PleD exists in two different forms that have different targeting properties. To test whether phosphorylation of PleD is required for dynamic localization, we fused GFP to an inactive PleD mutant form that lacks the aspartic acid phosphoryl acceptor residue at position 53 (Asp 53). Immunoblot analysis confirmed that the resulting fusion protein $PleD_{D53N}$–GFP is stable and produced at wild-type levels (data not shown). However, in contrast to PleD–GFP, $PleD_{D53N}$–GFP is homogenously distributed in all cells and fails to accumulate at the stalked pole (Fig. 4A; Table 1), irrespective of the genetic background (data not shown). PleD–GFP also fails to localize in a mutant strain lacking both the DivJ and PleC kinases (Fig. 4B; Table 1). In this mutant PleD phosphorylation is reduced below detectable levels in vivo (Aldridge et al. 2003). A partial loss of PleD localization to the stalked pole was observed in mutants lacking either PleC or DivJ (Table 1).

These data suggest that phosphorylation plays a critical role in sequestering PleD to the pole. To support this and to distinguish between the possibilities that phosphorylation itself might constitute the targeting signal or, alternatively, that PleD preferentially binds to the cell pole in its active conformation, we analyzed the dynamic behavior of a constitutively active PleD mutant protein, $PleD^{\star}_{D53N}$. $PleD^{\star}_{D53N}$ is dominant over wild-type PleD, but because it lacks the aspartic acid phosphoryl acceptor site at position 53, its activity is not dependent on phosphorylation (Aldridge et al. 2003). As shown in Figure 4C, $PleD^{\star}_{D53N}$–GFP localizes almost exclusively to the cell poles in both the wild-type and the *divJ pleC* double mutant. The polar foci of $PleD^{\star}_{D53N}$–GFP are considerably stronger than the foci observed for wild-type PleD–GFP, and the cytoplasmic signal is significantly reduced. The fact that cellular levels of the two fusion proteins are similar (data not shown) suggests that a larger fraction of the activated PleD protein is concentrated at the pole compared with the wild-type PleD–GFP fusion. Pole selectivity of $PleD^{\star}_{D53N}$–GFP is similar to wild-type PleD in that the protein has a strong preference for the stalked pole and is absent from newly formed poles at division and from flagellated swarmer poles (Fig. 4C; Table 1). Pole selectivity of $PleD^{\star}_{D53N}$–GFP is unaltered even in the absence of the stalked pole-specific kinase DivJ (Fig. 4C),

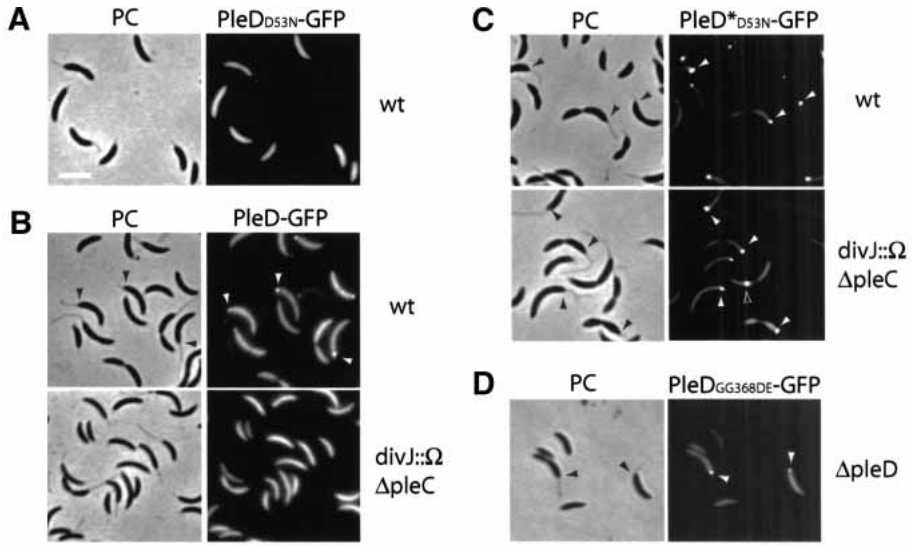


Figure 4. Dynamic localization of PleD to the stalked pole requires its activation by phosphorylation and is dependent on the polar kinases DivJ and PleC. (*A*) A nonphosphorylatable PleD mutant is impaired in polar localization. Phase contrast (PC) and corresponding fluorescent image of wild-type strain CB15N producing $PleD_{D53N}$–GFP from a low-copy number plasmid. Bar: *left* panel, 2 µm. (*B*) DivJ and PleC are required for the polar positioning of wild-type PleD–GFP. Phase contrast (PC) and corresponding fluorescent images of wild-type and *divJ pleC* mutant strain producing PleD–GFP from a low-copy number plasmid. (*C*) DivJ and PleC are not required for localization of the constitutive mutant $PleD^{\star}_{D53N}$–GFP. Phase contrast (PC) and corresponding fluorescent images of wild-type and *divJ pleC* mutant strain producing $PleD^{\star}_{D53N}$–GFP from a low-copy number plasmid. (*D*) An active GGDEF output domain is not required for localization of PleD. Phase contrast (PC) and corresponding fluorescent image of a *pleD* mutant strain producing $PleD_{GG368DE}$–GFP from a low-copy number plasmid. Filled arrows point to stalked poles in the phase contrast images (black) and to polar PleD–GFP foci in the fluorescent images (white). Open arrows point to nonpolar PleD–GFP foci.

indicating that DivJ contributes to PleD localization mainly by activating the response regulator. Together, these data are consistent with the idea that activated PleD protein is specifically targeted to the emerging stalked pole. The observation that a $PleD_{GG368DE}$–GFP fusion protein, which lacks an active C-terminal output domain (see below), still localizes to the stalked pole (Fig. 4D) suggests that an activated conformation of PleD, rather than the PleD readout itself, is required for polar sequestration of the regulator.

The PleD response regulator is a di-guanylate cyclase

The experiments described above suggest that PleD accumulates at the old pole of the cell only in its activated state. Because genetic data indicated that PleD~P is required for the differentiation of a flagellated into a stalked pole (Hecht and Newton 1995; Aldridge et al. 2003), PleD could act locally at this subcellular site, coordinating the developmental events involved in pole remodeling. However, what could be the output signal generated by the activated PleD response regulator, which in turn controls these downstream events? A report has established a link between a multidomain protein family containing the GGDEF domain and the metabolism of cyclic di-GMP (c-di-GMP), a compound discovered as a cofactor of cellulose synthase in *Gluconacetobacter xylinum* (Ross et al. 1991; Tal et al. 1998). To examine the possibility that the PleD output domain harbors di-guanylate cyclase activity, we attempted to biochemically assay its ability to convert GTP into c-di-GMP. Although extracts of *C. crescentus* wild-type strain CB15N and CB15N Δ*pleD* showed no activity (Fig. 5A), GTP was readily converted into a novel nucleotide compound when extracts of a strain containing the *pleD** or $pleD^{*}_{D53N}$ alleles were used (Fig. 5A; data not shown). To demonstrate that PleD was responsible for this activity, PleD with a C-terminal His-tag was overexpressed in *Escherichia coli* and purified to homogeneity. In the presence of purified PleD protein, GTP rapidly disappeared and was replaced by a nucleotide with a retardation factor (RF) value identical to the one observed with crude extracts (Fig. 5B).

To confirm that the novel spot indeed corresponds to the cyclic dimeric form of GMP, the reaction product of PleD and GTP was analyzed by mass spectrometry. The major peak resulting from the mass fingerprinting corresponds to a molecular mass of 689, which exactly matches the molecular weight of c-di-GMP (Fig. 5C). To gather additional evidence for the proposed enzymatic reaction, product inhibition was investigated. When chemically synthesized c-di-GMP was added to the reaction mix in a concentration range similar to the GTP substrate, strong inhibition was observed (Fig. 6A). This suggests that c-di-GMP effectively competes with GTP for the binding site. To exclude the possibility that PleD also possesses phosphodiesterase (PDE) activity and catalyzes the cleavage of c-di-GMP into two GMP mono-

