Crystal Structure of the CheA Histidine Phosphotransfer Domain that Mediates Response Regulator Phosphorylation in Bacterial Chemotaxis

Lionel Mourey, Sandra da Re, Jean-Denis Pedelacq, Tatiana Tolstykh, Cécile Faurie, Valérie Guillet, Jeffry B Stock, Jean-Pierre Samama

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 276, No. 33, Issue of August 17, pp. 31074–31082, 2001

Printed in U.S.A.

Crystal Structure of the CheA Histidine Phosphotransfer Domain that Mediates Response Regulator Phosphorylation in Bacterial Chemotaxis*

Received for publication, March 5, 2001, and in revised form, May 23, 2001
Published, JBC Papers in Press, May 32, 2001, DOI 10.1074/jbc.M101943200

Lionel Mourey‡, Sandra Da Re§, Jean-Denis Pédelacq‡, Tatiana Tolstykh§, Cécile Faurie‡, Valérie Guillet‡, Jeffry B. Stock§¶, and Jean-Pierre Samama‡‖

From the ‡Groupe de Cristallographie Biologique, Centre National de la Recherche Scientifique/Institut de Pharmacologie et de Biologie Structurale, 205 route de Narbonne, 31077 Toulouse Cedex, France and the §Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544-1014

The x-ray crystal structure of the P1 or H domain of the *Salmonella* CheA protein has been solved at 2.1-Å resolution. The structure is composed of an up-down up-down four-helix bundle that is typical of histidine phosphotransfer or HPt domains such as *Escherichia coli* $ArcB_C$ and *Saccharomyces cerevisiae* Ypd1. Loop regions and additional structural features distinguish all three proteins. The CheA domain has an additional C-terminal helix that lies over the surface formed by the C and D helices. The phosphoaccepting His-48 is located at a solvent-exposed position in the middle of the B helix where it is surrounded by several residues that are characteristic of other HPt domains. Mutagenesis studies indicate that conserved glutamate and lysine residues that are part of a hydrogen-bond network with His-48 are essential for the ATP-dependent phosphorylation reaction but not for the phosphotransfer reaction with CheY. These results suggest that the CheA-P1 domain may serve as a good model for understanding the general function of HPt domains in complex two-component phosphorelay systems.

The signal transduction system that controls motor behavior in *Salmonella* and *Escherichia coli* provides a paradigm for understanding the biochemistry of intracellular information processing networks (for recent reviews see Refs. 1–3). Sensory inputs are mediated by a family of transmembrane chemoreceptor proteins that interact with the histidine protein kinase CheA. The central region of CheA is a histidine kinase module composed of a four-helix bundle dimerization domain formed from the parallel association of two up-down helices, each linked to an ATP-binding catalytic domain that contains the kinase active site (4). This structural arrangement seems to be a general feature of the over 500 different histidine protein kinases that have been identified in a wide range of different signal transduction systems in microorganisms and plants (5–7). In most members of the histidine kinase superfamily, the histidine that is phosphorylated is located at a solvent-exposed position termed the H box located in the N-terminal helix of the dimerization domain (8). In CheA proteins, however, the conserved H box histidine in the dimerization domain is replaced by glycine, and the site of histidine phosphorylation is displaced to a distinct N-terminal α-helical domain termed the P1 or H domain (9, 10).

Histidine protein kinases generally function together with a second type of signal transduction module, the response regulators (11). The phospho-histidine generated at a given histidine kinase H box is subsequently transferred to a highly conserved aspartate residue at the active site of a cognate response regulator (12). Phosphorylation generally activates the response regulators to produce a response. Thus, the phospho-histidine group in the P1 domain of CheA is readily transferred to the chemotaxis response regulator, CheY, and phospho-CheY binds to proteins at the flagellar motor to control swimming behavior and effect chemotaxis responses.

P1-like domains have been identified in a number of different phosphorelay signal transduction systems besides chemotaxis. They are monomeric domains with a conserved four-helix bundle that contains the phosphoaccepting histidine at a solvent-exposed position in the second helix. High resolution structures of the so-called histidine phosphotransfer (HPt) domains from the *E. coli* histidine kinase ArcB, $ArcB_C$ (13) and from *Saccharomyces cerevisiae* Ypd1 (14, 15), indicated some similarity to the low resolution NMR structure of the P1 domain of CheA (10). Here we report the crystal structure of the CheA-P1 domain at 2.1-Å resolution. As expected from the solution NMR structure, the domain is composed of 5 α-helices. The phosphoaccepting histidine is positioned near the middle of the B helix, with the imidazole side chain exposed to solvent. The topology of the four-helix bundle formed by the A, B, C, and D helices is essentially identical to the corresponding HPt motif of $ArcB_C$ (13) and Ypd1 (14, 15). Unlike in CheA, the dimerization domain of the ArcB histidine kinase has an H box histidine that accepts phosphoryl groups from ATP. Moreover, ArcB also has a response regulator CheY-like domain between its dimerization domain and the C-terminal HPt domain, $ArcB_C$. The phosphoryl group in $ArcB_C$ is subsequently transferred to separate response regulator proteins that function to regulate transcription, ArcA and OmpR (16). Similar domain arrangements have been seen in numerous other regulatory systems, although frequently, as in yeast, the HPt domain is produced as a sep-

* This work was supported by Public Health Service Grant R01 GM57773 (to J. B. S) and the Program de Recherches Fondamentales en Microbiologie and Centre National de la Recherche Scientifique (to J. P. S). 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 1i5n) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ To whom correspondence may be addressed. Tel.: 609-258-6111; Fax: 609-258-6175; E-mail: jstock@princeton.edu.

‖ To whom correspondence may be addressed. Tel.: 33-5-61-17-54-44; Fax: 33-5-61-17-54-48; E-mail: samama@ipbs.fr.

