

The Structure of the Periplasmic Ligand-binding Domain of the Sensor Kinase CitA Reveals the First Extracellular PAS Domain*

Received for publication, June 4, 2003, and in revised form, July 10, 2003
Published, JBC Papers in Press, July 16, 2003, DOI 10.1074/jbc.M305864200

Stefan Reinelt^{‡§}, Eckhard Hofmann^{‡§¶}, Tanja Gerharz^{||}, Michael Bott^{||}, and Dean R. Madden^{‡**‡‡}

From the [‡]Ion Channel Structure Group, Max Planck Institute for Medical Research, D-69120 Heidelberg, Germany, [¶]Ruhr-Universität Bochum, Proteinkristallographie, Lehrstuhl für Biophysik D-44780 Bochum, Germany, the ^{||}Institut für Biotechnologie 1, Forschungszentrum Jülich, D-52425 Jülich, Germany, and the ^{**}Department of Biochemistry, Dartmouth Medical School, Hanover, New Hampshire 03755-3844

The integral membrane sensor kinase CitA of *Klebsiella pneumoniae* is part of a two-component signal transduction system that regulates the transport and metabolism of citrate in response to its environmental concentration. Two-component systems are widely used by bacteria for such adaptive processes, but the stereochemistry of periplasmic ligand binding and the mechanism of signal transduction across the membrane remain poorly understood. The crystal structure of the CitAP periplasmic sensor domain in complex with citrate reveals a PAS fold, a versatile ligand-binding structural motif that has not previously been observed outside the cytoplasm or implicated in the transduction of conformational signals across the membrane. Citrate is bound in a pocket that is shared among many PAS domains but that shows structural variation according to the nature of the bound ligand. In CitAP, some of the citrate contact residues are located in the final strand of the central β -sheet, which is connected to the C-terminal transmembrane helix. These secondary structure elements thus provide a potential conformational link between the periplasmic ligand binding site and the cytoplasmic signaling domains of the receptor.

Specific recognition of environmental conditions is an important prerequisite for the adaptive regulation of gene expression in response to external stimuli. In bacteria, this response is often mediated by two-component regulatory systems consisting of a sensor kinase and a response regulator. Transmembrane sensor kinases are used for the recognition of external stimuli and typically contain an extracytoplasmic sensor domain, flanked by two transmembrane helices, and a cytoplasmic histidine kinase domain. In many cases, they also contain additional cytoplasmic modulatory and one or more signal transfer domains (1, 2). Coupling between the sensor and kinase domains enables these proteins to transduce a signal

across the membrane to the cytoplasm, initiating a downstream cascade that involves the autophosphorylation of a conserved histidine residue within the kinase domain and the subsequent transfer of the phosphoryl group to an aspartate residue in the corresponding response regulator (1).

In the past two decades, hundreds of two-component systems have been identified. Based on knock-out experiments, few two-component systems appear to be essential. However, some have been implicated in the control of bacterial virulence and drug resistance. Furthermore, their adaptive importance and complete absence in mammals could provide an attractive set of antimicrobial drug targets (2). Nevertheless, many aspects of their molecular function remain obscure. In particular, for the ligand-binding sensor kinases little is known about such fundamental processes as the stereochemistry of ligand-binding, the nature of the associated conformational changes, and the mechanism of signal transduction to the cytoplasmic kinase domain.

To obtain a more detailed understanding of these processes, we have analyzed the crystal structure of the ligand-binding domain of the sensor kinase CitA from *Klebsiella pneumoniae*. Together with the response regulator CitB, CitA is required for the induction of citrate fermentation genes by the presence of environmental citrate under anoxic conditions. These genes are organized in a cluster that also includes, for example, subunits of citrate lyase, the citrate carrier CitS and oxaloacetate decarboxylase (3, 4). Synthesis of these proteins requires careful regulation, because inappropriate expression would interfere with central metabolic pathways. Thus, even though CitA appears to be nonessential (1, 2), targeted dysregulation of its pathway could provide therapeutic approaches.

CitA exhibits a domain organization similar to that described above, *i.e.* a periplasmic citrate-binding domain flanked by two transmembrane helices and followed by linker and kinase domains located in the cytosol (3). The periplasmic domain of CitA can be expressed as a soluble polyhistidine fusion protein (CitAP)¹ that exhibits highly specific citrate binding properties (5).

Biochemical investigations of the interactions between citrate and CitAP wild-type and mutant proteins have been reported previously (5, 6). In this study we present the structure of CitAP in complex with citrate and molybdate, revealing the ligand-binding stereochemistry of an integral membrane sensor kinase and providing initial insights into possible mechanisms of signal transduction.

* This work was supported by the Max Planck Society. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[¶] The atomic coordinates and structure factors (code 1P0Z) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

[§] Both authors contributed equally to this work.

^{‡‡} Supported by Howard Hughes Medical Institute Award 76200-560801 to Dartmouth Medical School under the Biomedical Research Support Program for Medical Schools. To whom correspondence should be addressed: Dept. of Biochemistry, Dartmouth Medical School, 7200 Vail Bldg., Hanover, NH 03755-3844. Tel.: 603-650-1164; Fax: 603-650-1128; E-mail: dean.madden@dartmouth.edu.

¹ The abbreviations used are: CitAP, CitA periplasmic domain; MES, 2-[N-morpholino]ethanesulfonic acid; NCS, non-crystallographic symmetry; SeMet, selenomethionine; PYP, photoactive yellow protein.

EXPERIMENTAL PROCEDURES

Crystallization—CitAP was expressed in *Escherichia coli* strain BL21 (DE3) as a 15.3-kDa polyhistidine fusion protein and purified as previously described (5). Selenomethionine (SeMet)-labeled CitAP was produced in BL21 (DE3) cells grown in M9 medium (0.04% glucose, w/v) as described (7). Purified CitAP was dialyzed extensively against 10 mM NaCl, 10 mM Tris, pH adjusted to 8.0 with HCl at 20 °C. 1 mM Na-citrate was added to protein (16 mg/ml) prior to crystallization. Initial conditions for crystal growth were identified by microbatch trials performed at the High Throughput Crystallization Laboratory at the Hauptmann-Woodward Institute (Buffalo, NY) (8). Following optimization of conditions, crystals were grown in hanging drops by vapor diffusion against buffer containing 100 mM Na₂MoO₄, 100 mM MES, pH 6.0, and either 8% (w/v) polyethylene glycol (PEG) 4000 and 2.5% (v/v) glycerol or 14% PEG-4000 and 15% glycerol for the native and SeMet-labeled protein, respectively. Protein and reservoir buffer were mixed 1:1. The best crystals were obtained at 16 and 27 °C for the native and labeled proteins, respectively. Prior to freezing in liquid N₂ the crystals were soaked briefly in a solution of 0.5 mM Na-citrate, 25% (w/v) PEG-4000, and 20% (v/v) glycerol.