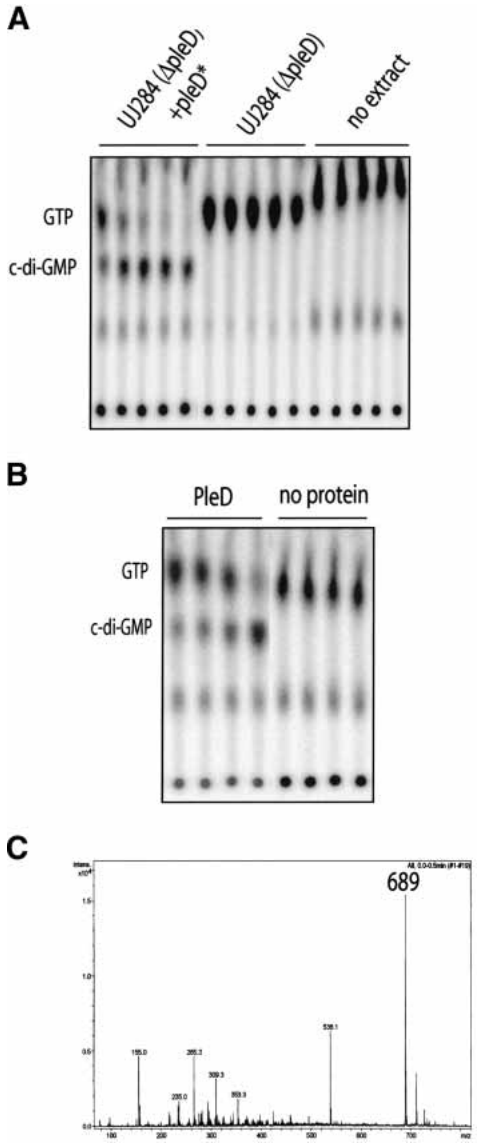


Figure 5. PleD is a di-guanylate cyclase. (*A*) Cyclic-di-GMP is produced by *C. crescentus* cell extracts. The soluble fraction of total cell extracts of CB15N Δ*pleD* (UJ284) or CB15N Δ*pleD* producing PleD* from a low-copy number plasmid were used to assay for di-guanylate cyclase activity. Control reactions without cell extracts are also shown. Samples were taken at 5, 10, 15, 20, and 30 min after addition of extracts and were analyzed on thin-layer chromatography plates. (*B*) Cyclic-di-GMP is produced by purified PleD protein. PleD–His6 (50 µg) was tested for di-guanylate cyclase activity, and the products of the enzymatic reactions were analyzed for 30, 45, 60, and 300 sec (lanes *1–4*) after addition of purified PleD–His6, as indicated in *A*. (Lanes *5–8*) Control reactions without PleD are also shown. (*C*) Analysis of products synthesized by PleD in vitro. The reaction products of PleD (200 µg) and GTP (1 mM) were separated by HPLC (cf. Fig. 6B). Peaks were collected and applied to mass spectrometry analysis. The fragmentation pattern shown corresponds to the main reaction product with a high-pressure liquid chromatography retention time of 6.73 min. The main peak had a molecular mass of 689 (theoretical molecular weight of c-di-GMP: 688.4). Reisolation and analysis of the substance with the molecular mass of 689 resulted in an identical fragmentation pattern.

mers, c-di-GMP concentration was measured quantitatively by HPLC after incubation with purified PleD–His6 protein for several hours. Neither a decrease of the c-di-GMP concentration nor a conversion of c-di-GMP into GMP or any other degradation product was observed during a prolonged incubation period (Fig. 6B). The absence of PDE activity is not caused by the loss of PleD enzyme activity because the PleD sample used in this experiment had a high di-guanylate cyclase (DGC) activity (Fig. 6B). Together, this is consistent with the view that the PleD protein harbors a di-guanylate cyclase activity, which specifically catalyzes the conversion of GTP into the di-cyclic form of guanosine monophosphate, and that this activity constitutes the signaling output of the PleD response regulator.

The PleD nucleotide cyclases activity is GTP specific

Nucleotide cyclases have been described for both adenosine and guanosine nucleotides (Domino et al. 1991; Johnson and Salomon 1991). To investigate whether the nucleotide cyclase activity of PleD is specific for GTP, we measured the synthesis of radiolabeled [^{32}P]c-di-GMP from [α^{32}P]GTP (0.1 mM) on addition of nonlabeled nucleotides (see Materials and Methods). As expected, an excess of unlabeled GTP efficiently inhibits the formation of [^{32}P]-labeled c-di-GMP (Fig. 6A). Similarly, deoxyGTP was able to effectively compete with radiolabeled GTP, suggesting that GTP and deoxyGTP bind to PleD with comparable affinities (Fig. 6A). In contrast, the addition of ATP had only a marginal effect on [^{32}P]c-di-GMP formation. Although the addition of 100 µM nonlabeled GTP reduced the formation of [^{32}P]-labeled c-di-GMP by about 50%, the activity was unchanged in the presence of 100 and 500 µM ATP and dropped by only 10%–20% in the presence of a 10-fold higher concentration of ATP (Fig. 6A). This suggests that guanosine nucleotides bind specifically to the PleD nucleotide cyclases, whereas the affinity for ATP is significantly lower. However, HPLC analysis showed that neither ATP nor deoxyGTP were converted into their respective dimeric forms (data not shown). In summary, the nucleotide cyclase domain of PleD seems to specifically bind to guanosine nucleotides, but only GTP serves as a substrate for the formation of a dimeric product.

PleD-dependent synthesis of c-di-GMP is stimulated by phosphorylation of the receiver domain and requires an intact GGDEF output domain

The modular architecture of the PleD response regulator suggests that the receiver domain or domains are involved in information input and that the C-terminal GGDEF domain constitutes the regulatory output of the molecule (Parkinson and Kofoid 1992). To test whether the guanylate cyclase activity is indeed localized in the GGDEF domain, we determined the activity of wild-type PleD with the activity of two mutant proteins with amino acid changes in the highly conserved GGDEF sig-

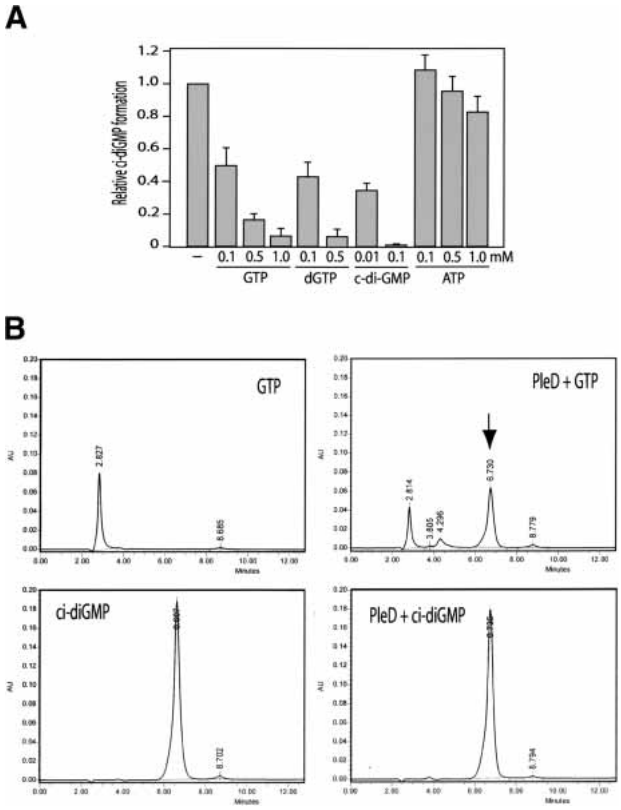


Figure 6. The PleD nucleotide cyclase activity is GTP-specific. (*A*) GTP, deoxyGTP, and c-di-GMP, but not ATP, specifically inhibit the PleD-dependent conversion of [^{32}P]GTP into [^{32}P]c-di-GMP. The relative c-di-GMP formation corresponds to the initial velocity determined for the enzymatic reactions. The reaction mixtures routinely contained 25 µg PleD and 100 µM GTP (see Materials and Methods) and were supplemented with nonlabeled nucleotides, as indicated *below* the graph. (*B*) PleD is a di-guanylate cyclase but lacks phosphodiesterase activity. High-pressure liquid chromatography analyses of GTP and c-di-GMP (*left*) and their reaction products with PleD (*right*) are indicated. The reaction mixtures contained 100 µg PleD and GTP and c-di-GMP at 200 µM each. Conversion of GTP into c-di-GMP by PleD is shown in the *top* panel by the appearance of a novel peak (arrow), which corresponds to chemically synthesized c-di-GMP (panels in *second* row). Incubation of PleD with chemically synthesized c-di-GMP for several hours did not lead to the cleavage or disappearance of the cyclic substance.

nature motif. The mutant alleles $pleD_{\Delta 368-372}$ (lacking the entire GGDEF motif) and $pleD_{GG368DE}$ (two highly conserved Gly residues in GGDEF replaced by Asp and Glu) failed to complement the *pleD* mutant phenotype, even though their products were stably expressed in *C. crescentus* (Aldridge and Jenal 1999). Consistent with their functional deficiency in vivo, both mutant proteins completely lack di-guanylate cyclase activity in vitro (Fig. 7A). This is in line with the idea that the C-terminal GGDEF domain represents the output domain of PleD and is responsible for the enzymatic activity observed.