TABLE I
Data collection and structure determination statistics

Protein datasets	CheA-P1	SeMet CheA-P1		
	Native	Peak λ_1	Remote λ_2	Edge λ_3
Data collection				
Wavelength (Å)	0.946	0.9802	0.9300	0.9805
Resolution limit (Å)	36.5–2.17	35.3–2.14	35.3–2.14	35.3–2.14
Measured reflections	100362	208310	206627	127566
Unique reflections	26401	31480	31247	31443
Completeness[a] (%)				
Overall/last shell	86.0/88.2	98.6/93.1	98.3/91.3	98.2/90.8
R_{sym}[a] (%)				
Overall/last shell	9.7/19.0	5.5/17.5	4.8/16.7	5.6/23.0
MAD phasing				
f′ (e^-) initial/refined[b]		−7.4/−8.4	−2.2	−9.3/−10.7
f″ (e^-) initial/refined[b]		5.3/5.3	3.4	3.0/3.6
P_p iso acen/cent[c]	2.7/1.9	7.2/4.5		6.2/3.7
P_p ano[c]		4.5	3.8	3.4
R_{Cullis} iso acen/cent[c]	0.91/0.91	0.32/0.35		0.33/0.36
R_{Cullis} ano[c]		0.45	0.54	0.62
FOM acen/cent[c]			0.83/0.72	

[a] The resolution ranges in highest bin are 2.29–2.17 Å and 2.26–2.14 Å for the Met and the SeMet protein, respectively.
[b] Initial values for λ_1 and λ_3 were calculated with the program Chooch (24), and values for λ_2 (the reference dataset) were from the Sasaki tables (63) and not refined.
[c] Phasing statistics provided by the program SHARP (25).

arate protein. The high degree of structural similarity between HPt domains suggests a common mechanism of action.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—A fragment of the *Salmonella cheA* gene that encodes CheA-(1–138) was cloned into a pQE12 vector (Qiagen) that adds two codons at the N terminus and eight codons including six His codons at the C terminus of the gene as described previously (17). Kinetics studies of the phosphotransfer activities of this protein indicate that the his_6 tag has no significant effect on phosphorylation (17). In the x-ray structure determination, there was no electron density corresponding to the his_6 tag at the end of helix E. In the protein structure, this region is very distant from the active site on the opposite face of the bundle. M15/pREP4/pHD cells were grown in LB medium at 37 °C with 100 μg/ml ampicillin and 25 μg/ml kanamycin. The selenium-substituted protein was similarly expressed in methionine auxotroph strain B834(DE3) using minimal medium supplemented with 17 amino acids, the bases for nucleic acids, various salts, sulfate, isopropyl-1-thio-β-D-galactopyranoside, and SeMet[1] (18). Cells from 1.0 liter of culture medium was harvested by centrifugation at 4 °C at 5500 rpm for 1 h, suspended in 60 ml of buffer A (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 3.5 mM EDTA) plus 1.0 mM phenylmethylsulfonyl fluoride, and lysed by sonication. The resulting extract was centrifuged at 100,000 × *g* for 1 h to remove membranes and cell debris, and the supernatant was dialyzed overnight against buffer A and subjected to nickel affinity column chromatography on a 7.5-ml column (Amersham Pharmacia Biotech) equilibrated in buffer A. Protein was eluted with a 300-ml linear gradient of 0–0.50 M imidazole in buffer A. Met CheA-P1 eluted between 100 and 300 mM imidazole. The SeMet protein was purified by the same procedure except all buffers were supplemented with 15 mM β-mercaptoethanol. The purified P1 was dialyzed overnight against buffer B (25 mM Tris-HCl, pH 7.5, 25 mM NaCl, and 3.5 mM EDTA), loaded on a 6.0-ml UnoQ6 column (Bio-Rad) equilibrated in the same buffer, and eluted with an 18-ml linear gradient of 175–275 mM NaCl in buffer B. The purification profiles of both the Met and SeMet P1 proteins indicated a contaminant that migrated during SDS-polyacrylamide gel electrophoresis with an apparent molecular weight of 17,000. Removal of the P1 fractions containing this contaminant was essential for obtaining good quality diffracting crystals. Pure P1 fractions were pooled and dialyzed overnight against buffer B. Electrospray mass spectrometry of the purified SeMet CheA-P1 domain established that all its methionines were replaced by SeMet.

To purify mutant P1 domains for biochemical analyses, fractions from nickel affinity columns were concentrated in an Ultrafree 5K centrifugal device (Millipore) and loaded onto a Sepharose S12 column (Amersham Pharmacia Biotech) equilibrated in 25 mM Tris-HCl, pH 7.5, 200 mM NaCl, and 1.0 mM dithiothreitol. Fractions containing pure P1 domains were pooled, concentrated, dialyzed against 25 mM Tris-HCl, pH 7.5, 25 mM NaCl, and 0.1 mM dithiothreitol and stored at −20 °C.

Crystallization—Met and SeMet proteins were dialyzed against buffer B and concentrated to ~7 mg/ml. In the case of the SeMet protein, 15 mM β-mercaptoethanol were also added. Small platelike crystals were obtained initially for the Met protein using the Crystal Screen sparse matrix (19). The optimized crystallization medium using the hanging drop vapor diffusion method at 12 °C was 20% polyethylene glycol monomethyl ether (w/v), 0.1 M sodium acetate, pH 5.0–5.2, and 0.3–0.4 M ammonium sulfate. Platelike crystals appeared after 3 weeks. The crystals were cryocooled in liquid nitrogen after a rapid transfer in 35% polyethylene glycol monomethyl ether (w/v) or in a drop of reservoir solution (0.5 μl) containing 20% (w/v) glycerol for the Met and the SeMet protein, respectively. The crystals belong to the orthorhombic space group $P2_12_12_1$ with cell parameters of a = 55.3 Å, b = 88.8 Å, c = 116.3 Å (a = 54.9 Å, b = 88.7 Å, and c = 115.8 Å for the SeMet protein), with four molecules per asymmetric unit.

