Solution Structure of the PilZ Domain Protein PA4608 Complex with Cyclic di-GMP Identifies Charge Clustering as Molecular Readout*⒮◆

Judith Habazettl[1], **Martin G. Allan**[1,2], **Urs Jenal**[3], and **Stephan Grzesiek**[4]

From the Biozentrum, University of Basel, CH-4056 Basel, Switzerland

Cyclic diguanosine monophosphate (c-di-GMP) is a ubiquitous bacterial second messenger that controls the switch from a single-cell lifestyle to surface-attached, multicellular communities called biofilms. PilZ domain proteins are a family of bacterial c-di-GMP receptors, which control various cellular processes. We have solved the solution structure of the *Pseudomonas aeruginosa* single-domain PilZ protein PA4608 in complex with c-di-GMP by NMR spectroscopy. Isotope labeling by ^{13}C and ^{15}N of both the ligand and the protein made it possible to define the structure of c-di-GMP in the complex at high precision by a large number of intermolecular and intraligand NOEs and by two intermolecular hydrogen bond scalar couplings. Complex formation induces significant rearrangements of the C- and N-terminal parts of PA4608. c-di-GMP binds as an intercalated, symmetric dimer to one side of the β-barrel, thereby displacing the C-terminal helix of the apo state. The N-terminal R*XXX*R PilZ domain motif, which is flexible in the apo state, wraps around the ligand and in turn ties the displaced C terminus in a loose manner by a number of hydrophobic contacts. The recognition of the dimeric ligand is achieved by numerous H-bonds and stacking interactions involving residues Arg^8, Arg^9, Arg^{10}, and Arg^{13} of the PilZ motif, as well as β-barrel residues Asp^{35} and Trp^{77}. As a result of the rearrangement of the N and C termini, a highly negative surface is created on one side of the protein complex. We propose that the movement of the termini and the resulting negative surface form the basis for downstream signaling.

The second messenger cyclic diguanosine monophosphate (c-di-GMP)[5] is widespread in and unique to the bacterial kingdom. Elevated intracellular levels of c-di-GMP generally cause bacteria to change from a motile single-cell state to an adhesive surface-attached multicellular state called biofilm (1–3). Moreover, c-di-GMP controls the virulence of animal and plant pathogens (4–8), progression through the cell cycle (9), antibiotic production (10), and other cellular functions. In several pathogenic bacteria, c-di-GMP has been associated with the regulation of virulence factors and the development of chronic infections. During long term cystic fibrosis lung infections, increased persistence of *Pseudomonas aeruginosa* correlates with the generation of adaptive colony morphotypes. This includes mucoid colonies and small colony variants, auto-aggregative, hyper-adherent cells whose appearance correlates with poor lung function and persistence of infection. Both morphotypes are strongly linked to elevated levels of c-di-GMP, implicating a central role for this second messenger in chronic *P. aeruginosa* infections (8, 11–14). These studies emphasize the clinical significance of c-di-GMP and stress the importance for understanding c-di-GMP signaling in the pseudomonads.

c-di-GMP comprises two guanosine monophosphate molecules that are symmetrically connected by 5′-3′ phosphodiester bonds, forming a 12-membered ring (see Fig. 1*A*). It is synthesized from two molecules of guanosine triphosphate (GTP) by diguanylate cyclases (GGDEF-containing domains) and hydrolyzed into the linear dinucleotide 5′-phosphoguanylyl-(3′-5′)-guanosine by the activity of c-di-GMP-specific phosphodiesterases (EAL-containing domains) (15). The opposing activities of diguanylate cyclases and phosphodiesterases control the intracellular c-di-GMP concentrations, and hence c-di-GMP signaling. Several classes of c-di-GMP effector proteins have been identified so far, including PilZ domain proteins, response regulators, and degenerate GGDEF proteins that have lost catalytic activity (9, 16–19). The PilZ protein family (named after the type IV pilus control protein in *P. aeruginosa*) represents the best studied class of effectors. Members of this family have been implicated in a diverse range of cellular functions including exopolysaccharide biosynthesis, flagellar motor activity, and virulence gene expression. However, little information is available about the mechanistic details involved in c-di-GMP-mediated PilZ activation of these cellular processes.

Several structures of PilZ domain proteins have been solved in their apo and holo forms. One is the two-domain *Vibrio cholerae* protein VCA0042. A comparison of apo VCA0042 with its holo structure (20) shows that binding of a single c-di-

GMP molecule induces a conformational change in the loop connecting the C-terminal PilZ domain to the N-terminal YcgR-N domain. This brings the two domains into close proximity, thereby forming a new allosteric interaction surface with c-di-GMP at their mutual interface. It has been shown recently that YcgR, a homolog of VCA0042 in *Escherichia coli*, is able to bind to the flagellar motor in its c-di-GMP-bound form (21). Although VCA0042 appears to bind only one molecule of c-di-GMP and does not alter its quaternary structure upon ligand binding, PP4397 from *Pseudomonas putida* binding of two molecules of c-di-GMP in the junction of its YcgR-N and PilZ domains results in a dimer-monomer transition (22). This suggested that different PilZ domain proteins exhibit distinct binding stoichiometries and mechanistic principles.

All structures of c-di-GMP protein complexes solved until today involve multidomain proteins (for review, see Ref. 15). In these structures, the binding of c-di-GMP within a hinge-like region induces an allosteric rearrangement of the neighboring domains via fixation to the ligand. It was proposed that this c-di-GMP-induced domain immobilization represents a general mechanism for signal transduction. In contrast to these multidomain proteins, the PilZ domain PA4608 from *P. aeruginosa* is a single domain c-di-GMP-binding protein. The apo solution structure of PA4608 was solved by Ramelot *et al.* (23), who also confirmed binding of c-di-GMP by chemical shift analysis. NMR chemical shift mapping was then used to localize the ligand binding site of PA4608 to one side of the protein surface and to demonstrate that c-di-GMP binds to PA4608 as a dimer (16). Furthermore, these analyses demonstrated that PA4608 remains monomeric after ligand binding (16).

Here we report the structure of the complex of PA4608 with c-di-GMP solved by advanced solution NMR techniques. Uniform ^{13}C, ^{15}N isotope labeling of both c-di-GMP and PA4608 allowed the observation of a large number of intermolecular and intraligand NOE contacts, $^{1}D_{HN}$ and $^{1}D_{CH}$ ligand and protein residual dipolar couplings (RDCs), and intermolecular hydrogen bond scalar couplings. Both termini of the protein undergo large structural changes upon ligand binding while retaining a certain amount of flexibility. In particular, the C-terminal helix is displaced by the ligand, thereby opening one side of the β-barrel as a binding site, and the N terminus containing the R*XXX*R PilZ domain motif wraps around the ligand and in turn ties the C-terminal helix in a loose conformation induced by a number of hydrophobic contacts. The rearranged termini expose a highly negatively charged surface on one side of the complex to a possible effector protein. This induced folding of the flexible N terminus containing the PilZ signature R*XXX*R and the resulting clustering of surface charges may present a general mechanism for c-di-GMP readout by PilZ domains.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—Overexpression, ^{13}C, ^{15}N labeling, and purification of PA4608 were achieved as described previously (16). In the protein construct used, the PA4608 sequence is preceded by a hexahistidine tag with the sequence MGSSHHHHHHSSGLVPRGSH. c-di-GMP was produced enzymatically from GTP by using a non-feedback-inhibited mutant of the diguanylate cyclase DgcA named DgcA0646,

in which the inhibition site RESD is replaced by GRDC (24, 25). Uniformly ^{13}C- and ^{15}N-labeled c-di-GMP was produced in the same manner but starting with ^{13}C-, ^{15}N-labeled GTP (Spectra Stable Isotopes) as a precursor.

NMR Samples—Uniformly $^{13}C,^{15}N$- and ^{15}N-labeled samples of apo PA4608 (0.8–1.2 mM) were prepared in 250 mM NaCl, 10 mM Tris-HCl at pH 6.7, 0.01% NaN_3 (w/v), and 5% (v/v) D_2O as sample volumes of 200–400 μl. For complex formation, c-di-GMP was titrated to the apo PA4608 samples from a 20 mM stock solution, and the fraction of apo- and ligand-bound protein was monitored by ^{1}H-^{15}N heteronuclear single quantum correlation spectra. The titration was stopped at two equivalents of c-di-GMP, *i.e.* when the protein was saturated. For preparation of D_2O samples, the H_2O samples were lyophilized and redissolved in D_2O twice. Non-isotropic samples of complexed (apo) protein were prepared by adding 18 mg/ml (26 mg/ml) filamentous phage Pf1 (Asla Biotech).