Structure Determination—Oscillation data were collected at ID 14–4 of the European Synchrotron Radiation Facility (Grenoble, France) at 100 K using an ADSC Q4R CCD detector. For phasing, data were collected from crystals of SeMet-labeled protein at the Se K-edge ($\lambda = 0.9793$ Å, 1° oscillations, 0.5 s, 180-mm plate distance). Diffraction data for crystals of native protein were collected at $\lambda = 0.9393$ Å (1° oscillations, 1 s, 140-mm plate distance). Data were processed in space group P1 using XDS (9). A test set was established containing 3% of the reflections chosen in thin resolution shells. The Matthews volume was estimated at 2.8 Å³/Da, assuming the presence of 10 molecules per asymmetric unit. Self-rotation functions also indicated the presence of up to five independent non-crystallographic 2-fold axes.

Forty ordered Se sites were identified in the asymmetric unit using single wavelength anomalous techniques as implemented in SHELXD (10), corresponding to ten monomers. Initial phases were calculated using SHELXE (10) and improved using RESOLVE (11), which automatically identified and applied 2-fold non-crystallographic symmetry (NCS). Ten clusters of very strong positive peaks in the electron density were used to determine initial NCS operators (these were later identified as Mo₇O₂₄ clusters).

Model Building and Refinement—Using Arp/Warp 6.0 (12) and the CCP4 program suite (13), 1068 of the 1390 protein residues were placed in a solvent-flattened, 2-fold NCS-averaged initial electron density map. For the most completely traced monomer (132/139) all side chains were autodocked using the program GUISE (12). The other nine copies were generated using the initial NCS operators.

The resulting model was then refined against the native data set. Based on values of mean I/σ_I and $R_{\text{mrgd-F}}$ as a function of resolution (Table II), data to 1.6 Å resolution were used. In the first cycle the individual monomers were treated as rigid bodies. Subsequent cycles of simulated annealing, conjugate gradient minimization, and temperature-factor refinement against maximum-likelihood targets were followed by manual rebuilding, using the programs CNS (14) and O (15), respectively. In the first cycle, experimental phase restraints were used. Tight NCS definitions were applied during refinement and were partially released in the final cycles as necessary to account for alternate conformations and crystal contacts.

The final model contains 10 monomers composed of amino acid residues 5–135, 1 Mo₇O₂₄ cluster, 1 MoO₃-citrate, and 1 Na⁺ each. A total of 1573 water molecules were added. Refinement statistics are presented in Table I.

Size-exclusion Chromatography—Size-exclusion chromatography was performed using a Superose 12 HR 10/30 column (Amersham Biosciences), pre-equilibrated in 100 mM MES, 150 mM NaCl, 0.02% Na-azide, pH 6.0, in the presence or absence of 1 mM Na-citrate, respectively. Experiments were performed at 25 °C, and concentrations were held at 34 mg/ml. 100 μl of sample was loaded onto the column and eluted at 0.6 ml/min. Elution profiles were monitored by UV absorbance at 280 and 218 nm.

RESULTS

Overall Structure—The structure of CitAP in complex with citrate was determined at an initial resolution of 1.9 Å using single wavelength anomalous diffraction data from crystals of SeMet-labeled CitAP (Table I and Fig. 1). The asymmetric unit contains 10 monomers (referred to here as monomers A–J). A

TABLE I
Data collection, phasing, and refinement statistics

Data set	SeMet	Native
Wavelength (Å)	0.9793	0.9393
Resolution (Å)	20-1.9	38.0-1.6
Cell parameters (a,b,c; Å)	61.6, 83.9, 98.0	61.9, 84.6, 98.0
(α, β, γ ; degrees)	112.8, 107.4, 93.7	112.9, 107.5, 93.6
Space group	P1	P1
Completeness	0.893	0.958
Multiplicity	2.0	8.7
Average I/σ_I	8.7	11.9
R_{sym}^a	0.053	0.103
R_{meas}^b	0.075	0.110
$R_{\text{mrgd-F}}^c$	0.110	0.079
Phasing		
Figure of merit ^d (SHELXE)	0.663	
Figure of merit ^d (RESOLVE)	0.730	
Refinement		
Resolution (Å)	38.0-1.6	
R_{cryst}^e	0.168 (0.242; 1.70-1.60 Å)	
R_{free}^f	0.190 (0.228; 1.72-1.71 Å)	
Number of atoms (non-hydrogen)		
Protein	9980	
Alternate conformations	480	
Ligands	480	
Na ⁺ ions	10	
Water	1573	
Number of side chains with alternate conformations	61	
Average B-factor (Å ²)	28.2	
Root mean square deviation from ideality (protein atoms)		
Bonds (Å)	0.014	
Angles (degrees)	1.7	

$$^a R_{\text{sym}} = \sum_{h,j} |\hat{I}_h - I_{hj}| / \sum_{h,j} I_{hj}$$

$$^b R_{\text{meas}} = (\sum_h (n/(n-1))^{0.5} \sum_j |\hat{I}_h - I_{hj}|) / (\sum_{h,j} I_{hj}) \text{ with } \hat{I}_h = (\sum_j I_{hj})/n_j$$

$$^c R_{\text{mrgd-F}} \text{ as defined in Ref. 57.}$$

$$^d \text{Figure of merit} = \langle \int P(\alpha) e^{i\alpha} d\alpha / \int P(\alpha) d\alpha \rangle$$

$$^e R_{\text{cryst}} = \sum_h |F_{h,\text{calc}} - F_{h,\text{obs}}| / \sum_h F_{h,\text{obs}} \text{ for } h \in \{\text{refinement set}\}$$

$$^f R_{\text{free}} = \sum_h |F_{h,\text{calc}} - F_{h,\text{obs}}| / \sum_h F_{h,\text{obs}} \text{ for } h \in \{\text{test set}\}$$

model containing residues 5–135 of each CitAP monomer, ligand moieties, and 1573 water molecules was refined against native diffraction data to 1.6 Å resolution, yielding an R_{free} of 19.0% with excellent geometry (Table I). Crystals formed only in the presence of molybdate.

In the ligand-bound state, CitAP forms a mixed α/β -structure (Fig. 2, A and B) with a central five-stranded β -sheet containing residues 56–61, 66–68, 94–100, 104–110, and 121–128 (residue 2 of CitAP corresponds to residue 45 in the native CitA sequence). This sheet is flanked on one side by the protein N terminus, which forms two long, approximately parallel α -helices (residues 10–24 and 38–51) connected by a third, shorter α -helix nearly perpendicular to the first two (residues 27–34). The other side of the β -sheet is packed against the long peptide that connects strands S2 and S3 of the sheet and that contains a short 3_{10} -helix and a short α -helix. The sheet and the inter-strand peptide together form a deep concave pocket that contains electron density for a 1:1 complex of the citrate ligand with an additional moiety well fit by the structure of MoO₃ (16), consistent with the presence of molybdate in the crystallization buffer (Fig. 1). In this complex, the molybdenum is octahedrally coordinated by the three oxo groups and a tridentate interaction with citrate (Figs. 1 and 2B).