Data Collection and Phasing—A data set for the Met protein was collected at 100 K on the beam line ID14-EH4 at the European Synchrotron Radiation Facility (Grenoble, France). A multiwavelength anomalous dispersion experiment was performed on a single crystal (600 × 200 × 80 μm^3) of the SeMet protein. Diffraction data were collected at 100 K on the BW7A beam line of the European Molecular Biology Laboratory outstation at Deutsches Elektronen Synchrotron (Hamburg, Germany). A fluorescence spectrum recorded from the cryocooled crystal was used to select the wavelengths at the selenium K-absorption edge (λ_3 = 0.9805Å), at the peak (λ_1 = 0.9802 Å), and at a remote wavelength on the high energy side (λ_2 = 0.9300 Å). Complete and highly redundant data were collected to 2.1-Å resolution on a 30-cm MAR Research image plate with a frame size of 1.0° and ~5-min exposure time. Intensities were processed using the program MOSFLM (20). The CCP4 suite of programs (21) was used for merging and scaling (program SCALA) and for data reduction (Table I). Selenium atom positions were identified using Patterson map search procedures (22) as implemented within the SOLVE package (23). Initial values of anomalous scattering factors used in SOLVE were calculated from the fluorescence spectrum with the program Chooch (24). Seventeen selenium sites were found out of the 24 expected in the asymmetric unit. Refinement of the anomalous scatterers and multiwavelength anomalous dispersion phasing were conducted with the program SHARP (25) using all data between 35.3 and 2.1 Å and the remote wavelength (λ_2) data set as a reference (Table I). The native data set was also included in the phasing (26). The experimental phases, with a mean figure of merit of 0.83 (0.72) for acentric (centric) reflections, were improved by subsequent cycles of solvent flattening in SOLOMON (27) assuming a solvent content of 40% in the unit cell. This raised the average figure of merit to 0.93.

[1] The abbreviation used is: SeMet, selenomethionine.

TABLE II
Refinement statistics

Resolution	35.3–2.14 Å
No. of molecules in the A.U.[a]	4
No. of protein atoms	4014
No. of solvent molecules water/sulfate	215/1
No. of reflections work/test[b]	29631/1548
Crystallographic R factor/R_{free}	0.210/0.248
Rmsd bond lengths[c]	0.007 Å
Rmsd bond angles/dihedrals[c]	1.1/17.6 °
Mean temperature factor	23.4 Å^2

[a] A.U., asymmetric unit.
[b] The R_{free} value was calculated from a random set of reflections (5%) that were omitted from structure refinement. The R factor was calculated from the remaining reflections.
[c] Rmsd, root mean square deviation.

Model Building and Crystallographic Refinement—Initial automated tracing, performed with ARP/wARP using the "warpNtrace" procedure (28), led to the identification of 302 residues belonging to 20 different main chain fragments. These fragments clearly accounted for the four molecules in the asymmetric unit. The construction of one CheA-P1 molecule was then pursued manually, leading to a molecular model containing 126 of 142 residues and 114 side chains. The three other molecules were then positioned in the asymmetric unit by molecular replacement using AMoRE (29) yielding an R factor and correlation coefficient of 0.39 and 60.2% (8–3.5 Å), respectively. Structure refinement was performed using CNS (30). Anisotropic B factor and bulk solvent corrections as well as the cross-validation method (31) were applied throughout refinement using reflections of the high energy data set (λ_2). No noncrystallographic symmetry restraints were applied, and the four molecules in the asymmetric were modeled and refined independently. Model building and corrections were carried out in SIGMAA (32) weighted maps using TURBO-FRODO (33). The final model comprises 125–129 of 142 residues in the recombinant protein, excluding the C-terminal His_6 tag, one sulfate ion, and 215 water molecules. The crystallographic R factor and R_{free} values were 0.210 and 0.248, respectively (Table II). All residues are in the allowed regions of a Ramachandran plot, and 96.8% of them have the most favored backbone Φ,Ψ angles, as defined by PROCHECK (34). The average B factors are 23.2 Å^2 for protein atoms (21.5 Å^2 for main chain atoms and 24.9 Å^2 for side chain atoms) and 55.9 Å^2 and 26.6 Å^2 for the sulfate ion and water molecules, respectively.

Oligonucleotide-directed Polymerase Chain Reaction Mutagenesis—Point mutations in *cheA* P1 were generated using the Quick Change™ site-directed mutagenesis kit (Stratagene). Briefly, a supercoiled doubled-stranded DNA vector containing the insert of interest was amplified using two complementary oligonucleotide primers containing the desired mutation. After amplification, the parental DNA encoding a His-tagged CheA-P1 domain (17) was digested with *Dnp*I to select for mutation-containing synthesized DNA. The final clones, pHDK51A and pHDE70A, were confirmed by sequencing. The primer sequences and their complements were CGCACTCCATTGCAGGCGGCGCGGG (K51A) and CCACGCATTTAATGGCAAACCTGCTGGATG (E70A).

ATP-dependent Phosphorylation of CheA-P1—Phosphorylation assays were performed at 24 °C. Reactions contained 100 μM CheA-P1, 10 μM of CheA catalytic C domain (17), 5.0 mM $MgCl_2$, 50 mM KCl, and 50 mM Tris-HCl, pH 7.5. Reactions were initiated by the addition of [γ-^{32}P]ATP to a final concentration of 0.50 mM. At the indicated times, 10-μl aliquots were removed, added to the SDS-polyacrylamide gel electrophoresis sample buffer containing 100 mM EDTA, and subjected to SDS-polyacrylamide gel electrophoresis. Radioactivity co-migrating with P1 was measured using a PhosphorImager.