NMR Spectroscopy and Resonance Assignments—NMR spectra were recorded at 293 K on Bruker DRX 600 and DRX 800 NMR spectrometers equipped with TXI and TCI probe heads, respectively. For protein assignment, structure information, and assessment of backbone dynamics by ^{15}N relaxation, standard two-dimensional and three-dimensional NMR spectra were acquired similar to the ones described (26). c-di-GMP ribose resonances were connected by three-dimensional HCCH-COSY and -TOCSY spectra recorded on $^{13}C/^{15}N$-labeled c-di-GMP bound to $^{13}C/^{15}N$-labeled PA4608. Ribose moieties and guanine H1/N1 and H8/C8 resonances were assigned via their NOE contacts. Details of the acquired spectra are indicated in the supplemental Table S1. NMR data were processed using the NMRPipe suite of programs (27). Spectra were displayed and analyzed with the programs SPARKY (28) and PIPP (29).

Structure Calculations—Structure calculations were performed with the program XPLOR-NIH (30) using two consecutive simulated annealing protocols (31). In a first step, calculations started from an extended strand with c-di-GMP bound to the protein via the measured H-bond between Hε1 of Trp[77] and N7 of Gua1, using intraprotein restraints only; the second step included all restraints. c-di-GMP molecules were defined as two-residue, circular strands of RNA. The 20-residue His tag and linker were not included in structure calculations. A total of 200 structures were calculated, and the 20 lowest energy structures were selected for deposition (Protein Data Bank (PDB) code 2L74). The structural statistics are given in Table 1.

RESULTS AND DISCUSSION

NMR Spectroscopy, Resonance Assignments, and Structure Determination—Standard triple-resonance NMR spectra recorded on PA4608 saturated with c-di-GMP were used to assign backbone and side chain resonances of PA4608. Resonance assignments were obtained for about 92% of the protein backbone in the PA4608·c-di-GMP complex. Missing backbone assignments comprise parts of the N-terminal His tag sequence as well as residues His[12] and Asp[108]–Arg[114]. In almost all of these cases, resonances were not visible in the ^{1}H-^{15}N heteronuclear single quantum correlation spectra, presumably

Solution Structure of the PilZ Protein PA4608 in Complex with c-di-GMP

due to chemical exchange broadening resulting from flexibility on the microsecond to millisecond time scale.

The chemical shifts of c-di-GMP bound to PA4608 are very distinct from free c-di-GMP. In particular, four different imino H1 proton resonances are observed, indicative of four guanosines in an asymmetric environment. This is consistent with the binding of one c-di-GMP dimer to one protein monomer. Of these 4 guanosine residues, all one-bond ^{1}H1/^{15}N1 and ^{1}H/^{13}C resonance pairs could be assigned with the exception of Gua1-H5″ and Gua4-H5″ by using information from three-dimensional HCCH-TOCSY/-COSY and ^{15}N- or ^{13}C-edited NOESY experiments. For both Gua2 and Gua3, one of the two N2-bound protons could also be clearly observed. In free guanosines, these proton resonances are usually strongly broadened from the rotation around the C2–N2 bond. The observation of the resonances together with the measured NOEs indicates H-bond formation between the intercalated dimers in the following way (Fig. 1B): Gua1-O1P⋯H1–N1–Gua3, Gua1-O1P⋯H21–N2–Gua3, Gua4-O1P⋯H1–N1–Gua2, and Gua4-O1P⋯H21–N2–Gua2. Two additional H-bonds between c-di-GMP and the protein, Gua1-N7⋯Hε1–Nε1-Trp[77] and Gua2-N7⋯Hε—Nε-Arg[9], were detectable by the HNN-COSY experiment (32) (Fig. 1C) and are indicated in Fig. 1B as red dashed lines. The position of the c-di-GMP dimer on the protein is well defined by 103 intermolecular NOEs. A small section of the two-dimensional NOESY spectrum (Fig. 1C) shows some of these NOEs from the four imino protons of the c-di-GMP dimer and the Trp[77] Hε1 proton, indicating contacts between the ligand and protein residues Arg[8], Arg[9], Phe[11], Arg[13], Arg[12], Ile[41], Leu[42], Ala[76], and Trp[77]. The c-di-GMP dimer itself is well defined by 59 intramolecular NOEs and a further 15 ^{1}D$_{CH}$ RDCs. Structural restraints of the protein comprise 1734 intramolecular NOEs and 152 ^{1}D$_{HN}$ and ^{1}D$_{CH}$ RDCs.

Based on these restraints, a well defined structure of the PA4608·c-di-GMP complex was calculated using conventional simulated annealing protocols. Its quality is reflected in a heavy atom coordinate r.m.s.d. of 0.20 Å for the backbone in the ordered region of the protein and of 0.38 Å for the c-di-GMP dimer. Further structural statistics are given in Table 1.

Ligand-induced Changes in the Backbone Structure of PA4608—As compared with apo PA4608, the complexed protein shows strong chemical shift changes in the N-terminal region around residue Arg[9], in the C-terminal region around residue Val[120], as well as in the central β-barrel around residue Val[36] (16). An analysis of the secondary structure (Fig. 2) and backbone dynamics (see below) shows that structural changes mainly occur in the N- and C-terminal regions, whereas the central β-barrel is largely unaffected. Ligand binding induces a transition of the N-terminal region from an unfolded state to an ordered structure and a strong destabilization of the C-terminal region in particular between residues Asp[108] and Arg[114]. For residues Glu[110] to Arg[114], no resonances can be detected at all due to exchange broadening, indicating the disappearance of the apo 3$_{10}$ helix γ$_1$ (Glu[109]–Glu[113]). Residues Glu[115] up to the C terminus are again observable but correspond to an α-helix (α$_2$) between residues Ala[117] and Ser[121] rather than to the apo 3$_{10}$ helix γ$_2$ (Leu[116]–Leu[118]).

Fig. 2 shows a sequence and secondary structure alignment of PA4608 with the other PilZ domains, for which structures complexed with c-di-GMP have been solved, *i.e. P. putida* PP4397 (22) and *V. cholerae* VCA0042 (20). In addition, *Caulobacter crescentus* DgrA and DgrB and *Salmonella typhimurium* YcgR have been included because they have similar nanomolar affinities for c-di-GMP as PA4608, and key residues for ligand binding are identified by point mutations (16). Despite a rather low sequence similarity, the β-barrel and helix α$_1$ secondary structure elements of the solved structures are largely conserved. Most importantly, in all aligned proteins, the PilZ domain signature motifs R*XXX*R and D*XSXX*G are completely conserved apart from a single amino acid replacement (S/A) for DgrA. Residues that have van der Waals contacts to the ligand in the solved holo structures (*yellow*) are also largely conserved. These comprise the two R*XXX*R and D*XSXX*G motifs and residues around Trp[77] (PA4608) and Gly[84] (PA4608), which are located on adjacent β-strands. Together with the DVSLHG[40] region, the latter form the surface of the β-barrel that is in direct contact with the ligand. Not surprisingly, mutations of the conserved residues Arg[8], Arg[9], Asp[35], or Trp[77] (PA4608 numbering) strongly reduce or abolish c-di-GMP binding in DgrA (16).

Description of the PA4608·c-Di-GMP Complex Structure—The core structure of PA4608 in complex with c-di-GMP consists of a six-stranded antiparallel β-barrel with an elongated α-helix across one barrel opening (Fig. 3, A and B). The antiparallel β-strands are in the order 1-2-3-6-5-4 with 1 and 4 closing the barrel in an antiparallel manner. The intercalated c-di-GMP dimer is located at one side of the barrel. It makes contacts with residues of the barrel strands β5, β6, β3, and β2 as well as with side chains from the N-terminal part of the sequence, which is unstructured in the apo state. This N-terminal end wraps around the external side of the intercalated c-di-GMP dimer. The face of the N-terminal region, which points away from the ligand, makes contacts to the C-terminal region, thereby also anchoring helix α2 toward the ligand. Fig. 3, C and D, show bundle representations of the 20 lowest energy structures in a best fit superposition of the barrel backbone atoms and the ligand atoms, respectively. The high coordinate precision of the protein core, c-di-GMP, and the N terminus is evident. In contrast, the C-terminal region is less well defined, but clearly not random.