The discovery of a MoO₃-citrate complex in the binding site was unexpected, because CitAP binds citrate *in vitro* in the absence of metal ions other than sodium (present in the buffer) and because Mg²⁺ appears to inhibit the interaction (5). How-

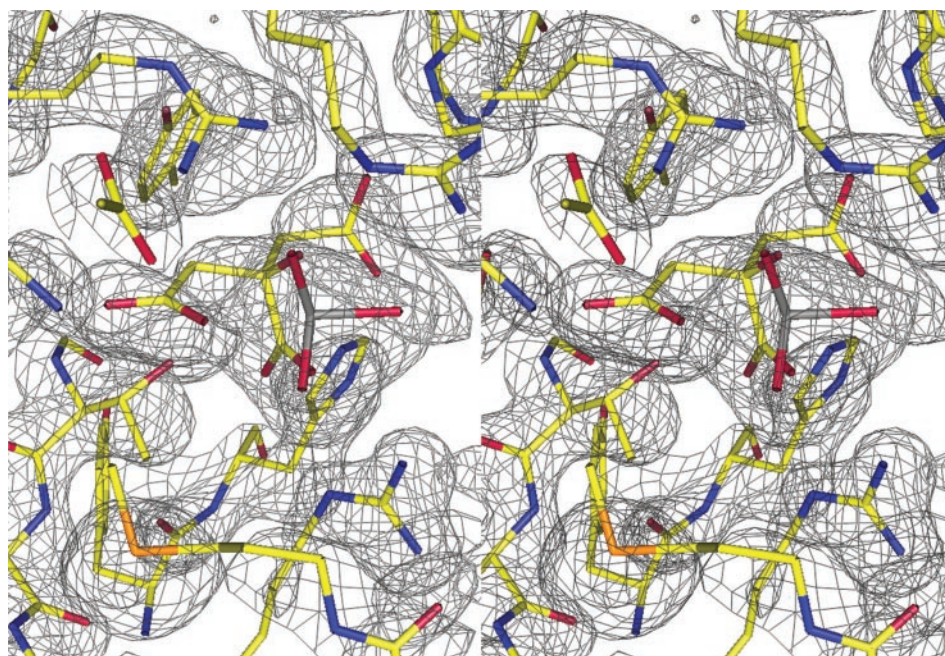


FIG. 1. **Stereo model of the citrate-binding site with experimental data.** A stick figure of the refined structure is shown, with citrate in the center and MoO_3 (gray and red tripod) in the foreground. Bonds are colored by atom type: carbon, yellow; oxygen, red; nitrogen, blue; molybdenum, gray; selenium, orange. A SeMet residue is seen at the bottom (with orange selenium atom). Three water molecules have been omitted for clarity. The solvent-flattened, 2-fold NCS-averaged single wavelength anomalous diffraction experimental electron density is also shown contoured at 1.1σ , revealing clear density for all features of the binding site. No cover radius was applied. All structural figures were generated using PYMOL (55).

ever, the role of metal ions in regulating the expression of the citrate fermentation genes has not been systematically investigated, except in the case of sodium, which is required (3). Thus, the observation that CitAP can accommodate MoO_3 -citrate suggests that other metal-citrate complexes might also be recognized by CitA. Furthermore, in several systems, metal-citrate complexes have been implicated in regulating the transport of metals across the cell membrane. For example, in *Bacillus subtilis*, the homologous CitS/CitT two-component system regulates the expression of a Mg-citrate transport protein (17). In *E. coli*, the outer membrane receptor/transporter FecA up-regulates its own transcription in response to ferric citrate binding (18). By analogy, the ability of *K. pneumoniae* CitA to bind MoO_3 -citrate raises the possibility that the associated citrate carrier CitS (3, 19) may be able to transport not only free citrate but also metal-citrate complexes as well.

In addition to the citrate-bound MoO_3 group, two other non-protein moieties were identified (Fig. 2A). One forms an octahedral complex involving three water molecules, the main-chain carbonyl group of Pro-111, and the hydroxyl moieties of Ser-110 and Ser-24, and thus links the central β -sheet with the C-terminal end of the H1 α -helix. Its position is marked with a green sphere in Fig. 2A. Based on its coordination geometry (20) and the composition of the mother liquor, this moiety has been modeled as a sodium ion. A second site with very strong electron density was found near the N and C termini and could be well fit by the structure of an isopolymolybdate cluster (Mo_7O_{24} ; Fig. 2A, bottom) (21). To our knowledge, it represents the first example of a Mo_7O_{24} cluster bound to a protein structure. Although unlikely to be physiologically relevant, the cluster mediates crystal packing interactions, consistent with the essential role of molybdenum for crystal formation.

CitAP Has a PAS Fold—A structural data base search using DALI (22) revealed significant similarity between the CitA periplasmic ligand-binding domain described here and several members of the PAS superfamily. Homologues ($Z \geq 4.0$) include the C-terminal domain of the IclR-type transcriptional regula-

TABLE II
Quality of native data set

Resolution	Completeness	$\langle I/\sigma \rangle$	R_{sym}^a	R_{meas}^b	$R_{\text{mrgd-F}}^c$	No. reflections ^d
20.00 Å	0.591	18.61	0.091	0.104	0.073	301
10.00	0.970	26.61	0.066	0.073	0.035	5320
6.00	0.981	24.56	0.068	0.073	0.032	26,912
3.00	0.978	22.57	0.079	0.084	0.036	251,420
2.30	0.971	16.76	0.103	0.110	0.051	352,879
2.00	0.964	12.27	0.156	0.165	0.080	331,839
1.90	0.959	9.36	0.231	0.245	0.122	160,885
1.80	0.956	7.46	0.318	0.338	0.166	197,680
1.70	0.953	5.95	0.438	0.465	0.224	246,507
1.60	0.927	4.24	0.646	0.687	0.338	296,880
total	0.958	11.88	0.103	0.110	0.079	1,870,623

$$^a R_{\text{sym}} = \sum_{h_j} |\hat{I}_h - I_{h_j}| / \sum_{h_j} I_{h_j}$$

$$^b R_{\text{meas}} = (\sum_h (n/(n-1))^{0.5} \sum_j |\hat{I}_h - I_{h_j}|) / (\sum_{h_j} I_{h_j}) \text{ with } \hat{I}_h = (\sum_j I_{h_j}) / n_j$$

^c $R_{\text{mrgd-F}}$ as defined in Ref. 57.