CheY~P-dependent Phosphorylation of CheA-P1—CheY~P was generated by incubating 2.0 μM CheY for 40 min at 24 °C in the presence of 40 mM [^{32}P]acetylphosphate. Under these conditions, steady-state levels of CheY~P were produced that did not increase with the addition of higher concentrations of acetylphosphate. P1 was then added such that the final reaction mixture contained 1.0 μM CheY~P, 100 μM P1, 50 mM Tris-HCl, pH 7.5, 12.5 mM NaCl, 0.5 mM dithiothreitol, 20 mM $MgCl_2$, and 20 mM acetylphosphate. At the indicated times, 10-μl aliquots were removed and analyzed for P1 phosphorylation as described above.

RESULTS

Structure of the CheA-P1 Domain—The P1 domain of CheA forms a globular and rather compact all α structure, with 77.9% of its residues in helical conformation. The structure was analyzed using the program PROMOTIF (35). The fold (Fig. 1) is composed of a short 3_{10} helix at the N terminus followed by 5 α-helices (A, B, C, D, and E) 29–36 Å long. Helices A, B, C, and D form an antiparallel four α-helix bundle with an up-and-down topology and a left-handed twist, an arrangement shown to be favored by electrostatic and nonbonded interactions (36). The average values of Ω_0, the angle that describes the relative orientation of pairs of α-helices forming the bundle, are $-156.2 \pm 3.5°$ and $33.8 \pm 2.7°$ for adjacent and diagonal neighbors, respectively. The average value for the closest distances between adjacent helices is 9.2 ± 0.9 Å. The C-terminal E helix runs parallel to C and antiparallel to D at an equal distance (10.3 Å) to both of them. Loop-helix interactions play a critical role in stabilizing this type of fold (37). In the P1 domain, α-helices are connected by one γ turn and three short loop segments. On the top of the bundle (Fig. 1), the loop connecting helices A and B (residues 30–36) is formed from two consecutive β turns involving two proline residues that interact with the hydrophobic core. The connection between helices C and D (residues 78–84) is rather extended with an alternate pattern of exposed polar and buried hydrophobic residues. At the bottom of the bundle, the loop connecting helices B and C (residues 57–59) has a marked hydrophobic character. Two phenylalanines (Phe-57 and Phe-59) together with Tyr-9, Phe-12, Phe-13, and Tyr-105 form a cluster of stacked aromatic side chains (Fig. 1). This patch covers the cross-section of the bundle and is partly topped by polar and charged residues from the N-terminal 3_{10} helix and from the loop connecting helices D and E (residues 107–112). The N termini of helices B, D, and E are capped by the side chains of Asp-36, Asn-84, and Asp-112, which provide hydrogen bonds to the main chain nitrogen atoms of residues 39, 87, and 115, respectively. All five helices are amphipathic. Charged and polar groups are mostly exposed on the surface whereas most hydrophobic side chains either participate in lateral inter-helix interactions or are buried inside the hydrophobic core of the helical bundle. A cluster of methionine side chains from residues 23, 69, and 98 occupies the exact center of the core. The only polar inter-helix interactions were found between helices A and D and between helices B and C. On the face of the bundle formed by the A and D helices, the A helix Glu-24 γ-carboxylate is at 3.1 Å from the D-helix Lys-95, and the A helix Asp-17 β-carboxylate is engaged in a network of salt bridges and hydrogen-bonding interactions involving the Lys-106 ϵ-amino group at a 2.8-Å distance and the side chains of Gln-99 and Asp-103 all from helix D (Fig. 1). On the opposite side of the helical bundle, the side chains of His-48 (the site of phosphorylation) and Lys-51 from helix B interact with the carboxylate oxygen atoms of Glu-70 from helix C, and a hydrogen bond is formed between the side chains of Asn-41 from helix B and Arg-77 from helix C (Fig. 1).

CheA-P1 Is an "Orthodox" HPt Domain—The two-component paradigm was thought initially to involve only two conserved modules, histidine kinases and response regulators (11). In recent years it has become clear that many systems have a third conserved module, the histidine phosphotransfer or HPt domain. The structures of HPt domains from *S. cerevisiae*, Ypd1, and *E. coli*, $ArcB_C$, are remarkably similar to one another despite a low level of sequence identity (14, 15). The crystal structure of CheA-P1 clearly establishes that this domain is a member of the HPt family. In the structure reported here, the four-helix bundle formed by αA, αB, αC, and αD is very similar to the corresponding four-helix bundle in the crystal structures of $ArcB_C$ and Ypd1 (root mean square difference values of 1.5 and 1.2 Å, respectively for 73 common C-α atoms) (Fig. 2). Whereas all three proteins differ significantly in loops