Because no phospho-donors for PleD were present in the assays, our data indicate that activation of PleD by phosphorylation is not strictly required for the in vitro synthesis of c-di-GMP. This is supported by the fact that the activity of the mutant protein $PleD_{D53N}$, lacking the phosphoryl acceptor side Asp 53, is comparable to wild-type PleD (Fig. 7A,B). However, when purified DivJ kinase was added to the reaction mix, the di-guanylate cyclase activity of wild-type PleD was significantly stimulated (Fig. 7A). This stimulation was not only dependent on the presence of ATP but also required residue Asp 53, as a DivJ-dependent increase of enzyme activity was not observed for the $PleD_{D53N}$ mutant protein (Fig. 7A). Consistent with an increase of PleD enzyme activity on activation of the molecule by phosphorylation, we found that the constitutively active mutant proteins $PleD^{\star}_{D53N}$ and PleD* had a considerably higher specific activity than unphosphorylated wild-type PleD (Fig. 7B). This dramatic increase of the in vitro di-guanylate cyclase activity suggests that the dominant phenotypic effects of the *pleD** and $pleD^{*}_{D53N}$ alleles is caused by an uncontrolled overproduction of c-di-GMP (Aldridge et al. 2003). The observation that PleD in the presence of DivJ and ATP is by far less active than purified PleD* may be the result of the modest efficiency of PleD phosphorylation in vitro (Fig. 2A). Although the molecular mechanism of PleD activation remains to be elucidated, these data clearly indicate that phosphorylation of the first receiver domain leads to an increased activity of the C-terminal guanylate cyclase domain.

Discussion

Localization of signaling molecules is a conserved mechanism for the establishment of cell polarity in both prokaryotes and eukaryotes (Shapiro et al. 2002; Nelson 2003). In *C. crescentus*, cell polarity and the developmental program are controlled by sensor histidine kinases, which are asymmetrically positioned at the ends of the cell (Shapiro et al. 2002). Here we present data indicating that two of these polar kinases, DivJ and PleC, control the activity and the dynamic localization of the soluble response regulator PleD during the cell cycle. The activated form of PleD, PleD~P, possesses catalytic guanylate cyclase activity and is specifically sequestered to one pole of the cell, arguing that spatially confined synthesis of a diffusible secondary messenger might contribute to the temporal and spatial control of pole development in this organism.

Time-lapse experiments with a PleD–GFP fusion revealed a highly dynamic behavior and precise localization mechanism for the PleD response regulator during the *C. crescentus* cell cycle. PleD–GFP is randomly dispersed in the cytoplasm of swarmer cells but then localizes to the emerging stalked pole during cell differentiation. This dynamic positioning coincides with the localization of DivJ to the same pole (Wheeler and Shapiro 1999) and presumably precedes flagellar release and stalk formation, both of which are dependent on activated PleD (Hecht and Newton 1995; Aldridge and Jenal 1999). We have presented several lines of evidence indicating that activation of PleD by phosphorylation is critical for polar targeting and that only the activated form of the

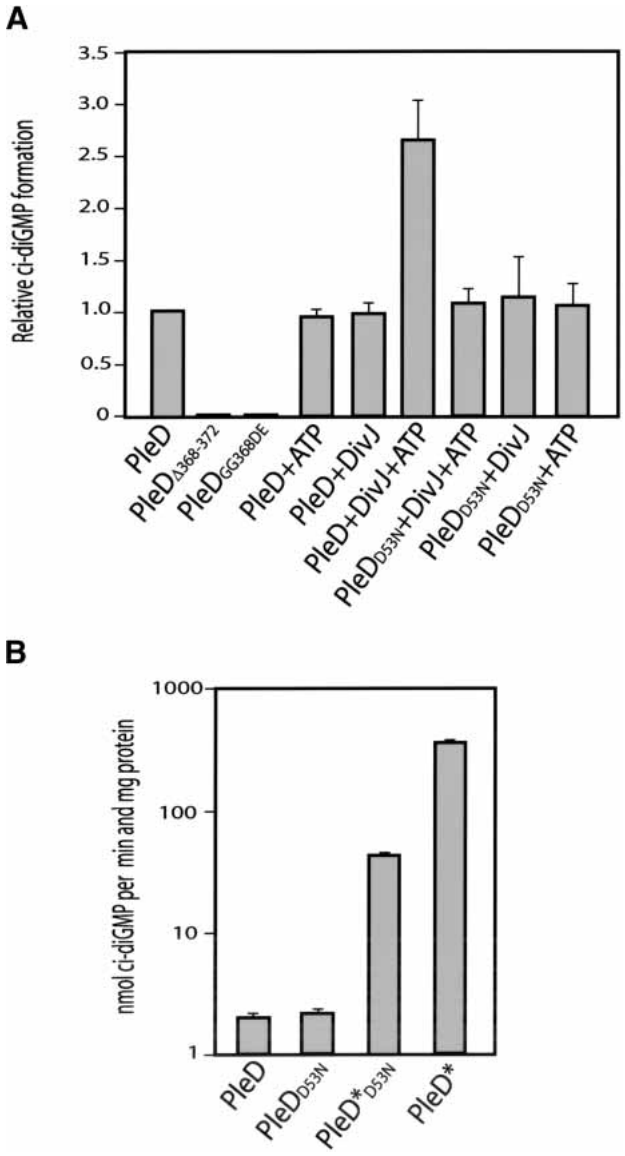


Figure 7. Activation of the PleD di-guanylate cyclase by phosphorylation and the requirement of an intact GGDEF output domain. (*A*) Influence of phosphorylation and mutations in the GGDEF domain on the PleD in vitro di-guanylate cyclase activity. The enzymatic reactions contained 100 µM GTP and were carried out with 25 µg of the following purified proteins: PleD, PleD$_{D53N}$, PleD$_{\Delta 368–372}$, and PleD$_{GG368DE}$. DivJ (12.5 µg) and ATP (200 µM) were added where indicated. The relative c-di-GMP formation corresponds to the initial velocity measured for the enzymatic reactions. (*B*) Constitutive active forms of PleD mimic the activated state of the di-guanylate cyclase. The enzymatic reactions contained 200 µM GTP and were carried out with 12.5 µg of the following purified proteins: PleD, PleD$_{D53N}$, PleD*$_{D53N}$, and PleD*. The relative c-di-GMP formation corresponds to the initial velocity measured for the enzymatic reactions.

response regulator accumulates at the pole. Localization experiments with PleD*$_{D53N}$–GFP suggest that it is not phosphorylation itself but, rather, an activated conformation of the protein that provides the information for polar localization. Two different mechanisms can be envisioned to explain the coupling between activity and polar localization of PleD. PleD could auto-catalytically control its own subcellular positioning, for instance, by altering the nature of the cell pole. Alternatively, pole recognition might be restricted to the activated form of PleD. The observation that the inactive PleD$_{D53N}$–GFP fusion protein does not accumulate at the stalked pole in the presence of a chromosomal *pleD* wild-type gene, and the finding that GFP fused to PleD$_{GG368DE}$, which is unable to generate c-diGMP, still sequesters to the pole in a *pleD* mutant, favors the second mechanism. Preliminary evidence suggests that purified PleD is able to form dimers, raising the possibility that activation and targeting might be a consequence of PleD dimerization. The observed coupling between PleD activity and polar localization is reminiscent of the mechanism observed for the single-domain response regulator DivK (Jacobs et al. 2001; Lam et al. 2003). However, although DivK~P localizes to both the swarmer and the stalked pole, PleD~P shows no detectable affinity for the flagellated pole. The affinity of DivK for the cell poles is also mediated through the DivJ and PleC kinases, but in contrast to PleD, which requires both kinases to be sequestered to the stalked pole, DivK targeting to the poles is mediated by DivJ, whereas PleC controls its release from the swarmer pole late in the cell cycle (Jacobs et al. 2001). Neither PleD–GFP nor the constitutive active form, PleD*$_{D53N}$–GFP, localize to the pole in swarmer cells, irrespective of the presence or absence of PleC. This argues for a marker at the stalked pole that appears or is unmasked during the swarmer-to-stalked cell transition and is recognized by activated PleD. Such a marker has already been postulated for the localization of DivJ (Wheeler and Shapiro 1999), but its molecular identity remains unknown.