^d Number of reflections included in R factor calculations.

tor TM0065 (23), the photoactive yellow protein (PYP) (24), the GAF domain of yeast open reading frame YKG9 (25), the human erg potassium channel fragment (HERG) (26), a 30 S ribosomal protein subunit (27), the oxygen sensor of FixL (28), and profilin (29). Based on sequence homology, a PAS domain has been proposed in the cytoplasmic portion of CitA between the second transmembrane and the kinase domains (30), but not for the periplasmic CitAP domain. Indeed, neither sequence-based fold prediction by BLAST (31), 3D-PSSM (32), or LOOPP (33) nor a manual search for S1/S2 boxes (34) indicated the PAS fold of CitAP.

Based on the structural similarity of CitAP and PYP (Fig. 3A), we aligned their sequences and that of FixL, because PYP and FixL represent the first structures determined for a PAS domain and for a ligand-binding PAS domain, respectively (24, 28) (Fig. 3B). Also included in the alignment are the sequences of the ligand-binding domains of the sensor kinases DcuS and PhoQ, both of which are homologous to CitA and are subjects of parallel structural studies (35, 36).

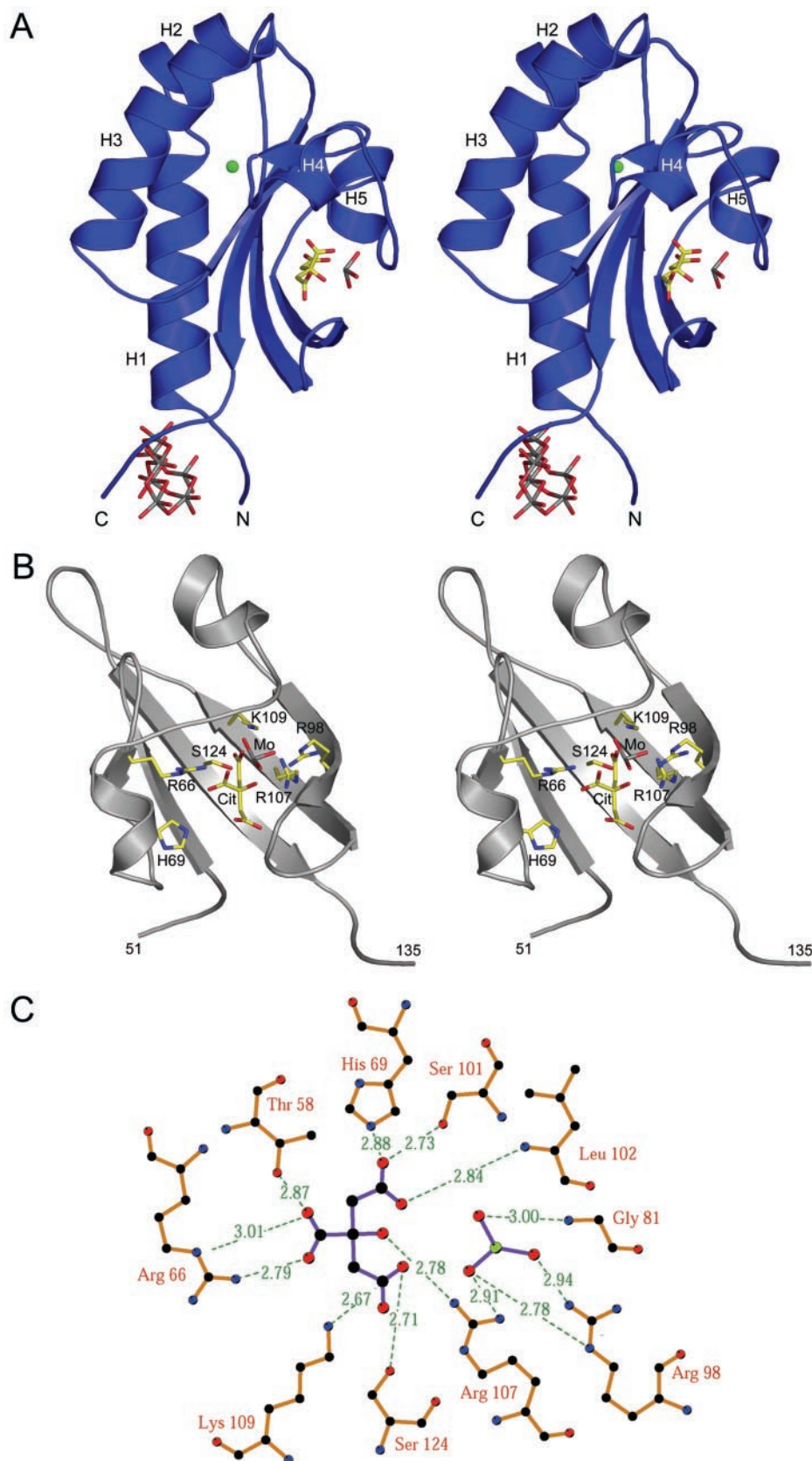


FIG. 2. Structure of the CitAP-citrate complex. *A*, the overall structure of CitAP is shown in stereo as a *ribbon diagram* viewed parallel to the plane of the central β -sheet. The citrate-binding pocket is on the *right*. Citrato-molybdate (*right*) and isopolymolybdate (*bottom*) groups are shown as *stick figures*. A putative sodium ion is shown as a *green sphere*. Helices H1–H5 are labeled according to Fig. 3*B*. *B*, the CitAP structural elements forming the citrate-binding pocket (residues 51–134) are shown in stereo as a *ribbon diagram*, together with *stick figures* representing the bound citrate (*Cit*) and molybdate (*Mo*) groups and the protein side chains that contact them. The N-terminal helices are omitted for clarity. Atom coloring as in Fig. 1. *C*, the hydrogen-bonding network in the citrate-binding pocket is represented schematically using the program LIGPLOT (56). Ligand-ligand interactions and hydrophobic contacts are not shown. Atom colors as in panels *A* and *B*, except for carbon (*black*) and molybdenum (*green*). The MoO_3 and citrate moieties are shown with *purple bonds*, protein side chains with *yellow bonds*.