Details of the recognition of c-di-GMP dimer by PA4608 are shown in Fig. 4. The main residues involved are Trp[77] and Arg[8] through Arg[13]. Trp[77] recognizes the side of the Gua1 base by the ^{h2}J$_{NN}$-detected H-bond from its indole Nε1 donor to the Gua1 N7 acceptor (Fig. 1). This locates the Gua1 base in the same plane as the aromatic ring of Trp[77]. A further H-bond may be possible from the Gua1 H1 hydrogen to the side chain of Asp[35] but is not clearly detectable due to low definition of the Asp[35] side chain (not shown). The Gua1 base also packs against the side chain of Arg[13]. The pyrimidine ring of Gua3 stacks below the pyrrole ring of Trp[77]. The N7 and O6 atoms of Gua3 are recognized by two further H-bonds from the Arg[13] Nη1 and Nη2 atoms. Besides the structural proximity, evidence for these H-bonds comes from the observation of resonances and NOE contacts from the Nη1 and Nη2 atoms of Arg[13]. Usually these proton resonances are not detectable in free arginine side
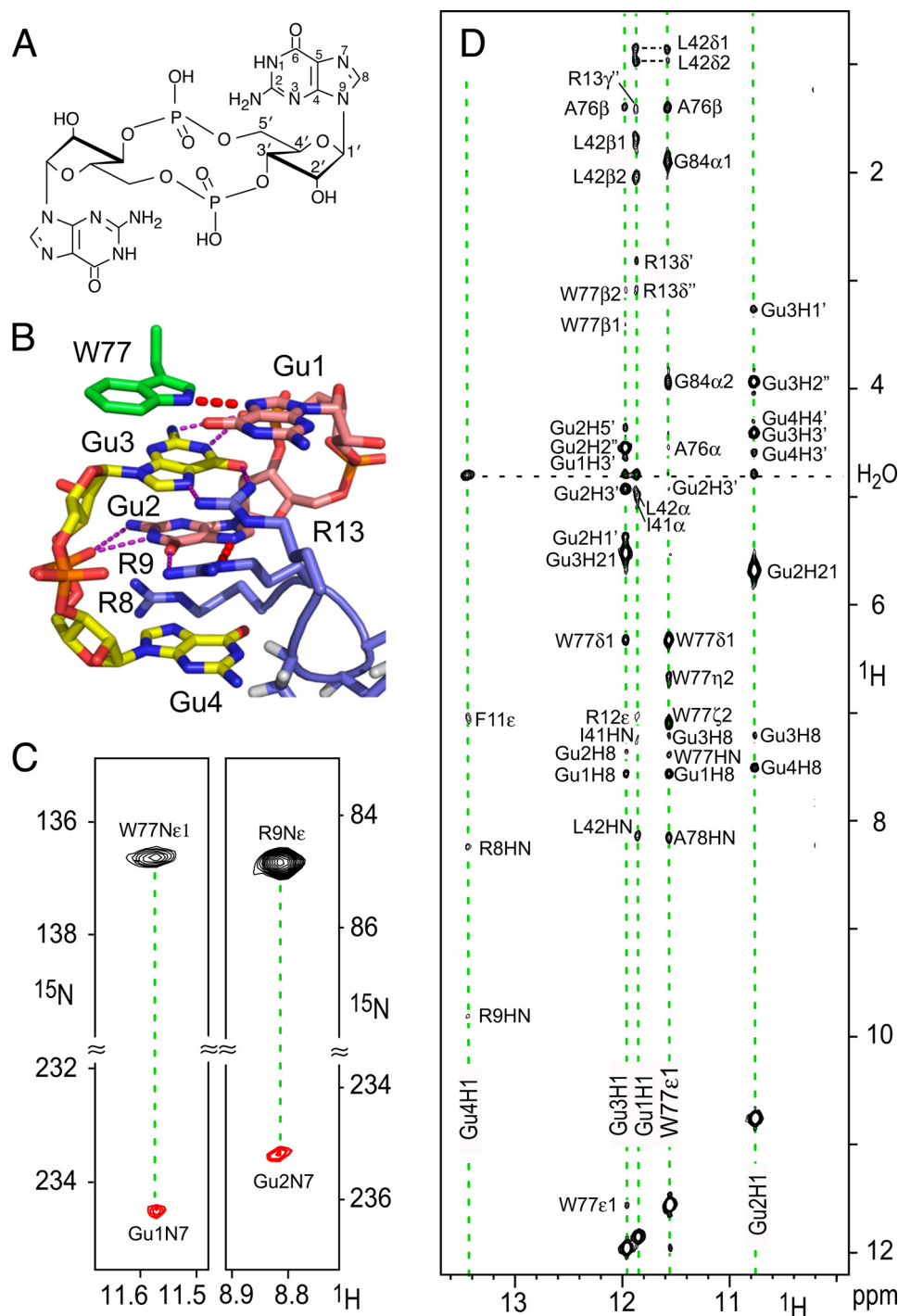
FIGURE 1. **Detection of the dimeric c-di-GMP structure and coordination in the complex with PA4608 by heteronuclear NMR methods.** *A*, the chemical structure c-di-GMP consisting of two GMP moieties linked by O3′-P phosphodiester bonds. *B*, the intercalated dimeric c-di-GMP as it is bound to PA4608. Because c-di-GMP comprises two guanosine monophosphate molecules that are symmetrically connected by 5′-3′ phosphodiester bonds to a single molecule, we kept the nomenclature of the atoms but numbered the guanosine moieties as guanosines 1 and 2 for the first monomer and guanosines 3 and 4 for the second monomer. For space reasons, guanosines are abbreviated as *Gu* in the figures. Hydrogen bonds, for which there is experimental evidence, are shown as *dashed lines*. Two of these result from direct detection via $^{h2}J_{NN}$ couplings (see *C*) and are shown by *thick red dashes*. *C*, extracted regions of two HNN-COSY spectra for direct detection of N–H···N H-bonds (32). The *red* (negative) signals are cross-peaks resulting from H-bond J-coupling transfer from the H-bond donor nucleus Trp[77]–^{15}Nε1 to the acceptor nucleus Gua1–^{15}N7 ($^{h2}J_{NN}$ = 4.7 ± 0.3 Hz) magnetization and from Arg[9]–^{15}Nε to Gua2–^{15}N7 ($^{h2}J_{NN}$ = 4.8 ± 0.4 Hz), respectively. *D*, extracted region from the ^{1}H two-dimensional NOESY spectrum showing the NOE contacts from the four imino protons H1 of guanines 1–4 and of Hε1 of Trp[77]. A large number of intermolecular c-di-GMP-to-protein as well as c-di-GMP intradimer contacts are observed that define the c-di-GMP coordination in the complex with PA4608.

chains due to chemical exchange broadening from the rotation around the Cζ–Nη1 and Cζ–Nη2 bonds. The imino N1 and the amino N2 groups of Gua3 are in H-bonding distance to the Gua1 phosphate group of the first c-di-GMP monomer.

The pyrimidine ring of Gua2 stacks below the pyrimidine ring of Gua3. Its imino N1 and the amino N2 groups are in H-bonding distance to the Gua4 phosphate group of the second c-di-GMP monomer. The N7 and O6 atoms of Gua2 are recog-

Solution Structure of the PilZ Protein PA4608 in Complex with c-di-GMP

TABLE 1
Assignment and structure statistics

The statistics comprise an ensemble of the final 20 simulated annealing structures. Individual structures were fitted to each other using residues in the β-sheets of the barrel: 19–24, 27–37, 40–44, 57–62, 68–79, and 82–90. The numbers of the various constraints are given in parentheses.