Targeting of active PleD to the differentiating stalked pole might serve to position the output domain of the response regulator in close proximity to the machinery that is responsible for the morphogenetic changes during cell differentiation. PleD is a di-guanylate cyclase, which on activation by phosphorylation, synthesizes c-di-GMP. This low-molecular weight molecule was originally identified as a positive allosteric effector of cellulose synthase in *G. xylinum* and *Agrobacterium tumefaciens* (Ross et al. 1987; Amikam and Benziman 1989). The conversion of glucose moieties into cellulose polymers is energetically costly for the cell, and it has been postulated that the committing step is tightly regulated by c-di-GMP to adjust the polymerization process to the cell's metabolism (Ross et al. 1991). The intracellular concentration of c-di-GMP in *G. xylinum* seems to be controlled by the opposing activities of DGCs and c-di-GMP-specific PDEs (Ross et al. 1987). Tal and coworkers were able to identify three operons in *G. xylinum* involved in cellular turnover of c-di-GMP (Tal et al. 1998; Chang et al. 2001). Each operon contains a pair of paralogous genes termed *pde* and *dgc*, which code for multidomain proteins with an N-terminal PAS/PAC domain (Ponting and Aravind 1997), a central GGDEF (DUF1)

domain, and a C-terminal EAL (DUF2) domain (Galperin et al. 2001). Although genetic data suggested that a PleD homolog in *Rhizobium* had DGC activity (Ausmees et al. 2001), the following evidence presented in this work strongly suggests that the GGDEF domain of PleD possesses DGC, but no PDE activity: First, in vitro synthesis of c-di-GMP with *C. crescentus* cell extracts is dependent on the presence of a constitutively active form of PleD. Second, purified PleD protein is able to efficiently convert GTP into a nucleotide species with a molecular mass corresponding exactly to that expected for c-di-GMP. Third, the DGC activity of purified PleD is dependent on an intact GGDEF domain and the PleD cyclase activity is specific for GTP. Fourth, purified PleD possesses no detectable PDE activity (in contrast, PDE activity can readily be detected in *C. crescentus* whole-cell extracts; data not shown). Fifth, the in vitro DGC activity of PleD is stimulated several fold by phosphorylation through its cognate kinase DivJ. Sixth, the specific DGC activity of a phosphorylation-independent form of PleD was up to two orders of magnitude higher than that of wild-type PleD. This dramatic increase of c-di-GMP synthesis is consistent with the dominant phenotype of the *pleD** allele with respect both to motility and to stalk formation (Aldridge et al. 2003). Taken together, these results support the view that the GGDEF domain represents a novel signaling domain with a bona fide DGC activity. This is in line with a recent structure prediction (Pei and Grishin 2001), which shows an excellent correspondence between GGDEF and the catalytic domain of adenylate cyclases.

The GGDEF proteins constitute one of the largest known families of orthologs with undefined function and three-dimensional structure (Schultz et al. 1998; Tatusov et al. 2001). Whereas proteins containing a GGDEF domain are found in most bacterial species for which the genome sequence is available, they are absent in archea and eukaryotes. The analysis of the domain architecture of GGDEF proteins listed in the nonredundant protein databases reveals an intriguing pattern. The GGDEF domain seems to be highly "promiscuous," as it is found associated as a module with a multitude of different domains. Intriguingly, all of these domains are known or proposed to be involved in signal sensing in the periplasm, the membrane, or the cytoplasm (Fig. 8). Although the nature of the signals is unclear in most cases, it has been well defined for others, like the PAS domain or hemerythrin (Gong et al. 1998; Terwilliger 1998). We propose that the GGDEF domains represent the output of a complex bacterial signal transduction network, which converts signals from different cellular compartments into the production of a secondary messenger, c-di-GMP (Fig. 8). The only two domains, which are often found associated with GGDEF and do not seem to be involved in signal sensing, are specialized metal-dependent phosphohydrolases (HD-GYP) and EAL domains. One could speculate that proteins containing both a GGDEF and an HD-GYP or EAL domain might have opposing cyclase and hydrolase activities, which contribute to the cellular level of c-di-GMP (Chang et al. 2001).

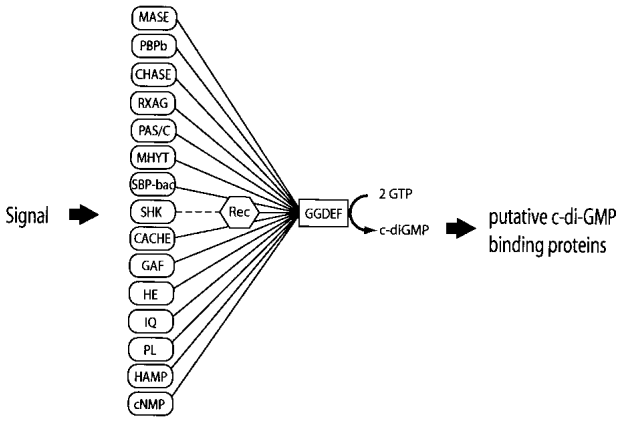


Figure 8. The GGDEF domain is coupled in a modular fashion with different sensory input or information transfer domains. The domain composition of GGDEF proteins listed in the SMART protein database (Schultz et al. 1998) is shown schematically. Recognized or putative signal sensing domains are indicated in rounded rectangles, the signal transfer domain (Rec) is indicated as a polygon, and the GGDEF domain is indicated as a rectangle. Connecting bars indicate the association of domains found in a single protein. The broken line symbolizes information transfer between sensor histidine kinases (e.g., DivJ) and the receiver domain of their cognate response regulator (e.g., PleD). (Rec) Receiver domain of response regulators; (MASE) membrane-associated sensor (MASE1 and MASE2; Nikolskaya et al. 2003); (PBPb) high-affinity periplasmic solute-binding protein of ABC-type amino acid transport system; (CHASE) cyclases/histidine kinases associated sensory exracellular (Anantharaman and Aravind 2001; Mougel and Zhulin 2001; Zhulin et al. 2003); (RXAG) permease component of ribose, xylose, arabinose, galactoside ABC transporter; (PAS/C:PAS) *Drosophila* period clock, aryl hydrocarbon receptor, and single-minded proteins (putative signaling domain; Ponting and Aravind 1997); (PAC) PAS C-terminal motif (Ponting and Aravind 1997); (MHYT) integral membrane sensor domain (Galperin et al. 1999); (SBP-bac) bacterial extracellular solute binding protein; (SHK) Sensor histidine kinase; (CACHE) signaling domain common to Ca^{2+} channels and chemotaxis receptors (Anantharaman and Aravind 2000); (GAF) cGMP-specific and -stimulated phosphodiesterases/adenylate cyclases (*Anabaena*)/FhlA (*E. coli*; Galperin et al. 2001); (HE) Hemerythrin, oxygen-binding protein (Stenkamp et al. 1978); (IQ) sequence motifs for calmodulin recognition (Rhoads and Friedberg 1997); (PL) phospholamban, small protein that regulates the affinity of the cardiac sarcoplasmic reticulum Ca^{2+}-ATPase for calcium (Smith et al. 2001); (HAMP) Histidine kinases, adenylyl cyclases, methyl-accepting proteins, phosphatases (Aravind and Ponting 1999); (cNMP) cyclic nucleotide-monophosphate-binding domain.

The presence of a large number of potential DGCs in single bacterial species (e.g., 39 in *Vibrio cholerae*) raises the question of how the output specificity of parallel signaling pathways might be achieved. Our finding that *C. crescentus* polarized cells spatially restrict the distribution of an active DGC to the site of morphogenetic changes could offer an explanation for this dilemma. Physical proximity between c-di-GMP synthesis and action could very well be of general regulatory significance. This is in agreement with the observation that in *G.*

xylinum, most of the c-di-GMP present in the cell seems to exist in a protein-associated rather than in a freely diffusible form (Ross et al. 1991; Weinhouse et al. 1997). An example for the compartmentalized production of a secondary messenger has been presented recently by Kriebel and collaborators, who showed that in *Dictyostelium discoideum*, asymmetric cellular distribution of adenylate cyclase is essential for cells to stream, possibly by contributing to the local secretion of the chemoattractant cAMP (Kriebel et al. 2003). Similarly, the observation that type VIII adenylate cyclase is enriched at cell–cell borders of endothelial cells could explain how localized changes in calcium-dependent cAMP concentrations regulate intercellular gap formation (Cioffi et al. 2002).

The finding that bacterial cells produce c-di-GMP as a regulatory compound highlights an added layer of complexity in bacterial signaling networks. What could be the cellular functions controlled by these regulatory mechanisms? Bacterial genetics has so far provided only a limited number of functional analyses of GGDEF proteins, but the results have revealed a recurring theme. In all cases, GGDEF proteins seem to be involved in the regulation of cell adhesion or cell surface colonization (Ausmees et al. 1999; Romling et al. 2000; Gronewold and Kaiser 2001; Boles and McCarter 2002; D'Argenio et al. 2002; Spiers et al. 2002; Bomchil et al. 2003). An interesting but so far poorly understood example is the HmsT protein, which allows the colonization and blockage of the flea foregut by *Yersinia pestis* and, as a result, the effective transmission of the plague bacillus to the mammal (Jones et al. 1999). A specific role for these novel regulatory components in adhesive behavior of bacterial cells could also help to explain why decades of (planktonic) bacterial genetics did not lead to the identification of the GGDEF network, which had to await the arrival of large-scale microbial genome analysis, bacterial cell biology, and an intensified scientific interest in microbial surface colonization.