The most strongly conserved feature of the PAS domain proteins is the presence of a central β -sheet (30). Superposition of the CitAP β -scaffold composed of the strands S1–S5, with the corresponding strands $A\beta$, $B\beta$, $G\beta$, $H\beta$, and $I\beta$ of PYP reveals a root mean square deviation of 1.2 Å for 39 $C\alpha$ atoms (nomen-

clature for PYP PAS elements according to Ref. 28). In contrast, the structural elements on either or both sides of the β -scaffold show a much greater variability (Fig. 3*A*). The two major helices H1 and H3 of CitAP are more elongated than the $A\alpha$ - and $B\alpha$ -helices in PYP and are connected by a short helical linker

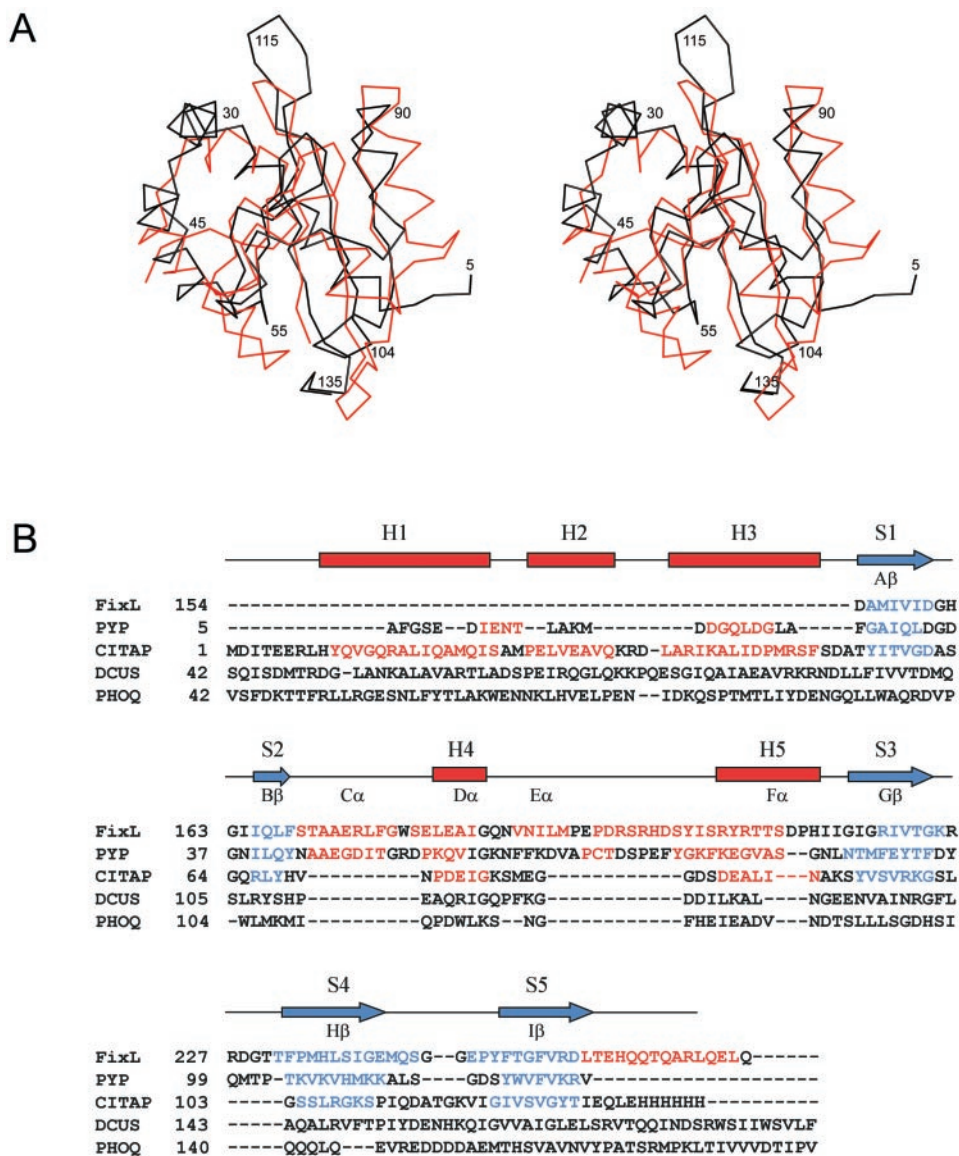


FIG. 3. CitAP has a PAS fold. *A*, a stereo $C\alpha$ trace of CitAP (black) was superimposed on that of PYP (Ref. 24; red) by the program DALI (22). Numbers refer to CitAP sequence. *B*, the sequence of CitAP was aligned with those of other PAS domains using structural (*Ectothiorhodospira halophila* PYP, residues 5–125, and *Bradyrhizobium japonicum* FixL, residues 154–270), or sequence homology (*E. coli* PhoQ, residues 42–183 and *E. coli* DcuS, residues 42–189). For known structures, residues in α -helices and β -sheets are marked in red and blue, respectively. Schematics above the sequence represent secondary structure elements of CitAP (α -helices, bars; β -sheets, arrows) with corresponding labels above the schematics. Labels below the schematics indicate the published nomenclature for secondary structural elements of FixL (28).

instead of a loop. N-terminal α -helices corresponding to those of CitAP are found in the YGK9 GAF domain, the 30 S ribosomal subunit, and (cyclically permuted) in TM0065 and profilin. They are completely absent in the FixL and HERG PAS domains and are not, in fact, required for PYP function (37).

Ligand-binding Site—Additional differences between CitAP and PYP are observed on the opposite side of the β -scaffold (Fig. 3A), where the ligand-binding site is located (Fig. 2). In this region, CitAP lacks the $C\alpha$ -helix and the short $B\beta$ -strand, and the helical connector H5 is significantly shorter than the corresponding $F\alpha$ -helix of PYP or FixL (Fig. 3B). Because of the smaller size of the interstrand loops in this part of CitAP compared with PYP, citrate is bound closer to the β -sheet and its pocket is more closed than is the case for the PYP 4-hydroxycinnamoyl moiety. Overall, the PAS binding pockets characterized to date exhibit considerable variability, consistent with the need to accommodate their respective structurally diverse cognate ligands (23, 24, 28, 38–41). Nevertheless, the binding pockets are typically located in the same general area

between the β -sheet and the interstrand loops. Putative effector binding pockets similar in size and location to that of CitAP have also been found in the TM0065 signal-binding domain, the YGK9 GAF domain, and the HERG PAS domain, although the cognate ligands have not yet been identified for these domains (23, 25, 26).

In detail, the CitAP ligand-binding groove is formed by the β -sheet together with helix H4 (residues 72–76) and two flanking random coils (residues 69–71 and 77–84) (Fig. 2). The relevant contacts between the protein and the MoO_3 -citrate can be divided into several groups. First, the binding of citrate is supported by hydrophobic contacts between citrate and the residues Tyr-56 and Met-79. Main-chain hydrogen bond interactions to citrate are mediated by Thr-58, Ser-101, and Leu-102. However, the majority of citrate-protein interactions are formed by hydrogens bonds originating from the side-chain atoms of Arg-66, His-69, Arg-107, Lys-109, and Ser-124. Except for Lys-109 and Ser-124, these residues are highly conserved among the CitA subfamily of histidine kinases (5, 6). Ser-124 is

partially conserved in the CitA subfamily and is located in the final β -strand S5. Hydrogen bonds to the oxygens of MoO_3 are formed by Gly-81, Arg-107, and especially by Arg-98, which interacts with two MoO_3 oxygens. Among these residues only Gly-81 is partly conserved.