Assignment statistics	
Completeness of resonance assignments (%)	
Protein, backbone[a] (461/502)	91.8
Protein, all atoms[b] (1195/1342)	89.0
c-di-GMP[c] (58/68)	85.0
Structure statistics	
r.m.s.d. value from experimental distance restraints (Å)	
All (1951)	0.036 ± 0.001
Protein, intraresidue ($i = j$) (163)	0.017 ± 0.002
Protein, sequential ($|i − j| = 1$) (598)	0.031 ± 0.002
Protein, medium range ($1 < |i − j| \leq 5$) (348)	0.029 ± 0.003
Protein, long range ($|i − j| > 5$) (625)	0.042 ± 0.002
Protein to ligand[d] (103)	0.058 ± 0.006
Within ligand (59)	0.021 ± 0.005
Hydrogen bonds[e] (55)	0.033 ± 0.002
r.m.s.d. values from RDCs (Hz)	
Protein[f] (152)	0.14 ± 0.01
Ligand[g] (15)	0.15 ± 0.03
r.m.s.d. values from TALOS torsion angle restraints[h] (°) (169)	0.88 ± 0.05
r.m.s.d. values from experimental dihedral restraints[i] (°) (16)	0.16 ± 0.17
Deviations from idealized covalent geometry[j]	
Bonds (Å)	0.003 ± 0.0001
Angles (°)	0.65 ± 0.01
Impropers (°)	0.57 ± 0.01
Coordinate precision[k] (Å)	
Protein backbone non-hydrogen atoms	0.20 ± 0.05
Protein all non-hydrogen atoms	0.88 ± 0.07
c-di-GMP non-hydrogen atoms	0.58 ± 0.18
c-di-GMP non-hydrogen atoms[l]	0.38 ± 0.14
PROCHECK quality indicators[m]	
In most favored regions of Ramachandran plot (%)	86.9
In additional allowed regions of Ramachandran plot (%)	11.7
In generously allowed regions of Ramachandran plot (%)	1.3
In disallowed regions of Ramachandran plot (%)	0.1
χ_1 pooled S.D. (°)	22.9
χ_2 pooled S.D. (°)	22.5

[a] Considering $^1H^N$, ^{15}N (except Pro), $^{13}C^\alpha$, and $^1H^\alpha$ resonances of residues Met[1]–Asp[125]. No resonances were assigned for residues Asp[108] through Arg[114]. The first numbers in parentheses are the assigned resonances, and the second numbers are all occurring resonances.

[b] Considering routinely assignable ^1H, ^{15}N, and ^{13}C resonances of residues Met[1]–Asp[125], excluding N-terminal and Lys amino groups, Arg guanidino groups, hydroxyl protons of Ser, Thr, Tyr, thiol protons of Cys, carboxyl resonances of Asp and Glu, non-protonated aromatic carbons, and Pro ^{15}N. ^1H belonging to the same methyl group and Phe, Tyr $^1H^\delta$, $^1H^\epsilon$ are counted as one.

[c] Considering H1′, C1′, H2′, C2′, H3′, C3′, H4′, C4′, H5′, H5″, C5′, H1, N1, N2, N7, H8, and C8. Two additional H21 resonances were observed due to hydrogen bonds.

[d] Derived from two-dimensional, isotope-filtered two-dimensional, and three-dimensional ^{13}C and ^{15}N-edited NOESY spectra.

[e] There are 49 hydrogen bonds within the protein, 4 within the ligand, and 2 directly measured between protein and ligand. For each backbone hydrogen bond constraint, there are two distance restraints: r_{NH-O}, 1.8–2.5Å, and r_{N-O}, 2.8–3.5Å.

[f] These comprise 147 RDCs from N-HN and C^α-H^α, and 5 RDCs from aromatic rings.

[g] 11 RDCs are from the ribose moieties and four of the guanine rings.

[h] Backbone torsion angle restraints were derived from $^{13}C^\alpha$ and $^{13}C^\beta$ chemical shifts using the program TALOS+ (43).

[i] χ_1 (15) and χ_2 (5) torsion angles were obtained from H^N-H^β ROE distances and $^3J_{H\alpha H\beta}$, $^3J_{NC\gamma}$, and $^3J_{C'C\gamma}$ couplings.

[j] The improper torsion restraints maintain planarity and chirality.

[k] The coordinate precision is the average r.m.s. difference between the individual simulated annealing structures and the mean coordinates. The calculation includes all residues that do not exhibit large amplitude internal motions as evidenced from the ^{15}N relaxation experiments. These consist of all residues from 16 to 103 with the exception of 18, 24, 25, 38, 39, 47–53, and 64–67. The latter are predominantly located in loops.

[l] From an optimal superposition of all non-hydrogen atoms of c-di-GMP.

[m] These values are calculated with the program PROCHECK-NMR (38). Values are reported for all non-mobile residues, i.e. Glu[7]–Phe[11], Arg[13]–Gly[107], and Leu[116]–Ser[121], excluding glycines and prolines.

nized by H-bonds from the Nε and the Nη2 atoms of Arg[9], respectively. The Arg[9]-Nε–Hε⋯N7–Gua2 H-bond is directly detected by $^{h2}J_{NN}$ couplings (Fig. 1), whereas again the close proximity and the detection of both Arg[9] Hη21 and Hη22 pro-

tons indicate the presence of the Arg[9]-Nη2-Hη21⋯O6-Gua2 H-bond. Finally, the base of the last guanosine Gua4 is packing against the side chain of Arg[9] and is turned away from the stacked Trp[77]/Gua3/Gua2 ring system. Nevertheless, all four guanine bases and the indole ring of Trp[77] are nearly coplanar.

The observed H-bond interactions of the N7 and O6 acceptors of Gua3 and Gua2 with the Nη and Nε donors of Arg[13] and Arg[9], respectively, are a very common mode of recognition of guanosines by arginines in protein-DNA complexes (33). A further common recognition motif is identified in the favorable cation-π interactions (34, 35) that can be assumed for the packing of Gua1 against Arg[13] and Gua4 against Arg[9]. The high degree of conservation of Arg[9] and Arg[13] (Fig. 2) is fully consistent with this observed key role in c-di-GMP dimer recognition. The side chains of Arg[8] and Arg[10] are not as well defined in the structure, and hence do not allow the identification of non-ambiguous H-bonds. However, the proximity of Arg[8] to the guanine ring of Gua4 and of Arg[10] to the negatively charged phosphate of Gua2 suggests that Arg[8] and Arg[10] are also involved in c-di-GMP recognition.

Residues Phe[11] and His[12] are in the center of the R*XXX*R[13] c-di-GMP recognition motif. Their backbone ^1H-^{15}N resonances are very weak (Phe[11]) or not observed at all (His[12]), indicating flexibility on the microsecond to millisecond time scale. Nevertheless, the side chains of Phe[11] and His[12] could be assigned, and 15 unambiguous NOEs are detected from these residues to the side chains of Leu[118] and Leu[119] in the C-terminal helix α2 (Fig. 4). Despite the apparently high structural flexibility in this region, these hydrophobic contacts clearly dock the C-terminal helix α2 onto the top of the N-terminal c-di-GMP recognition site. Thus, the R*XXX*R signature motif couples ligand binding directly to the reorientation of the C-terminal part of the protein.

Comparison of the PA4608 c-Di-GMP Binding Mode to Other Known Structures—Besides PA4608, PilZ structures have been solved in both apo and holo forms for two additional proteins, PP4397 (22) and VCA0042 (20). PP4397 consists of a C-terminal PilZ and an N-terminal YcgR domain. It dimerizes in its apo form via interactions between the R*XXX*R signature motif of one protomer and the PilZ domain of the other protomer. Binding of an intercalated dimer of c-di-GMP to the R*XXX*R motif in the linker region between the two PP4397 domains induces the separation of the protein dimer into monomers.