Materials and methods

Strains, plasmids, and media

Bacterial strains and plasmids used in this study are shown in Table 2. *C. crescentus* strains were grown in complex peptone-yeast extract or in minimal glucose media (Ely 1991). Cultures of *C. crescentus* were synchronized by density gradient centrifugation as described previously (Jenal and Shapiro 1996). For conjugal transfer into *C. crescentus*, *E. coli* strain S17-1 was used as donor strain. *E. coli* strains were grown in Luria Broth (LB) media supplemented with antibiotics for selection, where necessary. The exact procedure of strain and plasmid construction (Table 2) is available on request.

Purification of PleD, PleC′, and Div′

E. coli cells carrying the respective expression plasmid were grown in LB medium with ampicillin (100 μg/mL), and expression was induced by adding either arabinose (final concentration of 0.2%) or IPTG (final concentration of 0.4 mM). After harvesting by centrifugation, the cells were resuspended in TN-buffer (50 mM Tris-HCl at pH 8.0, 500 mM NaCl, 1 mM β-mercaptoethanol), lysed by passage through a French press cell, and the suspension was clarified by centrifugation. The supernatant was loaded onto Ni-NTA affinity resin (Qiagen), washed with TN-buffer, and eluted with an imidazol-gradient. All PleD fusion proteins and the DivJ′ fragment remained soluble and were purified in native form on Ni-NTA affinity resin (Qiagen) or Glutathione-agarose (Clontech), whereas the PleC′ fragment was solubilized from inclusion bodies in guanidine hydrochloride and renatured after purification as described previously (Hecht et al. 1995). Protein preparations were examined for purity by SDS-PAGE. Fractions containing pure protein were pooled and dialyzed. *C. crescentus* cell extracts were prepared after harvesting cells by centrifugation and resuspension in TN-buffer. Cells were lysed by passage through a French press cell, and the extract was clarified by centrifugation. Soluble and insoluble protein fractions were separated by a high-spin centrifugation step (100,000 × g, 1 h; Jenal et al. 1994).

Enzymatic assays

Di-guanylate cyclase assays were adapted from procedures described previously (Ross et al. 1987). The reaction mixtures with purified PleD protein or *Caulobacter* cell extracts contained 75 mM Tris-HCl at pH 7.8, 250 mM NaCl, 25m M KCl, 10 mM $MgCl_2$ in 50 μL volume and were started by the addition of a mixture of 0.1 mM GTP [α-^{32}P]GTP (Amersham Biosciences; 0.01 μCi/μL). To calculate the initial velocity of product formation, aliquots were withdrawn at regular time intervals and the reaction was stopped with an equal volume of 50 mM EDTA. Reaction products (2.5 μL) were separated on polyethyleneimine-cellulose plates (Macherey-Nagel) in 1.5 M KH_2PO_4 (pH 3.65). Plates were exposed to a phosphor-imager screen, and the intensity of the various radioactive species was calculated by quantifying the intensities of the relevant spots using the imageQuant software (Molecular Dynamics). Measurements were always restricted to the linear range of product formation. Reaction mixtures for HPLC analyses were incubated 90 min at 25°C and terminated by heating to 95°C.

In vitro phosphorylation assays were adapted from a method described previously (Hecht et al. 1995). The proteins were incubated at 25°C for 20 min in phosphorylation buffer (50 mM Tris-HCl at pH 7.8, 25 mM NaCl, 25 mM KCl, 5 mM $MgCl_2$) containing 5 μCi [γ-32P]ATP (Amersham Biosciences). The reactions were stopped by adding one-third volume SDS-PAGE sample buffer (250 mM Tris-HCl at pH 6.8, 40% glycerol, 8% SDS, 2.4 M β-mercaptoethanol, 0.06% bromophenol blue, 40 mM EDTA), and ^{32}P-labeled proteins were separated by electrophoresis on 10% SDS-PAGE gels followed by autoradiography.

Synthesis and analysis of c-di-GMP

c-di-GMP was chemically synthesized as described by Ross and coworkers (Ross et al. 1990) and was purified by semipreparative reverse phase high-pressure liquid chromatography. Merck Lichrospher RP18e was used at 37°C with a 0.01 M triethylammonium carbonate buffer pH 7 containing 7.5% of methanol as mobile phase. A flow rate of 5 mL/min was used on a Hewlett Packard 1050 series system with ultrviolet detection at 252 nm. Synthetic c-di-GMP was used in comparative HPLC runs to characterize the enzymatic assay products. The conditions described above were used with a flow rate of 1 mL/min on a Waters Alliance 2690 separative module connected to a Waters 2487 ultroviolet detector. The retention time of c-di-GMP ranges between 6 and 7 min.

Table 2. *Strains and plasmids*

Strain	Relevant genotype or description	Reference or source
Caulobacter crescentus		
CB15N	Synchronizable variant strain of CB15	Evinger and Agabian 1977
UJ284	CB15N Δ*pleD*	Aldridge et al. 2003
UJ417	CB15N and plasmid pPA41	Aldridge et al. 2003
UJ506	CB15N Δ*pleC*	Aldridge et al. 2003
UJ626	UJ284 and plasmid pPA53-4	This study
UJ627	CB15N and plasmid pPA53-4	This study
UJ998	CB15N *divJ*::Ω	Aldridge et al. 2003
UJ1000	CB15N Δ*pleC*::Ω	Aldridge et al. 2003
UJ1168	CB15N and plasmid pPA114-47	Aldridge et al. 2003
UJ1169	UJ284 and plasmid pPA114-47	Aldridge et al. 2003
UJ1420	UJ1000 and plasmid pPA53-4	This study
UJ1466	UJ998 and plasmid pPA53-4	This study
UJ1875	UJ506 and plasmid pPA534	This study
UJ1909	CB15N and plasmid pSW6	This study
UJ1910	UJ284 and plasmid pSW7	This study
UJ2222	UJ506 and plasmid pSW7	This study
UJ2223	UJ998 and plasmid pSW7	This study
UJ2224	UJ1000 and plasmid pSW7	This study
UJ2262	UJ284 and plasmid pSW8	This study
Escherichia coli		
DH10B	F⁻*mcr*A Δ(*mrr*⁻ hsd RMS⁻ *mcr*BC) ϕ80d*lacZ*ΔM15 Δ*lac*X74 endA1 recA1 deoR Δ(*ara, leu*)7697 *araD139 galU galK nupG rpsL thi pro hsdR*⁻ *hsd*⁺ *recA RP4-2-Tc*::*Mu-Tn7*	Simon et al. 1983
BL21 (DE3)	*E. coli*B F⁻ *dcm ompT hsdS*(rB⁻mB⁻)	Stratagene
pLysS	*gal* λ(DE3) [pLysS CAM^r]	
BL21-CodonPlus	*E. coli* B F⁻ *ompT hsdS*(rB⁻ mB⁻) *dcm*⁺	Stratagene
(DE3)-RIL	Tet^r *gal l* (DE3) *endA* Hte [*argU ileY leuW* Cam^r]	

Plasmid	Relevant genotype or description	Reference or source
pMR20	Tet^R low copy number vector	Roberts et al. 1996
pBAD	Amp^R expression plasmid	Invitrogen
pGEX4T3	Amp^R vector for creation of GFP-fusion proteins	Amersham Biotech
pEGFP-N1	Amp^R expression plasmid	Clontech
pET11	Amp^R expression plasmid	Stratagene
pPA53-4	pMR20; *pleD-GFP* under the control of *divK* promoter	This study
pSW6	pMR20; $pleD_{D53N}$-*GFP* under the control of *divK* promoter	This study
pSW7	pMR20; $pleD^*_{D53N}$-*GFP* under the control of *divK* promoter	This study
pSW8	pMR20; $pleD_{GG368DE}$-*GFP* under the control of *divK* promoter	This study
pPA69	pGEX4T3; *pleD,* N-terminal GST-tag	This study
pRP49	pBAD; *pleC,* N-terminal His_6 tag	This study
pRP63	pBAD; *divJ,* N-terminal His_6 tag	This study
pCC2	pET11; *pleD,* C-terminal His_6 tag	This study
pRP87	pET11; $pleD_{D53N}$, C-terminal His_6 tag	This study
pRP88	pET11; $pleD_{\Delta 368\text{-}372}$, C-terminal His_6 tag	This study
pRP89	pET11; *pleD**, C-terminal His_6 tag	This study
pRP90	pET11; $pleD^*_{D53N}$, C-terminal His_6 tag	This study
pRP91	pET11; $pleD_{GG368DE}$, C-terminal His_6 tag	This study

Both synthetic and enzymatic c-di-GMP were also analyzed by mass spectrometry. ESI mass spectrometry was conducted on a Brucker Daltonics Esquire 3000 plus instrument.