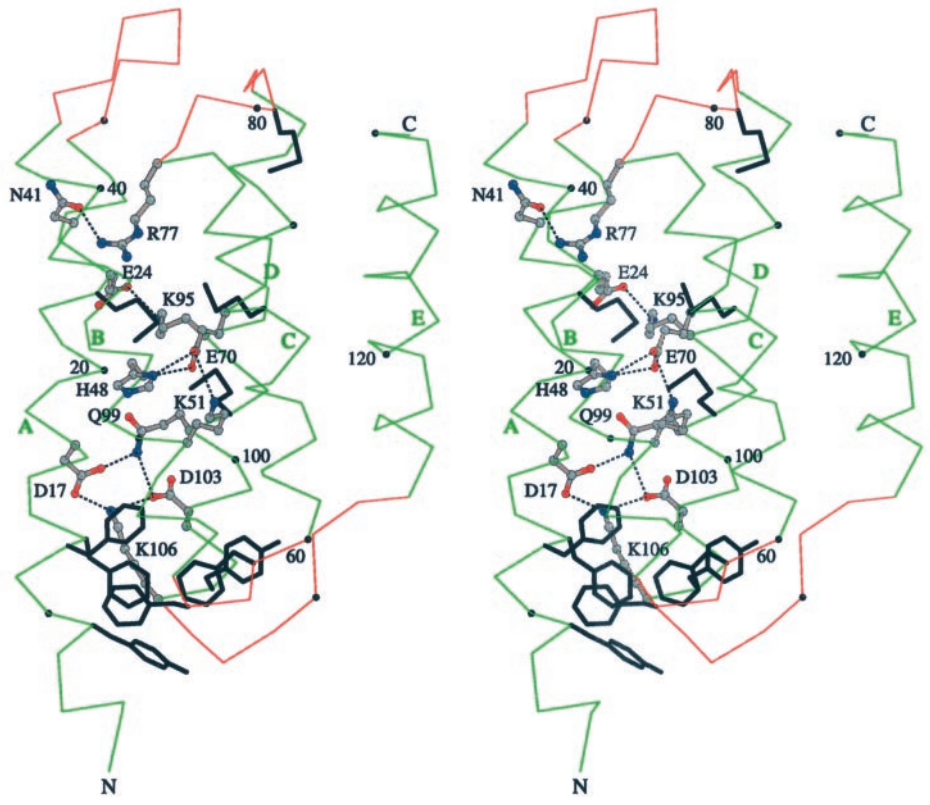

FIG. 1. **Stereo view of the α-carbon trace of the P1 domain of CheA.** The domain is composed of five helices (*green*) with connecting loops (*red*). Every 10th C-α is indicated by a *black dot*. The four SeMet residues and the aromatic patch at the bottom base of the bundle are displayed with *thick lines*. Residues involved in polar or charged inter-helix interactions (shown as *black dotted lines*) are displayed as ball-and-stick. Oxygen, nitrogen, and carbon atoms are shown in *red*, *blue*, and *gray*, respectively. The figure was produced with BOBSCRIPT (59).

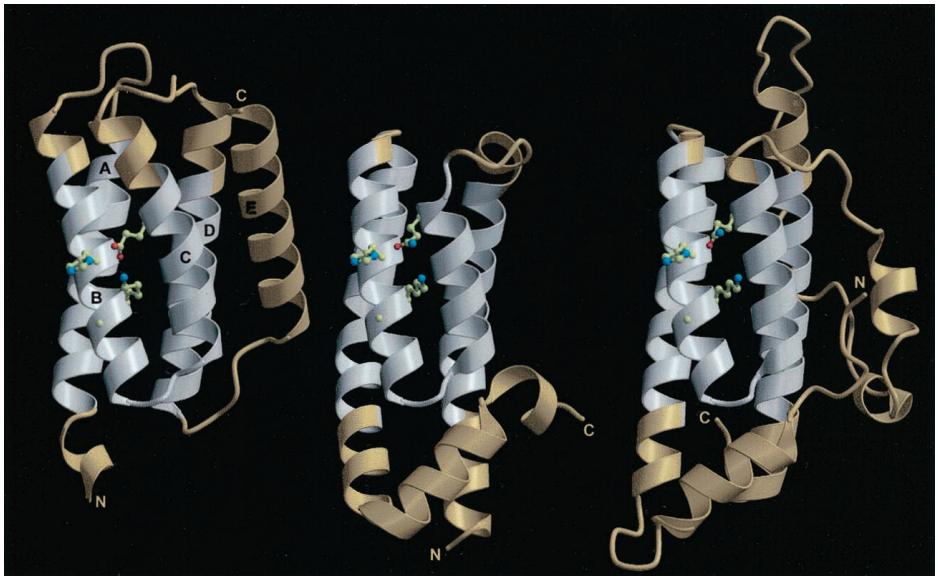

FIG. 2. **Comparison of P1 with known HPt domains.** Ribbon diagrams comparing the overall structures of P1 from *Salmonella* (*left*), the HPt domain of *E. coli* ArcB (*middle*, PDB entry code 2A0B), and yeast Ypd1 (*right*, PDB entry code 1QSP or 1C02). The three structures are shown in the same orientation. The structural motif used for the superposition of $ArcB_C$ (residues 683–704, 707–739, and 747–764) and Ypd1 (residues 31–52, 56–88, and 140–157) on P1 (residues 8–29, 40–72, and 90–107), leading to root mean square difference values of 1.5 and 1.2 Å for 73 C-α atoms, respectively, is shown in *white-gray*. The side chains of the phosphorylatable histidine residue and of the conserved Lys, Gly, and Glu/Gln residues of the active site are displayed as ball-and-stick. Oxygen, nitrogen, and carbon atoms are shown in *red*, *blue*, and *yellow*, respectively. The figure was produced with BOBSCRIPT (59) and rendered in RASTER3D (60).

between helices and in their N- and C-terminal flanking regions, the central four-helix bundle is completely conserved.

Active Site Conformation—Different HPt domains share very little sequence identity (<10%). There are a few conserved residues, however, and these are found in P1 as well as in other HPt proteins (Fig. 3). These amino acids appear to define critical features of the active site. The site of phosphorylation is invariably located at the position corresponding to that of His-48 in the B helix of CheA-P1. In the P1 structure, the electron density of His-48 is well defined in all four molecules of