Dimerization—In solution, CitAP exhibits a monomer-dimer equilibrium at millimolar protein concentrations in both the presence and absence of citrate (Fig. 4). Analysis of the lattice packing interactions in the crystal revealed two distinct candidate dimer interfaces (Fig. 5). In the first type (e.g. between monomers “E” and “G”), the interface is mainly built by the

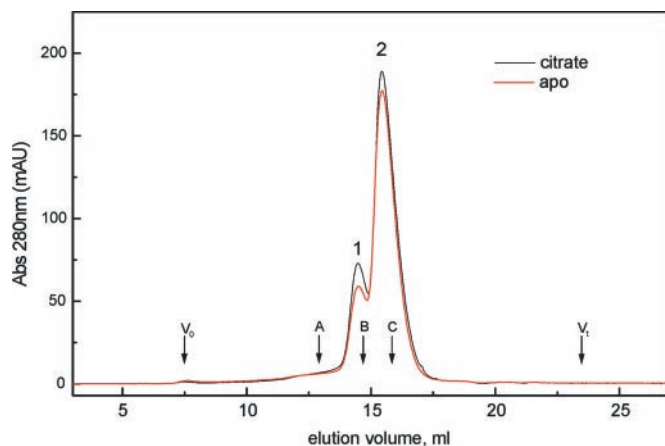


FIG. 4. CitAP exhibits a monomer:dimer equilibrium in solution. The 280-nm absorbance trace during size exclusion chromatography analysis shows the presence of CitAP dimers (*peak 1*) and monomers (*peak 2*) in both the presence (*black*) and absence (*red*) of 1 mM citrate. Void (V_0) and total (V_t) volumes are shown for the column, together with the elution volumes of molecular weight standards (A, bovine serum albumin; B, carbonic anhydrase; C, cytochrome *c*). The estimated molecular weight values of the CitAP monomer and dimer peaks are 17,600 and 35,600, respectively. The calculated monomer molecular weight is 15,300.

strands S3 and S4, the helical connector H5, and additional interactions in the N-terminal region (Fig. 5, A and B). This dimer interface buries a total of 1798 \AA^2 of accessible surface. The dominant feature of the second dimer interface (e.g. between monomers “G” and “J”) is the parallel association of the N-terminal helices H1 and H3 involving residues 10–24 and 38–51 (Fig. 5, C and D). This dimer formation decreases the total accessible surface by 1415 \AA^2 .

There are two major differences in the molecular architecture of the resulting dimers. In the EG dimer (Fig. 5, A and B), the interface is relatively planar and the two N-terminal helices are oriented roughly perpendicular to the molecular 2-fold, i.e. parallel to the presumed membrane surface (horizontal in Fig. 5, A and C; in the plane of page in Fig. 5, B and D). In the GJ dimer (Fig. 5, C and D), the interface is characterized by a deep pocket in each surface that binds the Phe-51 side chain of the other molecule (Fig. 5D) and the N-terminal helices are arranged roughly perpendicular to the putative membrane. Although it is not possible to determine which of the two dimer interactions is physiological on the basis of these data alone, the GJ assembly is more suggestive of a mechanism for signal transduction, because it clusters the attachment points for the transmembrane helices and permits them to form as extensions of domain secondary structure elements. Furthermore, a GJ-like dimer has also been reported for a preliminary structure of the PhoQ histidine kinase sensor domain in the apo form (36), which, like CitAP, exhibits only weak dimerization in solution (42).

DISCUSSION

The ability of organisms to adapt to changing environmental conditions is essential for their long-term survival and requires a set of receptors that can recognize external stimuli and activate the appropriate responses. In bacteria, two-component regulatory systems play an important role in these processes. In particular, membrane-bound sensor kinases allow the detection of ligands outside of the cytoplasmic membrane, so that

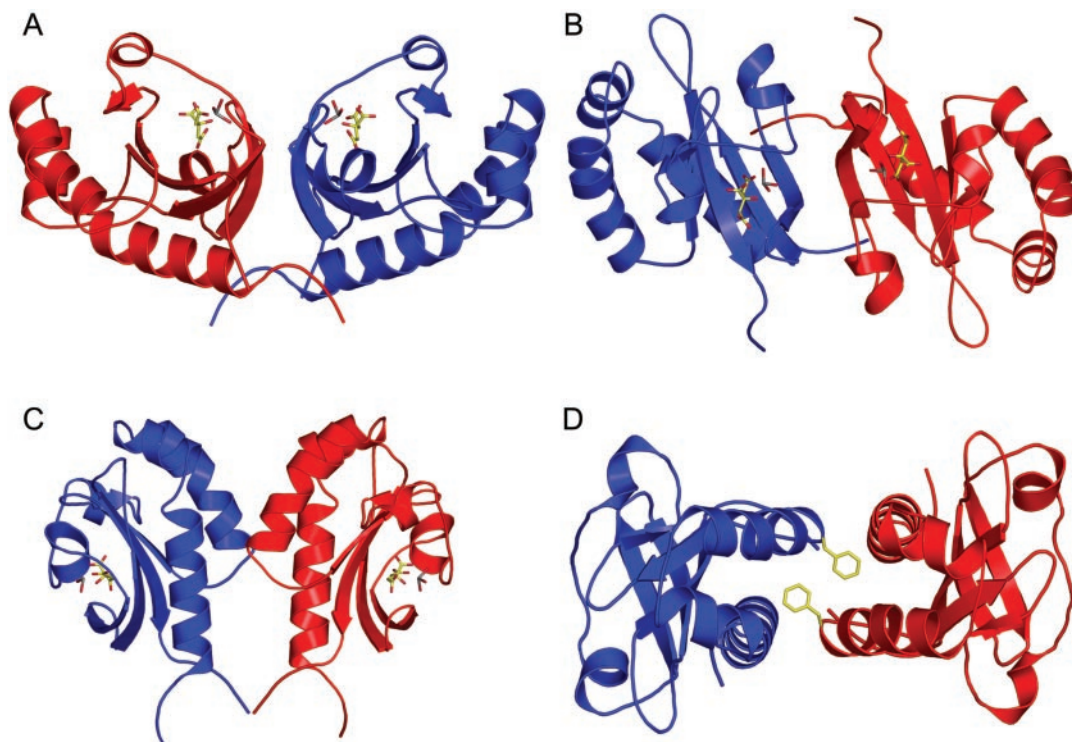


FIG. 5. Ribbon diagrams of two dimeric forms of CitAP observed in the crystal lattice. A, side view of the EG-type dimer. B, top view of the EG-type dimer. C, side view of the GJ-type dimer. D, top view of the GJ-type dimer, showing side chains Phe-51 at the dimer interface. Citrato-molybdate moieties are shown as *stick figures*. Within each dimer, individual monomers are *red or blue*.

the cellular response does not require transport or diffusion of these ligands into the cell. However, despite their key role in bacterial adaptation, many aspects of the molecular mechanisms of these proteins are unknown. Here, we present the first detailed crystallographic analysis of a sensor kinase periplasmic signal-binding domain and of the interaction of this domain with its cognate ligand.