Microscopy and photography

For fluorescence imaging, *C. crescentus* strains were grown in peptone-yeast extract media and placed on a microscope slide that was layered with a pad of peptone-yeast extract containing 1% agarose. The slide was placed on a microscope stage at room temperature (~22°C). Samples were observed on an Olympus AX70 microscope through a phase contrast 100× objective with a Hamamatsu C4742-95 digital camera. Images were taken and processed with Improvision Openlab and with Adobe Photoshop software.

Acknowledgments

We thank Dr. Phillip Aldridge for construction of plasmid pPA53-4 and Dr. Haim Weinhouse for technical assistance. The work was supported by Swiss National Science Foundation fellowships 31-59050.99 to U.J.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Aldridge, P. and Jenal, U. 1999. Cell cycle-dependent degradation of a flagellar motor component requires a novel-type response regulator. *Mol. Microbiol.* **32:** 379–391.

Aldridge, P., Paul, R., Goymer, P., Rainey, P., and Jenal, U. 2003. Role of the GGDEF regulator PleD in polar development of *Caulobacter crescentus*. *Mol. Microbiol.* **47:** 1695–1708.

Amikam, D. and Benziman, M. 1989. Cyclic diguanylic acid and cellulose synthesis in *Agrobacterium tumefaciens*. *J. Bacteriol.* **171:** 6649–6655.

Anantharaman, V. and Aravind, L. 2000. Cache—a signaling domain common to animal Ca^{2+}-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends Biochem. Sci.* **25:** 535–537.

———. 2001. The CHASE domain: A predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. *Trends Biochem. Sci.* **26:** 579–582.

Aravind, L. and Ponting, C.P. 1999. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol. Lett.* **176:** 111–116.

Ausmees, N., Jonsson, H., Hoglund, S., Ljunggren, H., and Lindberg, M. 1999. Structural and putative regulatory genes involved in cellulose synthesis in *Rhizobium leguminosarum bv. trifolii*. *Microbiology* **145:** 1253–1262.

Ausmees, N., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M., and Lindberg, M. 2001. Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. *FEMS Microbiol Lett.* **204:** 163–167.

Boles, B.R. and McCarter, L.L. 2002. Vibrio parahaemolyticus scrABC, a novel operon affecting swarming and capsular polysaccharide regulation. *J. Bacteriol.* **184:** 5946–5954.

Bomchil, N., Watnick, P., and Kolter, R. 2003. Identification and characterization of a *Vibrio cholerae* gene, *mbaA*, involved in maintenance of biofilm architecture. *J. Bacteriol.* **185:** 1384–1390.

Chang, A.L., Tuckerman, J.R., Gonzalez, G., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M., and Gilles-Gonzalez, M.A. 2001. Phosphodiesterase A1, a regulator of cellulose synthesis in *Acetobacter xylinum*, is a heme-based sensor. *Biochemistry* **40:** 3420–3426.

Cioffi, D.L., Moore, T.M., Schaack, J., Creighton, J.R., Cooper, D.M., and Stevens, T. 2002. Dominant regulation of inter-endothelial cell gap formation by calcium-inhibited type 6 adenylyl cyclase. *J. Cell Biol.* **157:** 1267–1278.

D'Argenio, D.A., Calfee, M.W., Rainey, P.B., and Pesci, E.C. 2002. Autolysis and autoaggregation in *Pseudomonas aeruginosa* colony morphology mutants. *J. Bacteriol.* **184:** 6481–6489.

Domino, S.E., Tubb, D.J., and Garbers, D.L. 1991. Assay of guanylyl cyclase activity. In *Methods in enzymology* (eds. R.A. Johnson and J.D. Corbin), pp. 345–355. Academic Press, San Diego, CA.

Ely, B. 1991. Genetics of *Caulobacter crescentus*. *Meth. Enzymol.* **204:** 372–384.

Evinger, M. and Agabian, N. 1977. Envelope-associated nucleoid from *Caulobacter crescentus* stalked and swarmer cells. *J. Bacteriol.* **132:** 294–301.

Galperin, M.Y., Natale, D.A., Aravind, L., and Koonin, E.V. 1999. A specialized version of the HD hydrolase domain implicated in signal transduction. *J. Mol. Microbiol. Biotechnol.* **1:** 303–305.

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

Gong, W., Hao, B., Mansy, S.S., Gonzalez, G., Gilles-Gonzalez, M.A., and Chan, M.K. 1998. Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction. *Proc. Natl. Acad. Sci.* **95:** 15177–15182.

Gronewold, T.M. and Kaiser, D. 2001. The *act* operon controls the level and time of C-signal production for *Myxococcus xanthus* development. *Mol. Microbiol.* **40:** 744–756.

Hecht, G.B. and Newton, A. 1995. Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in *Caulobacter crescentus*. *J. Bacteriol.* **177:** 6223–6229.

Hecht, G.B., Lane, T., Ohta, N., Sommer, J.M., and Newton, A. 1995. An essential single domain response regulator required for normal cell division and differentiation in *Caulobacter crescentus*. *EMBO J.* **14:** 3915–3924.

Jacobs, C., Hung, D., and Shapiro, L. 2001. Dynamic localization of a cytoplasmic signal transduction response regulator controls morphogenesis during the *Caulobacter* cell cycle. *Proc. Natl. Acad. Sci.* **98:** 4095–4100.

Jenal, U. and Shapiro, L. 1996. Cell cycle-controlled proteolysis of a flagellar motor protein that is asymmetrically distributed in the *Caulobacter* predivisional cell. *EMBO J.* **15:** 2393–2406.

Jenal, U., White, J., and Shapiro, L. 1994. *Caulobacter* flagellar function, but not assembly, requires FliL, a non-polarly localized membrane protein present in all cell types. *J. Mol. Biol.* **243:** 227–244.

Johnson, R.A. and Salomon, Y. 1991. Assay of adenylyl cyclase catalytic activity. In *Methods in enzymology* (eds. R.A. Johnson and J.D. Corbin), pp. 3–21. Academic Press, San Diego, CA.

Jones, H.A., Lillard Jr., J.W., and Perry, R.D. 1999. HmsT, a protein essential for expression of the haemin storage (Hms+) phenotype of *Yersinia pestis*. *Microbiology* **145:** 2117–2128.

Kriebel, P.W., Barr, V.A., and Parent, C.A. 2003. Adenylyl cyclase localization regulates streaming during chemotaxis. *Cell* **112:** 549–560.

Lam, H., Matroule, J.Y., and Jacobs-Wagner, C. 2003. The asymmetric spatial distribution of bacterial signal transduction proteins coordinates cell cycle events. *Dev. Cell.* **5:** 149–159.

Mougel, C. and Zhulin, I.B. 2001. CHASE: An extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. *Trends Biochem. Sci.* **26:** 582–584.

Nelson, W.J. 2003. Adaptation of core mechanisms to generate cell polarity. *Nature* **422:** 766–774.

Nikolskaya, A.N., Mulkidjanian, A.Y., Beech, I.B., and Galperin, M.Y. 2003. MASE1 and MASE2: Two novel integral membrane sensory domains. *J. Mol. Microbiol. Biotechnol.* **5:** 11–16.

Ohta, N. and Newton, A. 2003. The core dimerization domains of histidine kinases contain recognition specificity for the cognate response regulator. *J. Bacteriol.* **185:** 4424–4431.

Parkinson, J.S. and Kofoid, E.C. 1992. Communication modules in bacterial signaling proteins. *Annu. Rev. Genet.* **26:** 71–112.

Pei, J. and Grishin, N.V. 2001. GGDEF domain is homologous to adenylyl cyclase. *Proteins* **42:** 210–216.

Ponting, C.P. and Aravind, L. 1997. PAS: A multifunctional domain family comes to light. *Curr. Biol.* **7:** 674–677.

Rhoads, A.R. and Friedberg, F. 1997. Sequence motifs for calmodulin recognition. *FASEB J.* **11:** 331–340.

Roberts, R.C., Toochinda, C., Avedissian, M., Baldini, R.L., Gomes, S.L., and Shapiro, L. 1996. Identification of a *Caulobacter crescentus* operon encoding hrcA, involved in negatively regulating heat-inducible transcription, and the chaperone gene *grpE*. *J. Bacteriol.* **178:** 1829–1841.

Romling, U., Rohde, M., Olsen, A., Normark, S., and Reinkoster, J. 2000. AgfD, the checkpoint of multicellular and aggregative behaviour in *Salmonella typhimurium* regulates at least two independent pathways. *Mol. Microbiol.* **36:** 10–23.

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., Braun, S., de Wroom, E., van der Marel, G.A., van Boom, J.H., et al. 1987. Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325:** 279–281.