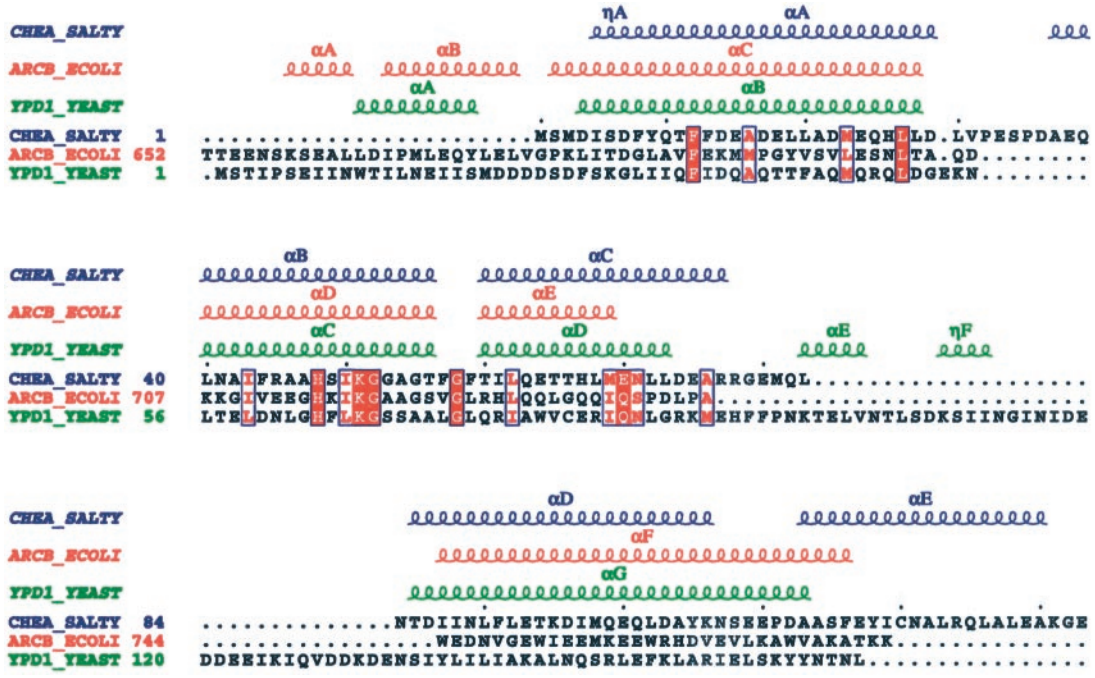

FIG. 3. **Structure-based sequence alignment.** Secondary structure elements were calculated using the program DSSP (61). α-Helices are shown as *coils* and labeled according to the literature. Homologous residues are in *red*, and identical residues are shown as *white letters on a red background*. This figure was created using ESPript (62).

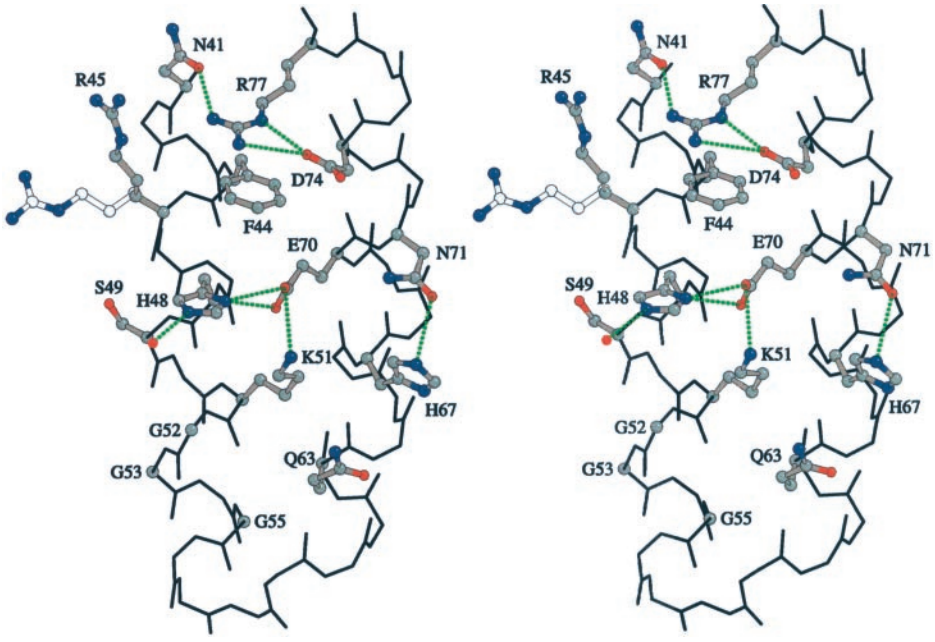

FIG. 4. **Stereo view of the α-helical hairpin and the environment around His-48.** Charged and hydrogen-bonding interactions are represented by *green dotted lines*. The two conformations observed for the side chain of Arg-45 caused by different crystal environments are represented with *solid* and *open bonds*. The figure was produced with BOBSCRIPT (59).

the asymmetric unit, and the imidazole side chain atoms have lower B factors than the average value (~18 *versus* 24.9 Å^2). In three of the four molecules in the asymmetric unit, His-48 is part of a hydrogen-bond network involving Lys-51 in helix B and Glu-70 in helix C (Fig. 4). One Glu-70 γ-carboxylate oxygen is within 2.8 Å of the ND1 atom of the His-48 imidazole, and the other is within 3.2 Å of the ϵ-amino of Lys-51. This geometry is in line with an ND1H tautomeric neutral form suggested from the NMR studies of the P1 domain of CheA from *E. coli* (38) and increases the nucleophilicity of the NE2 atom for a more efficient phosphotransfer chemistry. A Lys and a Glu or Gln occupy the corresponding positions in other HPt proteins. Another feature of the active site histidine is its exposure to solvent, especially in the distal region along the B helix, where three consecutive glycine residues (52, 53, and 55) provide an opening for ready access to the His-48 side chain. This feature also seems to be conserved among HPt domains.