Fig. 5, *A* and *B*, show a comparison of the c-di-GMP dimer recognition in PA4608 and PP4397. The similarity of the c-di-GMP coordination is very strong but may not be surprising based on the high sequence conservation of the recognition motifs. Thus, (i) Arg[123] and Arg[127] in PP4397 recognize the bases of Gua2 and Gua3 by the same H-bonds as Arg[9] and Arg[13] in PA4608; (ii) the side chain of His[201] in PP4397 replaces the Trp[77] side chain in PA4608 to recognize Gua1 by H-bonds; and (iii) Arg[122] in PP4397 has well defined H-bonds to Gua4 that correspond to H-bonds from Arg[8] in PA4608 but are not well detected in the NMR structure due to the low definition of the side chain. Further similarities are (iv) the H-bond recognition of the Gua1 imino group by the side chain of Asp[157] in PP4397 and (v) the H-bond rec-
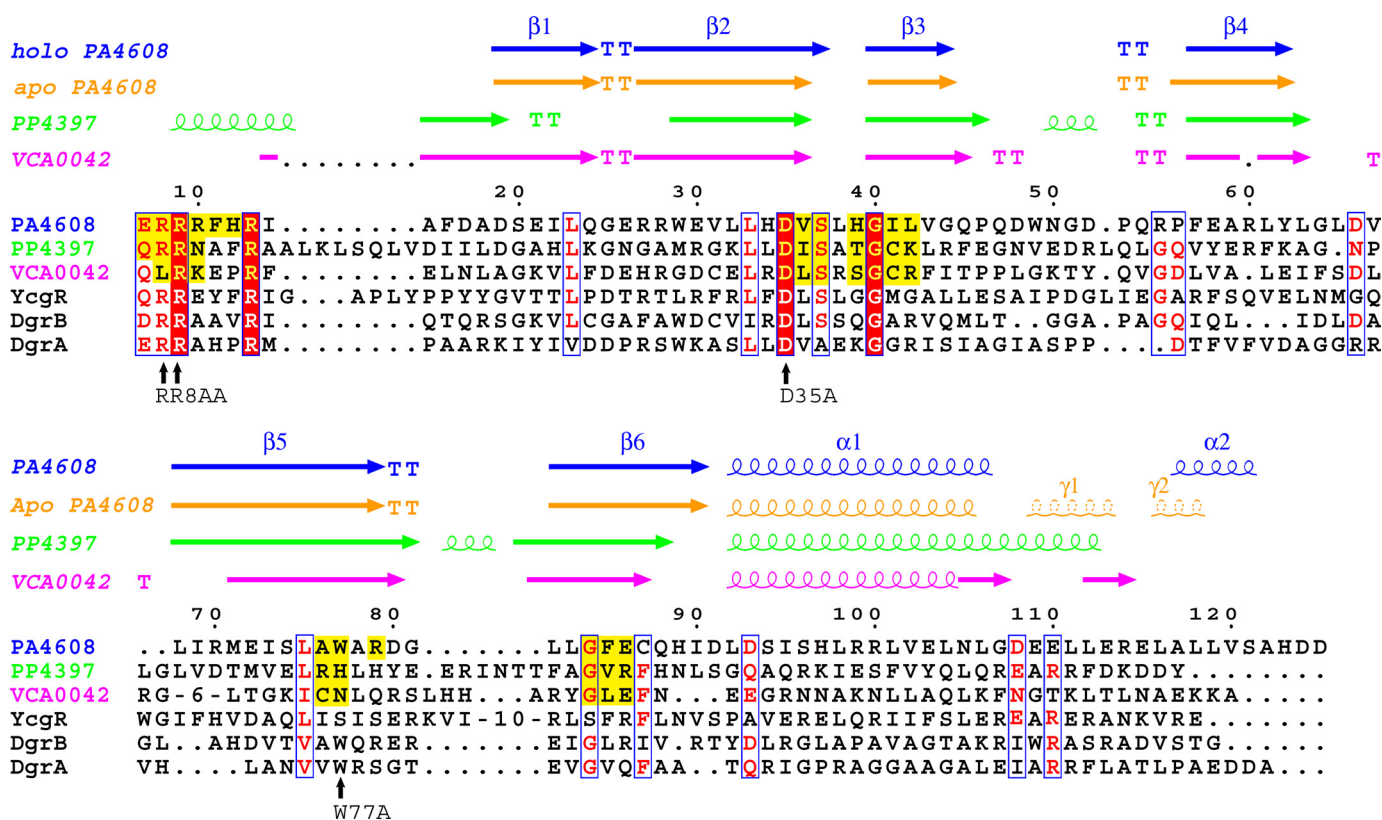
FIGURE 2. **Sequence-structure alignment of various c-di-GMP binding PilZ domains generated with MultAlin (42) and further adjusted manually.** *Arrows* represent β-sheets, and *coils* represent helices of holo PA4608 (this work), apo PA4608 (PDB ID 1YWU (23)), holo PP4397 (PDB ID 3KYF (22)), and holo VCA0042 (PDB ID 2RDE (20)). *Yellow residues* represent van der Waals contacts with c-di-GMP. The structures of the c-di-GMP-binding proteins DgrA and DgrB from *C. crescentus* and YcgR from *S. typhimurium* are unknown. The signature motifs of the PilZ domain (15) R*XXX*R and D*X*S*XX*G are almost fully conserved (*solid red*) in all six PilZ domains. The *black arrows* show DgrA residues that are important for c-di-GMP binding and in *vivo* function as identified by point mutations (16).

ognition of Gua2 phosphate by Asn[124] in PP4397, which are replaced by Asp[35] and Arg[10] in PA4608, respectively.

VCA0042 also consists of a C-terminal PilZ and an N-terminal YcgR domain, dimerizes via the YcgR domain, and binds c-di-GMP in the linker region between the two domains. However, in contrast to PP4397, the bound c-di-GMP is in monomeric form. It was recently proposed that the replacement of the arginine, which precedes the R*XXX*R motif, by a leucine (Fig. 2) is responsible for the recognition of a c-di-GMP monomer instead of a dimer (20, 22). It is interesting to note that the respective residues in PA4608 (Arg[8]) and PP4397 (Arg[122]) are both arginines and recognize the base of Gua4 in the second c-di-GMP monomer by hydrogen bonds, consistent with the binding of dimeric c-di-GMP to both PA4608 and PP4397.

PleD, a multidomain diguanylate cyclase from *C. crescentus*, is a member of an additional class of c-di-GMP-binding proteins (36, 37). In this protein family, the c-di-GMP binding site constitutes an allosteric site responsible for tight feedback inhibition of the enzymatic activity (24). Similar to PA4608 and PP4397, PleD binds a dimer of c-di-GMP. Fig. 5*C* shows the dimer coordination in the same orientation as for PA4608 in Fig. 5*A*. Again, the specific recognition is mainly achieved by arginines. Thus, Arg[390], Arg[359], and Arg[178] assume the role of the base-recognizing Arg[8], Arg[9], and Arg[13] in PA4608. However, these side chains originate obviously from very different regions of the primary sequence and even from different domains of the protein. In addition, no equivalent of the recognition of Gua1 by the aromatic residue Trp[77] (His[201]) in PA4608 (PP4397) is present in the PleD structure. Thus, although the principle of base recognition by arginine fingers is similar, the overall architecture of the c-di-GMP binding sites is distinct for the two protein families.

Comparison of PA4608 Apo and Holo Forms—Pronounced changes in the chemical shifts between the apo and holo form of PA4608 indicate significant structural rearrangements in the N- and C-terminal regions upon ligand binding. Fig. 6 shows a comparison between the two states. In the apo form, the c-di-GMP binding side of the β-barrel comprising strands β2, β3, β5, and β6 is partially covered by the C-terminal 3₁₀ helices γ₁ and γ₂. Binding of c-di-GMP to this site pushes the C-terminal region away from the β-barrel and causes the unfolding of helix γ₁. In turn, the N-terminal region, which is flexible in the apo state, buries the outer side of c-di-GMP with the residues RRRFHR[13] of the recognition motif. The end of the C-terminal region comprising the α-helix α2 covers this part of the N terminus by hydrophobic contacts in a lid-like fashion.

During this process, the β-barrel remains basically unchanged. This is reflected by an r.m.s.d. of 0.9 Å between the heavy backbone atom coordinates of the barrel in the apo and the holo structures. However, a certain rearrangement of helix α1 connecting the barrel to the C-terminal part is evident in Fig. 6. It corresponds to a rotation of 11° of the helix axis

FIGURE 4. **Structural details of the N- and C-terminal coordination of PA4608.** Parts of the N-terminal region (*blue*) comprising residues Arg[8], Arg[9], Arg[10], Phe[11], His[12], and Arg[13] and parts of the C-terminal region comprising helix α2 (*red*) are shown. The *green lines* represent the 15 NOE contacts observed between the side chains of Phe[11] and His[12] and the side chains of Leu[118] and Leu[119].