Ross, P., Mayer, R., Weinhouse, H., Amikam, D., Huggirat, Y., Benziman, M., de Vroom, E., Fidder, A., de Paus, P., Sliedregt, L.A., et al. 1990. The cyclic diguanylic acid regulatory system of cellulose synthesis in *Acetobacter xylinum*. Chemical synthesis and biological activity of cyclic nucleotide dimer, trimer, and phosphothioate derivatives. *J. Biol. Chem.* **265:** 18933–18943.

Ross, P., Mayer, R., and Benziman, M. 1991. Cellulose biosynthesis and function in bacteria. *Microbiol. Rev.* **55:** 35–58.

Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. 1998. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci.* **95:** 5857–5864.

Shapiro, L., McAdams, H.H., and Losick, R. 2002. Generating and exploiting polarity in bacteria. *Science* **298:** 1942–1946.

Simon, R., Prieffer, U., and Puhler, A. 1983. A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in gram negative bacteria. *Biotechnology* **1:** 784–790.

Smith, S.O., Kawakami, T., Liu, W., Ziliox, M., and Aimoto, S. 2001. Helical structure of phospholamban in membrane bilayers. *J. Mol. Biol.* **313:** 1139–1148.

Sommer, J.M. and Newton, A. 1991. Pseudoreversion analysis indicates a direct role of cell division genes in polar morphogenesis and differentiation in *Caulobacter crescentus*. *Genetics* **129:** 623–630.

Spiers, A.J., Kahn, S.G., Bohannon, J., Travisano, M., and Rainey, P.B. 2002. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* **161:** 33–46.

Stenkamp, R.E., Sieker, L.C., Jensen, L.H., and McQueen Jr., J.E. 1978. Structure of methemerythrin at 2.8-Angstrom resolution: Computer graphics fit of an averaged electron density map. *Biochemistry* **17:** 2499–2504.

Tal, R., Wong, H.C., Calhoon, R., Gelfand, D., Fear, A.L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., et al.. 1998. Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: Genetic organization and occurrence of conserved domains in isoenzymes. *J. Bacteriol.* **180:** 4416–4425.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. 2001. The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29:** 22–28.

Terwilliger, N.B. 1998. Functional adaptations of oxygen-transport proteins. *J. Exp. Biol.* **201:** 1085–1098.

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

Wheeler, R. and Shapiro, L. 1999. Differential localization of two histidine kinases controlling bacterial cell differentiation. *Mol. Cell* **4:** 683–694.

Zhulin, I.B., Nikolskaya, A.N., and Galperin, M.Y. 2003. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. *J. Bacteriol.* **185:** 285–294.

Appendix C

Curriculum Vitae

CARMEN CHAN

Personal details

Address: Biozentrum, Uni Basel, Klingelbergstr. 50-70, CH 4056, Basel
Phone: +41 61 267 2092
Email: carmen.chan@unibas.ch
Nationality: British
Date of birth: Aug 2, 1977

Education

PhD Biophysics, Prof T Schirmer Group, Dept of Structural Biology.
2000–Present, Biozentrum, University of Basel, Switzerland

Thesis: Structural elucidation of the multidomain response regulator PleD using X-ray crystallography.

In collaboration with: Prof U Jenal Group (Microbiology, Biozentrum) and Prof B Giese (Chemistry, Uni Basel).

MPhil Biochemistry, Dr I T Arkin Group.
1999–2000, University of Cambridge, Darwin College, UK

Thesis: Study of transmembrane helix packing using experimental and computational approaches.

BSc (Hons) Biochemistry (2:2)
1996–1999, Imperial College, University of London, UK

Funded summer research student, Dr M R K Alley Group, 1999

GCE A levels: Maths A, Chinese A, Physics A, Chemistry B
1989–1996, St Paul's Co-educational College, Hong Kong

Skills employed

Bioinformatics: data mining in gene and protein databases, comparative analyses of sequences and structures.

Molecular biology: cloning, protein expression.

Biochemistry: protein purification, biochemical and biophysical characterisation, enzymatic assays.

Structural biology: crystallisation of protein-ligand complexes, data collection (in-house and at SLS synchrotron), X-ray crystallography, structure elucidation and analyses.

Computing skills

Working knowledge on the crystallographic packages (CCP4 Suite, O), molecular display (DINO) and bioinformatics softwares.

Publications

Chan, C, Paul, R, Samoray, D, Amiot, N, Giese, B, Jenal, U and Schirmer, T. Structural basis of activity and allosteric control of di-guanylate cyclase. **Proc Natl Acad Sci USA**, in press.

Paul, R, Weiser, S, Amiot, N, **Chan, C**, Schirmer, T, Giese, B and Jenal, U. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. **Genes Dev.** (2004) 18(6):715-27

PDB Deposition

Coordinates of the non-activated PleD structure in complex with product molecules were deposited under the code 1W25 at Protein Data Bank.

Conference talks	Biozentrum Symposium, Basel, 2004. 18^{th} Regional Meeting on X-ray Crystallography of Biomacromolecules, Switzerland, 2004.
Poster presentations	Structural Biology Conference, EMBL Heidelberg, Germany, 2004. European Crystallography Meeting 22, Hungary, 2004.
Attended Workshop	EMBO Practical Course on 'Protein Expression, Purification and Crystallisation', EMBL Hamburg, Germany, 2002.
Teaching	Instructing and assessing final year undergraduates in small group practicals. Updating teaching materials and study examples.
Supervision	Supervising a lab technician to assist in research project, 2003-Present.
Extra-curricular activities	**Participant of Novartis' 'Women into Industry' programme** Novartis Pharmaceuticals, Switzerland, 2002-2003. **Attendant of intensive German courses** Switzerland, 2001-Present. Obtained Grade 1 in the Goethe Institut's 'Zertifikat Deutsch' exam, 2003.
Languages	Fluent English, Mandarin and Cantonese; conversational German.
References	
PhD supervisor	Prof T Schirmer: tilman.schirmer@unibas.ch
Collaborator	Prof U Jenal: urs.jenal@unibas.ch

Bibliography

[1] P. Aldridge and U. Jenal. Cell cycle-dependent degradation of a flagellar motor component requires a novel-type response regulator. *Mol Microbiol*, 32:379–91, 1999.

[2] I. Baikalov, I. Schroder, M. Kaczor-Grzeskowiak, D. Cascio, R. Gunsalus, and R. Dickerson. NarL dimerization? Suggestive evidence from a new crystal form. *Biochemistry*, 37(11):3665–76, Mar 17 1998.

[3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28:235–242, 2000.

[4] C. Birck, L. Mourey, P. Gouet, B. Fabry, J. Schumacher, P. Rousseau, D. Kahn, and J. P. Samama. Conformational changes induced by phosphorylation of the FixJ receiver domain. *Structure Fold Des*, 7(12):1505–15, Dec 15 1999.

[5] A. G. Blanco, M. Sola, F. X. Gomis-Ruth, and M. Coll. Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure*, 10(5):701–13, May 2002.

[6] B. Boeckmann, A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, and M. Schneider. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*, 31(1):365–70, Jan 1 2003.

[7] H. S. Cho, S. Y. Lee, D. Yan, X. Pan, J. S. P. S. Kustu, D. E. Wemmer, and J. G. Pelton. NMR structure of activated CheY. *J Mol Biol*, 297(3):543–51, Mar 31 2000.

[8] K. D. Cowtan and K. Y. Zhang. Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol.*, 72(3):245–70, 1999.

[9] T. E. Creighton. *Protein Structure: A Practical Approach*. IRL Press, second edition, 2002.

[10] D. A. D'Argenio and S. I. Miller. Cyclic di-GMP as a bacterial second messenger. *Microbiology*, 150(8):2497–502, Aug 2004.

[11] E. de la Fortelle and G. Bricogne. Maximum likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol*, 276:472–93, 1997.

[12] S. Djordjevic and A. M. Stock. Structural analysis of bacterial chemotaxis proteins: components of a dynamic signaling system. *J Struct Biol*, 124(2-3):189–200, Dec 15 1998.

[13] S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, and T. Ellenberger. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*, 391(6664):251–8, Jan 15 1998.

[14] G. D. V. Duyne, R. F. Standaert, P. A. Karplus, S. L. Schreiber, and J. C. J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol*, 229(1):105–24., Jan 1993.

[15] H. Edelhoch. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, 6(1948 - 1954), 1967.

[16] M. Egli, R. Gessner, L. Williams, G. Quigley, G. van der Marel, J. van Boom, A. Rich, and C. Frederick. Atomic-resolution structure of the cellulose synthase regulator cyclic diguanylic acid. *Proc Natl Acad Sci U S A*, 8(97):3235–9, April 1990.

[17] A. M. Eldridge, H. S. Kang, E. Johnson, R. Gunsalus, and F. W. Dahlquist. Effect of phosphorylation on the interdomain interaction of the response regulator, NarL. *Biochemistry*, 41(51):15173–80, Dec 24 2002.