The details of the active site presented above (Fig. 4) are virtually identical in three of the four P1 domains within the asymmetric unit. In these structures the guanidinium group of Arg-77 is involved in a cation-π interaction (39) with the aromatic ring of Phe-44, with the two side chains stacked and parallel to each other. In this geometry, the side-chain of Arg-77 is also involved in a polar interaction with the carboxylate group of Asp-74. The fourth P1 domain within the asymmetric unit has a distinctly different conformation in this re-

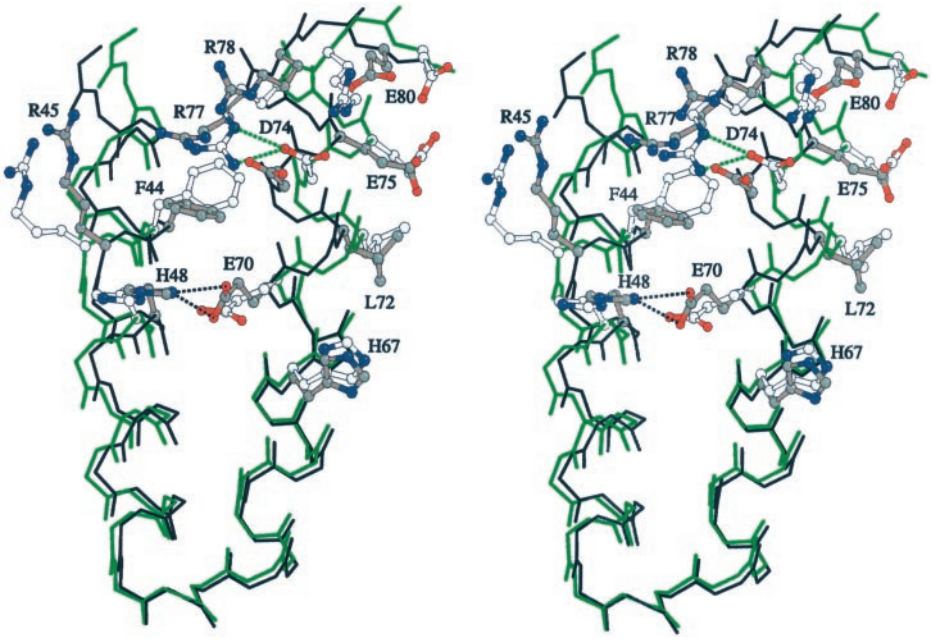

FIG. 5. **Stereo view of the helical hairpin in molecules A (*black backbone atoms* and *gray bonds* for side chains) and B (*green backbone atoms* and *open bonds* for side chains) of the asymmetric unit.** The side chains of residues that undergo the most important shifts are represented. The important polar interactions between the side chains of His-48 and Glu-70 and disrupted in molecule B are represented as *black dotted lines*. Polar interactions between the side chains of Asp-74 and Arg-77 are present in both molecules but were represented only in molecule B for the sake of clarity. The figure was produced with BOBSCRIPT (59).

gion. This arises from crystal packing interaction that brings a tyrosine side chain from a neighboring molecule at van der Waals contact to Arg-77. The displacement of the arginine side chain induces a movement of Phe-44, which preserves the cation-π interaction. This leads to significant rearrangements encompassing helices B and C and their connections to helices A and D. The side chain of Phe-44 adopts an internal rotamer and pushes apart the N-terminal half of helix B and the C-terminal half of helix C by 1–2 Å (Fig. 5). Other side chains with significantly altered positions include Arg-45 (root mean square difference = 3.3 Å), Leu-72 (2.5 Å), Asp-74 (1.9 Å), Arg-78 (4.8 Å), and Glu-80 (3.2 Å). These changes are associated with a profound effect at the active site. His-48 and Glu-70 are displaced by 1.8 and 1.4 Å, respectively, and the hydrogen-bond interaction otherwise formed between these two residues is disrupted. The observation of significant structural differences in the neighborhood of His-48 on the surface that interacts with protein partners suggests that protein-protein interactions, mimicked here by crystal packing, may induce functionally significant structural rearrangements.

Mutations of Conserved HPt Residues Block the ATP-dependent Phosphorylation of P1 by the CheA Histidine Kinase Core—Lys-51 and Glu-70 are conserved in HPt domains. Surprisingly, however, mutagenesis of the corresponding residues in Ypd1 (K67A and Q86A) seems to have little effect on the ability to transfer phosphoryl groups to and from response regulator proteins (40). Similar results have been obtained with the corresponding Gln→Ala mutation in $ArcB_C$ (41). We have used site-directed mutagenesis to construct the K51A and E70A variants of the CheA HPt domain and assayed the ability of these mutant proteins to be phosphorylated by ATP in the presence of the CheA kinase catalytic core. The results indicate a dramatic (>99%) reduction in the rate of phosphotransfer from ATP (Fig. 6). In contrast, mutations of these residues have little effect on the phosphotransfer reaction with the chemotaxis response regulator, CheY (Fig. 7). Phosphotransfer from phospho-CheY to P1 K51A was indistinguishable from transfer to the wild-type P1 domain. The initial rate of phosphotransfer to P1 E70A also seemed similar to wild type, but in this case the final level of P1 phosphorylation was significantly higher than wild type, indicating that the glutamate tends to favor phosphotransfer to CheY. Thus, in all HPt domains that have been examined, the highly conserved Lys and Glu/Gln residues