FIGURE 3. **Structure of PA4608 with bound c-di-GMP.** *A*, ribbon representation of the lowest energy structure of holo PA4608 with c-di-GMP in *stick representation*. *B*, the same structure rotated by 90° around the *y* axis. *C*, overlay of the 20 lowest energy structures in the same orientation and color coding as in *B*. The first 4 disordered residues are omitted. Structures are superimposed as a best fit of β-barrel residues 19–24, 27–37, 40–44, 57–62, 68–79, and 82–90. *D*, same as *C* except that structures are superimposed as a best fit of the c-di-GMP dimer.
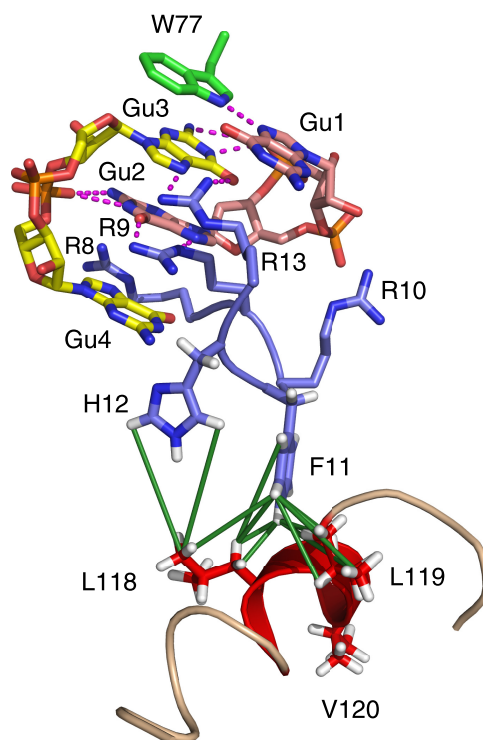
away from the barrel. The helix is structurally well defined with all its residues recognized as α-helical by PROCHECK-NMR (38) for all 20 lowest energy structures. Its change in orientation was confirmed experimentally from RDCs of the complex and the apo state, which were additionally acquired (supplemental Fig. S1 and Table S2). In the region of helix α1, the RDCs of the apo and holo form agree well with the apo and holo structures, respectively, but deviate for an interchange of structures. The difference in helix orientation is apparently caused by the hydrophobic interactions between helix residues Leu[98], Leu[101], and Val[102] and the N-terminal residues Ile[14] and Phe[16] in the apo state and rearrangement of the N- and C-terminal regions, which pull the helix away from the barrel in the complexed state.

The changes in the protein backbone dynamics between the apo and holo form were also characterized by ^{15}N relaxation experiments carried out at 293 K (Fig. 7). For most residues, the [15]N T_1 and T_2 relaxation times as well as the heteronuclear {[1]H}-^{15}N NOEs are very uniform and correspond to a well folded structure. Moreover, they are also almost identical between the apo and holo forms. An analysis by the program TENSOR (16, 39) yields isotropic rotational correlation times (τ_c) of 11.3 ns (12.3 ns) for the apo (holo) form, which are in reasonable agreement with values expected for monomeric PA4608 in its apo form (16.7 kDa) and bound to a dimer of c-di-GMP (18.1 kDa). Also, in agreement with the monomeric state of the complex, no NOEs were detected that would indicate intermolecular dimer
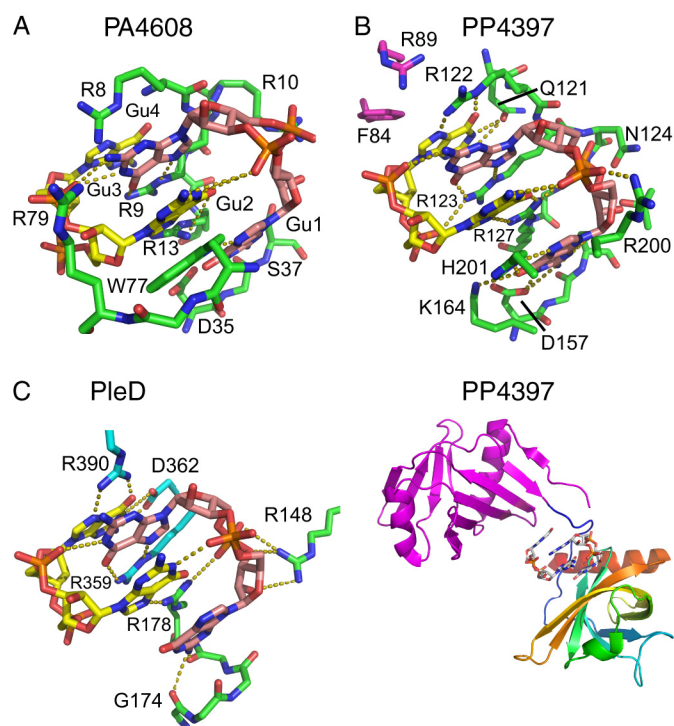
FIGURE 5. **Comparison of c-di-GMP dimer recognition in the stand-alone PilZ protein PA4608, the YcgR-PilZ protein PP4397, and the diguanylate cyclase PleD.** *A*, lowest energy NMR structure of PA4608 (*green*) coordinating the intercalated two c-di-GMP molecules (*pink* and *yellow*). H-bonds are shown as *gold dashed lines*. *B*, PilZ (*green*) and YcgR-N (*magenta*) domains of PP4397 from *P. putida* (PDB ID 3KYF (22)) with bound c-di-GMP dimer. The bottom part gives an overview how the c-di-GMP dimer fits into the junction between the YcgR-N (*magenta*) and PilZ (*green*, *gold*) domains of PP4397. *C*, binding of c-di-GMP dimer to the allosteric inhibition site of the diguanylate cyclase PleD (PDB ID 1W25 (36)). Residues Arg[390], Asp[362], and Arg[359] are from the GGDEF domain (*blue*), and Arg[148], Arg[178], and Gly[174] are from the D2 adaptor domain (*green*) of PleD.

contacts. Thus, PA4608 is monomeric in the apo and holo form.

The relaxation data also give evidence for changes in dynamics upon complex formation for certain parts of the sequence. Thus, the N-terminal residues around Arg[8] have low {[1]H}-^{15}N NOE and large T_2 values corresponding to large amplitude motions on the nanosecond time scale for the apo state, whereas these values become comparable with folded residues in the rest of the sequence in the holo state. This is consistent with folding upon ligand binding of the N terminus. In contrast, the C-terminal end around residue Ala[122] has the opposite behavior, indicating that the very C-terminal end becomes flexible upon ligand binding. Nevertheless, the adjacent helix α2 of the complexed state appears relatively stable because residues Ala[117] and Leu[118] of this helix have ^{15}N relaxation parameters that correspond to a folded structure. Consistent with this stability of helix α2, all its 5 residues (Ala[117] to Ser[121]) are found in α-helical conformations in 90% of the lowest energy structures. In the apo state, the region of residues Asp[108] to Leu[116], which contains the 3_{10} helix γ1, has relaxation parameters of a well folded protein. In contrast, resonances from these residues are strongly broadened or not detectable at all in the holo state, which must be caused by micro- to millisecond dynamics of the protein backbone. Thus, this region undergoes an unfolding transition upon ligand binding.

Clustering of PA4608 Surface Charges upon c-di-GMP Binding—The conformational changes in PA4608 induced by c-di-GMP binding lead to a severe rearrangement of surface charges (Fig. 8). Besides the four positive arginines (Arg[8], Arg[9], Arg[10], Arg[13]), the N-terminal region contains 3 negative residues (Asp[3], Asp[6], Glu[7]). The coordination of the ligand by the four arginines causes their side chains to point inward. The concomitant transition to a folded structure of this region reorients the side chains of Asp[3], Asp[6], and Glu[7] toward the
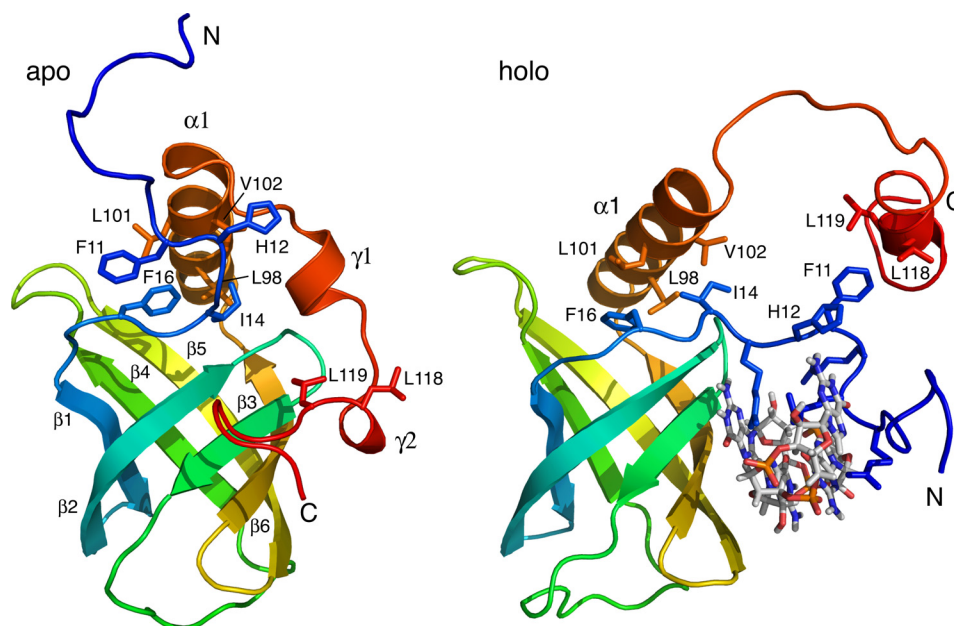