[18] M. Y. Galperin. Bacterial signal transduction network in a genomic perspective. *Environ Microbiol*, 6(6):552–67, Jun 2004.

[19] K. E. Galperin MY, Nikolskaya AN. Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol Lett*, 203(1):11–21, Sep 11 2001.

[20] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31:3784–3788, 2003.

[21] S. C. Gill and P. von Hippel. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, 182(2):319–26, Nov 1 1989.

[22] P. Gouet, E. Courcelle, D. I. Stuart, and F. Metoz. ESPript: multiple sequence alignments in PostScript. *Bioinformatics*, 15:305–8, 1999.

[23] G. B. Hecht and A. Newton. Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in *Caulobacter crescentus*. *J Bacteriol*, 177(21):6223–9, Nov 1995.

[24] A. Hoffmann and R. G. Roeder. Purification of his-tagged proteins in non-denaturing conditions suggests a convenient method for protein interaction studies. *Nucleic Acids Res*, 19(22):6337-8., Nov 25 1991.

[25] S. J. Hubbard and J. M. Thornton. 'NACCESS', Computer Program, 1993.

[26] U. Jenal. Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria? *Curr Opin Microbiol*, 7(2):185–91, Apr 2004.

[27] T. A. Jones and S. Thirup. Using known substructures in protein model building and crystallography. *EMBO J.*, 5(4):819–22, Apr 1986.

[28] W. Kabsch and C. Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–637, Dec 1983.

[29] D. Kern, B. F. Volkman, P. Luginbuhl, M. J. N. S. Kustu, and D. E. Wemmer. Structure of a transiently phosphorylated switch in bacterial signal transduction. *Nature*, 402(6764):894–8, Dec 23-30 1999.

[30] G. J. Kleywegt and T. A. Jones. *From First Map to Final Model*, chapter Halloween ... masks and bones., pages 59–66. SERC Daresbury Laboratory, Warrington, 1994.

[31] U. K. Laemmli. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(259):680–5, Aug 15 1970.

[32] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Cryst*, 26:283–291, 1993.

[33] J. Lebowitz, M. S. Lewis, and P. Schuck. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Science*, 11(9):2067–2079, September 2002.

[34] B. Lee and F. M. Richards. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*, 55(3):379–400, Feb 1971.

[35] S. Y. Lee, H. S. Cho, J. G. Pelton, D. Yan, E. A. Berry, and D. E. Wemmer. Crystal structure of activated CheY. Comparison with other activated receiver domains. *J Biol Chem*, 276(19):16425–31, May 11 2001.

[36] S. Y. Lee, H. S. Cho, J. G. Pelton, D. Yan, R. K. Henderson, D. S. King, L. Huang, E. A. B. S Kustu, and D. E. Wemmer. Crystal structure of an activated response regulator bound to its target. *Nat Struct Biol*, 8(1):52–6., Jan 2001.

[37] I. Letunic, R. R. Copley, S. Schmidt, F. D. Ciccarelli, T. Doerks, J. Schultz, C. P. Ponting, and P. Bork. SMART 4.0: towards genomic data integration. *Nucleic Acids Research*, 32:Database issue D142–D144, 2004.

[38] R. J. Lewis, J. A. Brannigan, K. Muchova, I. Barak, and A. J. Wilkinson. Phosphorylated aspartate in the structure of a response regulator protein. *J Mol Biol*, 294(1):9–15, Nov 19 1999.

[39] Y. C. Liaw, Y. G. Gao, H. Robinson, G. M. Sheldrick, L. A. Sliedregt, G. A. van der Marel, J. H. van Boom, and A. H. Wang. Cyclic diguanylic acid behaves as a host molecule for planar intercalators. *FEBS Lett*, 264(2):223–7, May 21 1990.

[40] G. G. Lu. Topp–an automatic topological and atomic comparison program for protein structures. *CCP4 suite*, 1996.

[41] A. E. Maris, M. R. Sawaya, M. Kaczor-Grzeskowiak, M. R. Jarvis, S. M. Bearson, M. L. Kopka, I. Schroder, R. P. Gunsalus, and R. E. Dickerson. Dimerization allows DNA target site recognition by the NarL response regulator. *Nat Struct Biol*, 9(10):771–8, Oct 2002.

[42] B. W. Matthews. Solvent content of protein crystals. *J Mol Biol*, 33(2):491–7, Apr 28 1968.

[43] J. P. Morth, V. Feng, L. J. Perry, D. I. Svergun, and P. A. Tucker. The crystal and solution structure of a putative transcriptional antiterminator from *Mycobacterium tuberculosis*. *Structure*, 12(9):1595–605, Sep 2004.

[44] C. C. P. Number. The CCP4 suite: Programs for Protein Crystallography. *Acta Crystallogr D Biol Crystallogr*, 50(Pt 5):760–763, Sep 1994.

[45] R. Paul, S. Weiser, N. C. Amiot, C. Chan, T. Schirmer, B. Giese, and U. Jenal. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev*, 18(6):715–27, Mar 15 2004.

[46] J. Pei and Grishin. GGDEF domain is homologous to adenylyl cyclase. *Proteins*, 42(2):210–6, Feb 1 2001.

[47] D. N. Perkins, D. J. Pappin, D. M. Creasy, and J. J. S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–3567, 1999.

[48] A. Philippsen. DINO: Visualizing Structural Biology, 2002.

[49] G. Ralston. *Introduction to Analytical Ultracentrifugation*. Beckman Instruments Inc., Fullerton, CA, 1993.

[50] V. L. Robinson, D. R. Buckler, and A. M. Stock. A tale of two components: a novel kinase and a regulatory switch. *Nat Struct Biol*, 7(8):626–33, Aug 2000.

[51] V. L. Robinson, T. Wu, and A. M. Stock. Structural analysis of the domain interface in DrrB, a response regulator of the OmpR/PhoB subfamily. *J Bacteriol*, 185(14):4186–94, Jul 2003.

[52] P. Ross, R. Mayer, and M. Benziman. Cellulose biosynthesis and function in bacteria. *Microbiol Rev*, 55(1):35–58, Mar 1991.

[53] P. Ross, R. Mayer, H. Weinhouse, D. Amikam, Y. Huggirat, M. Benziman, E. de Vroom E, A. Fidder, P. de Paus, and L. A. Sliedregt. The cyclic diguanylic acid regulatory system of cellulose synthesis in

Acetobacter xylinum. chemical synthesis and biological activity of cyclic nucleotide dimer, trimer, and phosphothioate derivatives. *J Biol Chem*, 265(31):18933–43, Nov 5 1990.

[54] P. Ross, H. Weinhouse, Y. Aloni, D.Michaeli, P. Weinberger-Ohana, R. Mayer, S.Braun, E. de Vroom, G. A. van Marel, J. H. van Boom, and M. Benizman. Regulation of cellulose synthesis in *Acetobacter xylinium* by cyclic diguanylic acid. *Nature*, 325:279–81, January 1987.

[55] T. R. Schneider and G. M. Sheldrick. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr*, 58(Pt 10 Pt 2):1772–9, Oct 2002.

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

[57] A. J. Spiers, J. Bohannon, S. M. Gehrig, and P. B. Rainey. Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Mol Microbiol*, 50(1):15–27, Oct 2003.

[58] A. Stock, J. Mottonen, J. Stock, and C. Schutt. Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature*, 337(6209):745–9, Feb 23 1989.

[59] A. M. Stock, E. Martinez-Hackert, B. F. Rasmussen, A. H. W. J. B. Stock, and D. R. G. Petsko. Structure of the Mg(2+)-bound form of CheY and mechanism of phosphoryl transfer in bacterial chemotaxis. *Biochemistry*, 32(49):13375–80, Dec 1993.

[60] G. Taylor. The phase problem. *Acta Crystallogr D Biol Crystallogr*, 59(Pt 11):1881–90, Nov 2003.

[61] J. Tesmer, R. K. Sunahara, R. A. Johnson, G. Gosselin, A. G. Gilman, and S. R. Sprang. Two-metal-ion catalysis in adenylyl cyclase. *Science*, 285(5428):756–60, Jul 30 1999.

[62] J. J. Tesmer and S. R. Sprang. The structure, catalytic mechanism and regulation of adenylyl cyclase. *Curr Opin Struct Biol*, 8(6):713–9, Dec 1998.

[63] J. D. Thompson, D. G. Higgins, and T. J. Gibson. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22):4673–80, 1994.

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

[65] S. Uzzau and A. Fasano. Cross-talk between enteric pathogens and the intestine. *Cell Microbiol*, 2(2):83–9, Apr 2000.

[66] F. M. D. Vellieux and R. J. Read. Non-crystallographic symmetry averaging in phase refinement and extension. *Methods in Enzymology*, 277:18–53, 1997.