The structure reveals that CitAP adopts a PAS fold, unexpected on the basis of sequence comparisons (Figs. 2 and 3). The PAS fold has now been identified in domains from a functionally heterogeneous group of prokaryotic and eukaryotic proteins, often exhibiting negligible sequence identity. Named for the founding members Per, ARNT, and Sim, the family includes signal-recognition modules that regulate cellular responses to light, oxygen, voltage (LOV) and other environmental stimuli (28, 30, 38, 41, 43, 44). In addition, PAS domains have been found in several other proteins, including a eukaryotic ion channel (26), a nucleotide-binding protein (25), and transcriptional regulators (23, 39, 40).

Based on sequence motifs, PAS domains have also been identified in a number of sensor histidine kinases, including CitA. However, these domains, like all PAS family members identified to date, are proposed to have an intracellular location. To our knowledge, the CitAP domain is the first representative of the PAS family located outside the cytoplasm. Because a large number of sensor kinase signal-binding domains exhibit sequence homology to CitAP (5, 6), we can anticipate that they will also adopt a PAS fold. Given its demonstrated flexibility in accommodating a wide variety of ligands in an evolutionarily protean binding pocket, the PAS fold appears well suited to the functional requirement that the sensor kinases be able to recognize chemically distinct environmental signals.

Citrate is generally bound to proteins via clusters of positively charged side chains that neutralize its multiple negatively charged carboxylate moieties (e.g. aconitase (45), yeast isocitrate dehydrogenase (46), citrate synthase (47), FecA (48)), although this is not strictly required (e.g. human aldose reductase (49), MoFe nitrogenase (50)). In the case of CitAP, two arginines, a lysine, and a histidine are found to interact with citrate in the pocket (Fig. 2), all of which had been previously identified by calorimetric binding studies of site-directed mutant CitAP domains (6). A third arginine that was found to influence citrate binding interacts indirectly with citrate via the MoO₃ group. Conservation of many of these countercharges suggests that the mode of binding is also likely to be found among the CitA subfamily of histidine kinases (5, 6). Because the mutational study was performed in the absence of molybdenum, it also suggests that the binding stereochemistry of the citrate-molybdate complex observed in the crystal structure is similar to that of citrate alone. The formation of this MoO₃-citrate complex is consistent with the presence of molybdenum in the crystallization buffer and represents, to our knowledge, the first structure of the MoO₃ species in complex with protein.

In the absence of a citrate-free structure, the molecular rearrangements responsible for kinase activation cannot be directly visualized. Effector-induced conformational changes have been characterized for the oxygen sensor FixL and the light-sensing LOV1 and LOV2 domains. In these cases, only small, local changes are observed (28, 38, 41, 44), and it appears that these changes may represent conserved signaling mechanisms (37, 51). Because these sensing domains interact directly with their respective histidine kinase domains, relatively small changes are sufficient for downstream signaling. In contrast, the CitA ligand-binding and histidine kinase domains are located on opposite sides of the cytoplasmic mem-

brane from each other, requiring distinct, long-range conformational changes for signal transduction.

Because CitAP is observed to dimerize in solution (Fig. 4) and because sensor kinase autophosphorylation generally occurs in *trans* within a homodimer (2), signal transduction is likely to occur within the context of a CitA dimer. However, the extent of CitAP dimerization does not appear to be strongly dependent on ligand binding (Fig. 4), so it is unlikely that signal transduction is mediated by a shift in the CitA monomer: dimer equilibrium, such as that seen in the human growth hormone receptor system (52).

Instead, CitA activation is more likely to be associated with rearrangements among the transmembrane helices that can, in turn, be transmitted to the cytoplasmic kinase domains. The citrate binding site includes the final strand (S5) of the β -sheet, which is connected to the transmembrane helix at the C terminus of the domain. As a result, ligand binding could directly affect the packing of the helices. Furthermore, alanine mutations of residues Arg-7, His-9, and Arg-15 have been shown to affect citrate-binding thermodynamics (6). These residues are located outside the binding pocket but near the N-terminal transmembrane helix. Finally, evidence suggests that sodium is required in addition to citrate for signal transduction to occur (3), and the putative sodium ion identified in our structure of CitAP is located between the N-terminal helix H1 and the β -sheet (Fig. 2A). All of these observations are consistent with the proposal that conformational rearrangements among the transmembrane helices and the central β -sheet are important for sensor kinase activation. Helical shifts have also been proposed as the mechanism for the bacterial chemoreceptors (53), although the nature of the conformational couplings is likely to differ because of differences in the structure and location of the ligand-binding site (54). As a result, a detailed understanding of PAS domain-mediated transmembrane signaling will ultimately require comparison of the ligand-bound structure presented here with a ligand-free state structure that remains to be determined.

Acknowledgments—We gratefully acknowledge the assistance of N. Fehrmann (HTS Lab, Hauptmann-Woodward Institute, Buffalo, NY) in performing high-throughput crystallization screens. We acknowledge the Department of Biophysics (MPIMF) and the European Synchrotron Radiation Facility (Grenoble, France) for provision of X-ray diffraction facilities. We thank Dr. J. McCarthy for assistance in the use of ESRF beamline ID14-4 and Drs. J. Fethiere (EMBL, Heidelberg) and W. Kabsch (MPIMF) for assistance with data collection and analysis.

Note Added in Proof—While this manuscript was in press, Pappalardo and colleagues reported the NMR structure of the apo form of the periplasmic binding domain of the sensor kinase DcuS (Pappalardo, L., Janausch, I. G., Vijayan, V., Zientz, E., Junker, J., Peti, W., Zweckstetter, M., Unden, G., and Griesinger, C. (2003) *J. Biol. Chem.* **278**, 39185–39188).