FIGURE 6. **Comparison of overall structure of apo and holo PA4608.** A ribbon representation highlighting the ligand-induced reorientation of the N- (*blue*) and C-terminal (*red*) parts of the protein is shown. The orientation of the barrel is the same for both structures. Residues involved in contacts between the N terminus and the C terminus and in anchoring helix α1 are shown in *stick representation*.
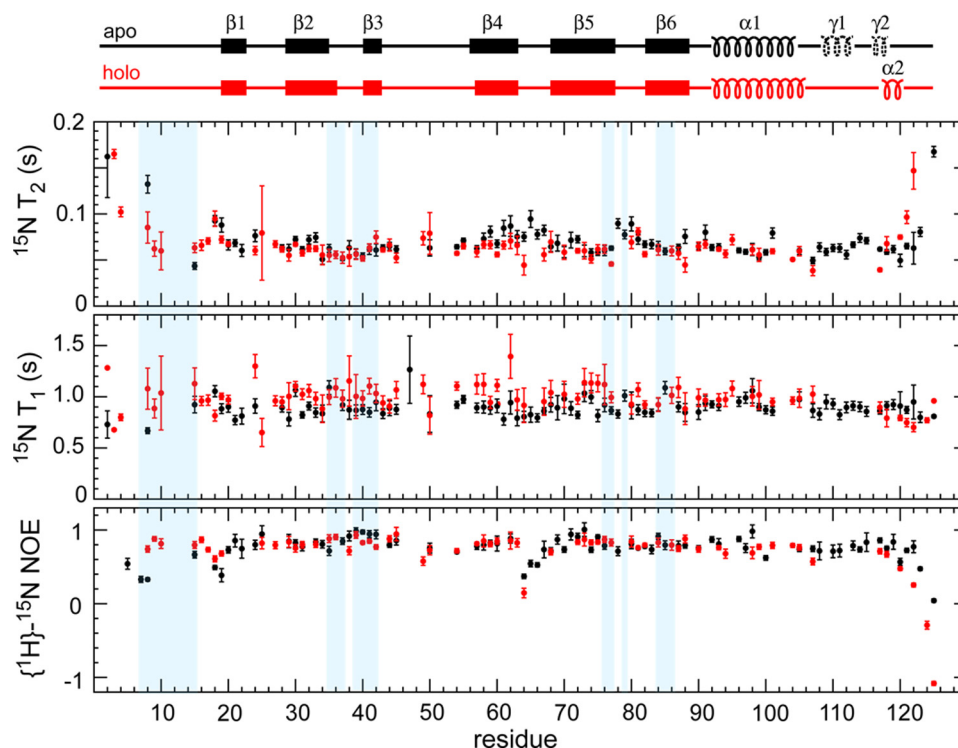
FIGURE 7. **^{15}N T_2, T_1, and heteronuclear {^1H}-^{15}N NOE relaxation data for apo (*black*) and holo (*red*) PA4608 as a function of residue number.** The secondary structures of apo and holo PA4608 are shown on *top*. Residues highlighted in *pale blue* are in van der Waals contact with the ligand. *Solid coils* represent α-helices, and *broken coils* represent 3_{10} helices γ_1 and γ_2.
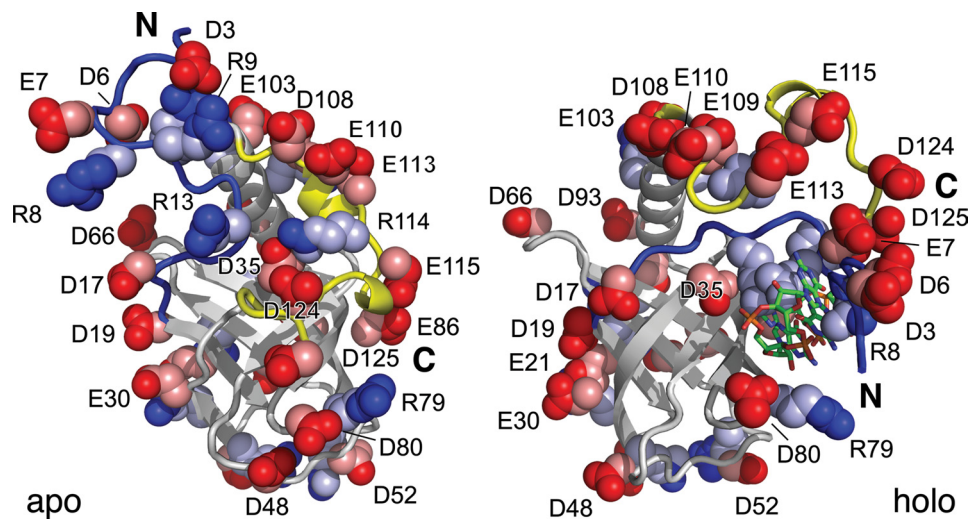

FIGURE 8. **Changes in surface charge distribution induced by c-di-GMP binding to PA4608.** Apo (*left*) and holo (*right*) structures are shown in the same orientation as in Fig. 6. Charged residues (Arg, Lys, Asp, and Glu) are indicated by *blue* and *red space-filling models* of the side chains with interior atoms painted in *fainter colors*. The protein backbone is indicated in *ribbon representation* with N-terminal and C-terminal regions in *blue* and *yellow*, respectively.

outside and creates a negative, surface-exposed cluster (Fig. 8). Likewise, the C terminus contains 7 negative residues centered on the 3_{10} helix $\gamma1$ (Asp[108], Glu[109], Glu[110], Glu[113], Glu[115]) and at the very C terminus (Asp[124], Asp[125]). The ligand-induced rearrangement of this region positions these negative charges on top of the negative cluster around residue Glu[7]. Together with further negative residues of the β-barrel (Asp[17], Asp[19], Glu[21], Glu[30], Asp[35], Asp[48], Asp[80]) and helix $\alpha1$ (Glu[103]), which do not move upon ligand binding, the combined structural changes create one face of the protein, which is strongly negatively charged and devoid of any positive charges. It is attractive

to speculate that this newly generated molecular surface constitutes the readout of this small signaling protein by providing a highly charged interaction surface for high affinity regulatory interactions with downstream target proteins.

Although the PilZ domain associates with a plethora of known signaling, catalytic, or transport domains (40), standalone or single domain PilZ proteins represent the largest subclass of this protein family (see NCBI/CDART (Conserved Domain Architecture Retrieval Tool)). It is interesting to note that some of the surface-exposed charged side chains that were found to cluster on one side of the c-di-GMP-bound form of

PA4608 show conservation within the subclass of stand-alone PilZ domain proteins. In particular, positions corresponding to Glu[7] and to a lesser extent Glu[115] have conserved negative charges within this family but not within the YcgRN-PilZ family represented by the VCA0042 and PP4397 proteins (supplemental Fig. S2). Consistent with this view, the C terminus of PilZ domains varies substantially between these two PilZ protein families (supplemental Fig. S2). Although the PilZ domains of YcgR homologs have rather conserved sequences of defined length, the C termini of single domain PilZ proteins share little sequence homology, vary substantially in length, and are usually strongly charged. This indicates a high degree of modularity within the C termini of single domain PilZ proteins and suggests that the clustering of key C-terminal residues engaged in potential electrostatic interactions might be a conserved principle for many of these c-di-GMP-binding proteins.

The motional freedom of the C-terminal region of the complexed form of PA4608 together with the abundant negative surface charges may indicate that more than one downstream effector protein with a positively charged surface site exists. Patterns of spatially clustered surface charges that have co-evolved with their respective interaction partners may provide binding specificity for a large variety of protein-protein interactions. This emphasizes the potential of this large family of proteins for functional diversification as signal transducer. Given the diverse cellular functions controlled by c-di-GMP in bacteria, rapid adaptation of this sensing module to different signaling pathways could explain the wide distribution and evolutionary success of stand-alone PilZ domain proteins.