REFERENCES

- Robinson, V. L., Buckler, D. R., and Stock, A. M. (2000) *Nat. Struct. Biol.* **7**, 626–633
- Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000) *Annu. Rev. Biochem.* **69**, 183–215
- Bott, M., Meyer, M., and Dimroth, P. (1995) *Mol. Microbiol.* **18**, 533–546
- Bott, M. (1997) *Arch. Microbiol.* **167**, 78–88
- Kaspar, S., Perozzo, R., Reinelt, S., Meyer, M., Pfister, K., Scapozza, L., and Bott, M. (1999) *Mol. Microbiol.* **33**, 858–872
- Gerharz, T., Reinelt, S., Kaspar, S., Scapozza, L., and Bott, M. (2003) *Biochemistry* **42**, 5917–5924
- Van Duyn, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J. Mol. Biol.* **229**, 105–124
- Luft, J. R., Wolfley, J., Jurisica, I., Glasgow, J., Fortier, S., and DeTitta, G. T. (2001) *J. Cryst. Growth* **232**, 591–595
- Kabsch, W. (1993) *J. Appl. Crystallogr.* **26**, 795–800
- Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 1772–1779
- Terwilliger, T. C. (2000) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **56**, 965–972
- Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6**, 458–463
- Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. Sect.*

- D Biol. Crystallogr.* **50**, 760–763
14. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 905–921
 15. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47**, 110–119
 16. Zhou, Z. H., Wan, H. L., and Tsai, K. R. (2000) *Inorg. Chem.* **39**, 59–64
 17. Yamamoto, H., Murata, M., and Sekiguchi, J. (2000) *Mol. Microbiol.* **37**, 898–912
 18. Harle, C., Kim, I., Angerer, A., and Braun, V. (1995) *EMBO J.* **14**, 1430–1438
 19. Van der Rest, M. E., Siewe, R. M., Abee, T., Schwarz, E., Oesterhelt, D., and Konings, W. N. (1992) *J. Biol. Chem.* **267**, 8971–8976
 20. Brese, N. E., and Okeeffe, M. (1991) *Acta Crystallogr. Sect. B Struct. Sci.* **47**, 192–197
 21. Gili, P., Lorenzo-Luis, P. A., Mederos, A., Arrieta, J. M., Germain, G., Castineiras, A., and Carballo, R. (1999) *Inorg. Chim. Acta* **295**, 106–114
 22. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233**, 123–138
 23. Zhang, R. G., Kim, Y., Skarina, T., Beasley, S., Laskowski, R., Arrowsmith, C., Edwards, A., Joachimiak, A., and Savchenko, A. (2002) *J. Biol. Chem.* **277**, 19183–19190
 24. Genick, U. K., Borgstahl, G. E. O., Ng, K., Ren, Z., Pradervand, C., Burke, P. M., Srajer, V., Teng, T. Y., Schildkamp, W., McRee, D. E., Moffat, K., and Getzoff, E. D. (1997) *Science* **275**, 1471–1475
 25. Ho, Y. S. J., Burden, L. M., and Hurley, J. H. (2000) *EMBO J.* **19**, 5288–5299
 26. Cabral, J. H. M., Lee, A., Cohen, S. L., Chait, B. T., Li, M., and Mackinnon, R. (1998) *Cell* **95**, 649–655
 27. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonnrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) *Nature* **407**, 327–339
 28. Gong, W. M., Hao, B., Mansy, S. S., Gonzalez, G., Gilles-Gonzalez, M. A., and Chan, M. K. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 15177–15182
 29. Cedergrenzeppenzauer, E. S., Goonesekere, N. C. W., Rozycki, M. D., Myslik, J. C., Dauter, Z., Lindberg, U., and Schutt, C. E. (1994) *J. Mol. Biol.* **240**, 459–475
 30. Taylor, B. L., and Zhulin, I. B. (1999) *Microbiol. Mol. Biol. Rev.* **63**, 479–506
 31. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
 32. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. E. (2000) *J. Mol. Biol.* **299**, 499–520
 33. Meller, J., and Elber, R. (2001) *Proteins* **45**, 241–261
 34. Zhulin, I. B., and Taylor, B. L. (1997) *Trends Biochem. Sci.* **22**, 331–333
 35. Parac, T. N., Coligaev, B., Zientz, E., Unden, G., Peti, W., and Griesinger, C. (2001) *J. Biomol. NMR* **19**, 91–92
 36. Regelmann, A. G., Lesley, J. A., Mott, C., Stokes, L., and Waldburger, C. D. (2002) *J. Bacteriol.* **184**, 5468–5478
 37. Vreede, J., Van der Horst, M. A., Hellingwerf, K. J., Crielaard, W., and van Aalten, D. M. F. (2003) *J. Biol. Chem.* **278**, 18434–18439
 38. Crosson, S., and Moffat, K. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2995–3000
 39. Amezcua, C. A., Harper, S. M., Rutter, J., and Gardner, K. H. (2002) *Structure* **10**, 1349–1361
 40. Vannini, A., Volpari, C., Gargioli, C., Muraglia, E., Cortese, R., De Francesco, R., Neddermann, P., and Di Marco, S. (2002) *EMBO J.* **21**, 4393–4401
 41. Fedorov, R., Schlichting, I., Hartmann, E., Domratcheva, T., Fuhrmann, M., and Hegemann, P. (2003) *Biophys. J.* **84**, 2474–2482
 42. Waldburger, C. D., and Sauer, R. T. (1996) *J. Biol. Chem.* **271**, 26630–26636
 43. Gu, Y. Z., Hogenesch, J. B., and Bradfield, C. A. (2000) *Annu. Rev. Pharmacol.* **40**, 519–561
 44. Crosson, S., and Moffat, K. (2002) *Plant Cell* **14**, 1067–1075
 45. Lloyd, S. J., Lauble, H., Prasad, G. S., and Stout, C. D. (1999) *Protein Sci.* **8**, 2655–2662
 46. Lin, A. P., McCammon, M. T., and McAlister-Henn, L. (2001) *Biochemistry* **40**, 14291–14301
 47. Remington, S., Wiegand, G., and Huber, R. (1982) *J. Mol. Biol.* **158**, 111–152
 48. Ferguson, A. D., Chakraborty, R., Smith, B. S., Esser, L., van der Helm, D., and Deisenhofer, J. (2002) *Science* **295**, 1715–1719
 49. Harrison, D. H. T., Bohren, K. M., Petsko, G. A., Ringe, D., and Gabbay, K. H. (1997) *Biochemistry* **36**, 16134–16140
 50. Mayer, S. M., Gormal, C. A., Smith, B. E., and Lawson, D. M. (2002) *J. Biol. Chem.* **277**, 35263–35266
 51. Crosson, S., Rajagopal, S., and Moffat, K. (2003) *Biochemistry* **42**, 2–10
 52. Cunningham, B. C., Ultsch, M., Devos, A. M., Mulkerrin, M. G., Clauser, K. R., and Wells, J. A. (1991) *Science* **254**, 821–825
 53. Falke, J. J. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6530–6532
 54. Milburn, M. V., Prive, G. G., Milligan, D. L., Scott, W. G., Yeh, J., Jancarik, J., Koshland, D. E., and Kim, S. H. (1991) *Science* **254**, 1342–1347
 55. DeLano, W. L. (2002) *The PyMOL Molecular Graphics Systems*, DeLano Scientific, San Carlos, CA
 56. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng.* **8**, 127–134
 57. Diederichs, K., and Karplus, P. A. (1997) *Nat. Struct. Biol.* **4**, 269–275