Despite the similar binding of c-di-GMP by the PilZ domains in PA4608, VCA0042, and PP4397 (Fig. 5), the three proteins show fundamental differences in how the ligand binding translates into a molecular readout. Although PA4608 is a monomer in its apo- and ligand-bound form, both VCA0042 and PP4397 form dimers. Ligand binding to YcgR homologs induces large conformational changes within the dimer (VCA0042 (20)) or causes a dimer-to-monomer transition (PP4397 (22)). In both structures, the R*XXX*R binding motif, also called the c-di-GMP switch (20), forms a hinge-like connector between the N-terminal YcgR and the C-terminal PilZ domain. In VCA0042, the switch is flexible and disordered in the apo structure but undergoes conformational changes upon ligand binding that cause a rotation of the entire PilZ domain and make it switch back toward the N-terminal YcgR domain (20). In contrast, in apo PP4397, the switch forms a short helix with residues within and next to the switch participating in dimerization. Upon c-di-GMP binding, this helix unwinds and wraps around the ligand, thereby contributing to the dimer-to-monomer transition. Thus, in all three proteins, the c-di-GMP switch region is critically involved in signal transduction by making extensive contacts both to c-di-GMP and to the flanking domains (VCA0042 and PP4397) or the C terminus of the same PilZ protomer (PA4608). Despite the different structural context, the central role of the c-di-GMP switch in all three PilZ proteins seems to be to "decode" ligand binding and to transfer this information to flanking regions within the same domain or interacting domains. Although this translates into repositioning of the two

domains in the YcgR homologs, it amounts to a striking clustering of charged residues on one side of the PilZ domain protein PA4608. This stresses the versatility and evolutionary flexibility of the PilZ c-di-GMP binding module, which allows for rapid functional expansion in cellular signaling pathways.

During the preparation of this manuscript, Shin *et al.* (41) published a structural model of PA4608 bound to c-di-GMP based on solution NMR data. The overall structural arrangement is similar. However, in contrast to our work, the position and coordination of the ligand are not well defined because the ligand and the intermolecular interactions were not assigned. Furthermore, NOEs indicating the coordination of the C terminus to the N terminus were not assigned or detected.

REFERENCES

1. Jenal, U., and Malone, J. (2006) *Annu. Rev. Genet.* **40,** 385–407
2. Hengge, R. (2009) *Nat. Rev. Microbiol.* **7,** 263–273
3. Yan, H., and Chen, W. (2010) *Chem. Soc. Rev.* **39,** 2914–2924
4. Cotter, P. A., and Stibitz, S. (2007) *Curr. Opin. Microbiol.* **10,** 17–23
5. Dow, J. M., Fouhy, Y., Lucey, J. F., and Ryan, R. P. (2006) *Mol. Plant Microbe Interact.* **19,** 1378–1384
6. Ryan, R. P., Fouhy, Y., Lucey, J. F., Jiang, B. L., He, Y. Q., Feng, J. X., Tang, J. L., and Dow, J. M. (2007) *Mol. Microbiol.* **63,** 429–442
7. Tamayo, R., Pratt, J. T., and Camilli, A. (2007) *Annu. Rev. Microbiol.* **61,** 131–148
8. Malone, J. G., Jaeger, T., Spangler, C., Ritz, D., Spang, A., Arrieumerlou, C., Kaever, V., Landmann, R., and Jenal, U. (2010) *PLoS Pathog* **6,** e1000804
9. Duerig, A., Abel, S., Folcher, M., Nicollier, M., Schwede, T., Amiot, N., Giese, B., and Jenal, U. (2009) *Genes Dev.* **23,** 93–104
10. Fineran, P. C., Williamson, N. R., Lilley, K. S., and Salmond, G. P. C. (2007) *J. Bacteriol.* **189,** 7653–7662
11. Merighi, M., Lee, V. T., Hyodo, M., Hayakawa, Y., and Lory, S. (2007) *Mol. Microbiol.* **65,** 876–895
12. Drenkard, E., and Ausubel, F. M. (2002) *Nature* **416,** 740–743
13. Meissner, A., Wild, V., Simm, R., Rohde, M., Erck, C., Bredenbruch, F., Morr, M., Römling, U., and Häussler, S. (2007) *Environ. Microbiol.* **9,** 2475–2485
14. Hickman, J. W., Tifrea, D. F., and Harwood, C. S. (2005) *Proc. Natl. Acad. Sci. U.S.A.* **102,** 14422–14427
15. Schirmer, T., and Jenal, U. (2009) *Nat. Rev. Microbiol.* **7,** 724–735
16. Christen, M., Christen, B., Allan, M. G., Folcher, M., Jenö, P., Grzesiek, S., and Jenal, U. (2007) *Proc. Natl. Acad. Sci. U.S.A.* **104,** 4112–4117
17. Lee, V. T., Matewish, J. M., Kessler, J. L., Hyodo, M., Hayakawa, Y., and Lory, S. (2007) *Mol. Microbiol.* **65,** 1474–1484
18. Hickman, J. W., and Harwood, C. S. (2008) *Mol. Microbiol.* **69,** 376–389
19. Krasteva, P. V., Fong, J. C., Shikuma, N. J., Beyhan, S., Navarro, M. V., Yildiz, F. H., and Sondermann, H. (2010) *Science* **327,** 866–868
20. Benach, J., Swaminathan, S. S., Tamayo, R., Handelman, S. K., Folta-Stogniew, E., Ramos, J. E., Forouhar, F., Neely, H., Seetharaman, J., Camilli, A., and Hunt, J. F. (2007) *EMBO J.* **26,** 5153–5166
21. Boehm, A., Kaiser, M., Li, H., Spangler, C., Kasper, C. A., Ackermann, M., Kaever, V., Sourjik, V., Roth, V., and Jenal, U. (2010) *Cell* **141,** 107–116
22. Ko, J., Ryu, K. S., Kim, H., Shin, J. S., Lee, J. O., Cheong, C., and Choi, B. S. (2010) *J. Mol. Biol.* **398,** 97–110
23. Ramelot, T. A., Yee, A., Cort, J. R., Semesi, A., Arrowsmith, C. H., and Kennedy, M. A. (2007) *Proteins* **66,** 266–271
24. Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P.,

Meuwly, M., and Jenal, U. (2006) *J. Biol. Chem.* **281,** 32015–32024

25. Christen, M., Christen, B., Folcher, M., Schauerte, A., and Jenal, U. (2005) *J. Biol. Chem.* **280,** 30829–30837
26. Grzesiek, S., Bax, A., Hu, J. S., Kaufman, J., Palmer, I., Stahl, S. J., Tjandra, N., and Wingfield, P. T. (1997) *Protein Sci.* **6,** 1248–1263
27. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6,** 277–293
28. Goddard, T., and Kneller, D. (2007) *SPARKY 3.* 114, University of California, San Francisco
29. Garrett, D., Powers, R., Gronenborn, A., and Clore, G. (1991) *J. Magn. Reson.* **95,** 214–220
30. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) *J. Magn. Reson.* **160,** 65–73
31. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) *FEBS Lett.* **229,** 317–324
32. Dingley, A. J., Nisius, L., Cordier, F., and Grzesiek, S. (2008) *Nat. Protoc.* **3,** 242–248
33. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) *Nucleic Acids Res.* **29,** 2860–2874
34. Dougherty, D. A. (1996) *Science* **271,** 163–168
35. Plevin, M. J., Bryce, D. L., and Boisbouvier, J. (2010) *Nat. Chem.* **2,** 466–471
36. Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U., and Schirmer, T. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101,** 17084–17089
37. Wassmann, P., Chan, C., Paul, R., Beck, A., Heerklotz, H., Jenal, U., and Schirmer, T. (2007) *Structure* **15,** 915–927
38. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J. Biomol. NMR* **8,** 477–486
39. Dosset, P., Hus, J. C., Blackledge, M., and Marion, D. (2000) *J. Biomol. NMR* **16,** 23–28
40. Amikam, D., and Galperin, M. Y. (2006) *Bioinformatics* **22,** 3–6
41. Shin, J. S., Ryu, K. S., Ko, J., Lee, A., and Choi, B. S. (2011) *Protein Sci.* **20,** 270–277
42. Corpet, F. (1988) *Nucleic Acids Res.* **16,** 10881–10890
43. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J. Biomol. NMR* **13,** 289–302