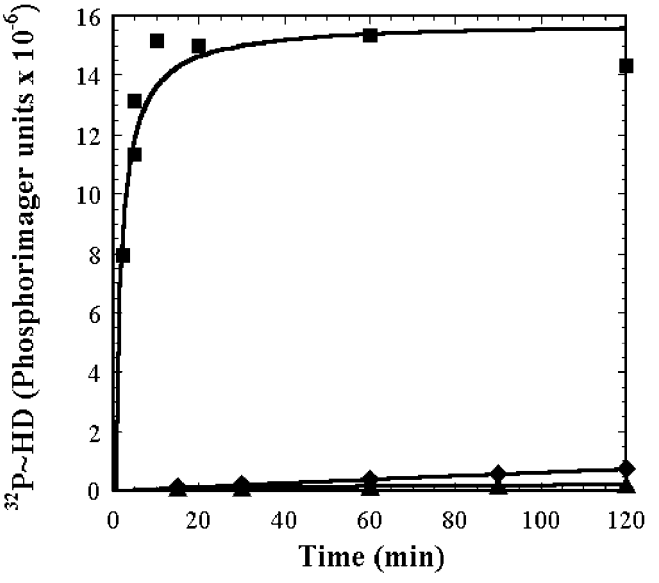

FIG. 6. **Time course of CheA-P1 phosphorylation by ATP.** A solution containing 100 μM of wild-type CheA-P1 (■) or the mutants K51A (♦) or E70A (▲) were incubated at 24 °C in the presence of 10 μM CheA catalytic domain and [γ-^{32}P]ATP. Aliquots were taken at the indicated times and analyzed as described under "Experimental Procedures."

at the active site do not seem essential for phosphotransfer reactions with response regulators. In the case of CheA, at least, these residues are required for kinase-mediated phosphotransfer from ATP.

DISCUSSION

The x-ray crystal structure of the CheA-P1 domain provides the first high resolution view of a phosphotransfer domain that is phosphorylated by ATP through the activity of a histidine kinase and subsequently acts as a phosphodonor for a response regulator. Previous structural studies had indicated that P1 was distantly related to HPt domains such as *E. coli* $ArcB_C$ and *S. cerevisiae* Ypd1, which shuttle phosphoryl groups between response regulator aspartates. The x-ray structure of P1 indicates a much higher degree of similarity than had been realized previously. P1, $ArcB_C$, and Ypd1 share a common up-down up-down four-helix bundle core with the site of histidine phos-

phorylation located at the same position on the second helix. This histidine is followed by several highly conserved residues that define the HPt active site (Figs. 2 and 4). Aside from a few other conserved residues that seem to play important structural roles, HPt domains exhibit very little sequence conservation, and there are generally additional helices formed from connecting loops or flanking sequences that add to this diversity.

In recent reviews, the term HPt domain has been used in a generic sense in reference to any two-component histidine phosphotransfer domain including the dimerization domains in EnvZ and Spo0B as well as monomeric structures such as CheA-P1, $ArcB_C$, and Ypd1 (3). This type of functional designation is misleading insofar as the monomeric and dimeric phosphotransfer domains are structurally distinct (Fig. 8), and there are numerous examples of other structurally unrelated domains with histidine phosphotransfer functions that are not part of two-component signal transduction systems. These would include the phospho-histidine proteins that mediate sugar transport and carbohydrate repression in bacteria (42) as well as a large number of metabolic enzymes such as nucleoside diphosphate kinase (43) and succinyl-CoA synthetase (44). Moreover, the CheA dimerization domain, which does not have a histidine phosphotransfer activity, seems homologous to the H box-containing dimerization domains of histidine kinases such as EnvZ (4, 8). Here we reserve the HPt designation solely for members of the P1, $ArcB_C$, and Ypd1 family. It should be noted that although the well established function of Spo0B in relaying phosphoryl groups between response regulators Spo0F and Spo0A to control sporulation in *Bacillus subtilis* has been used as a model for understanding the phosphotransfer activities of HPt domains like ArcBc and Ypd1, Spo0B has a very different structure (45, 46) (Fig. 8), and the role of the CheA-P1 domain in mediating phosphotransfer from ATP to response regulators does not seem to have a correlate in the functioning of Spo0B.

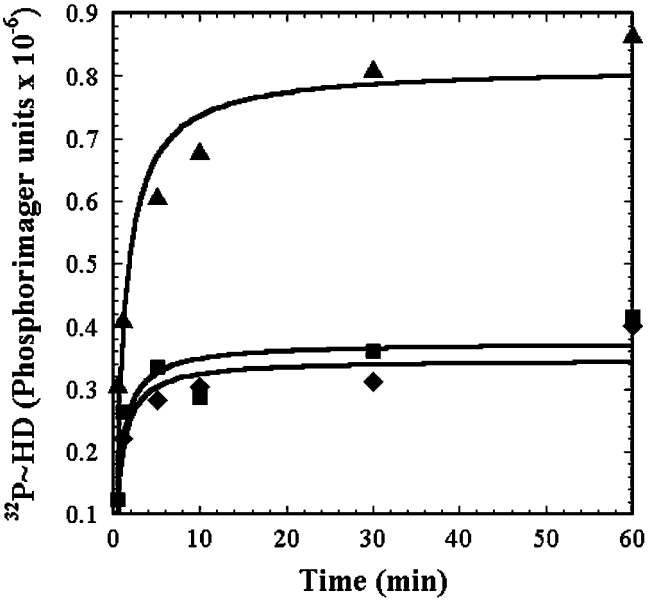

FIG. 7. **Time course of phosphotransfer from phospho-CheY to CheA-P1.** A 1.0 μM solution of phospho-CheY was incubated at 24 °C with 100 μM CheA-P1, wild type (■), mutants K51A (♦), or E70A (▲), and the appearance of phospho-P1 was monitored as a function of time as described under "Experimental Procedures."

The bacterial chemotaxis system is by far the most thoroughly investigated of all so-called two-component signal transduction networks, thus one would imagine that the function of the CheA HPt domain would serve as a model for understanding the role of HPt modules in less well characterized signaling systems like those involving ArcB and Ypd1. The high degree of similarity between the CheA-P1 domain and other HPt domains suggests a common underlying mechanism. All known P1-like HPt proteins can clearly function to shuttle phosphoryl groups between response regulators. Indeed, *in vitro* studies have shown that phospho-CheA-P1 from *Salmonella* can reversibly phosphorylate a wide range of different response regulators including NtrC and OmpR from *E. coli*, Spo0F and Spo0A from *B. subtilis*, and RegA from *Dictyoste-*

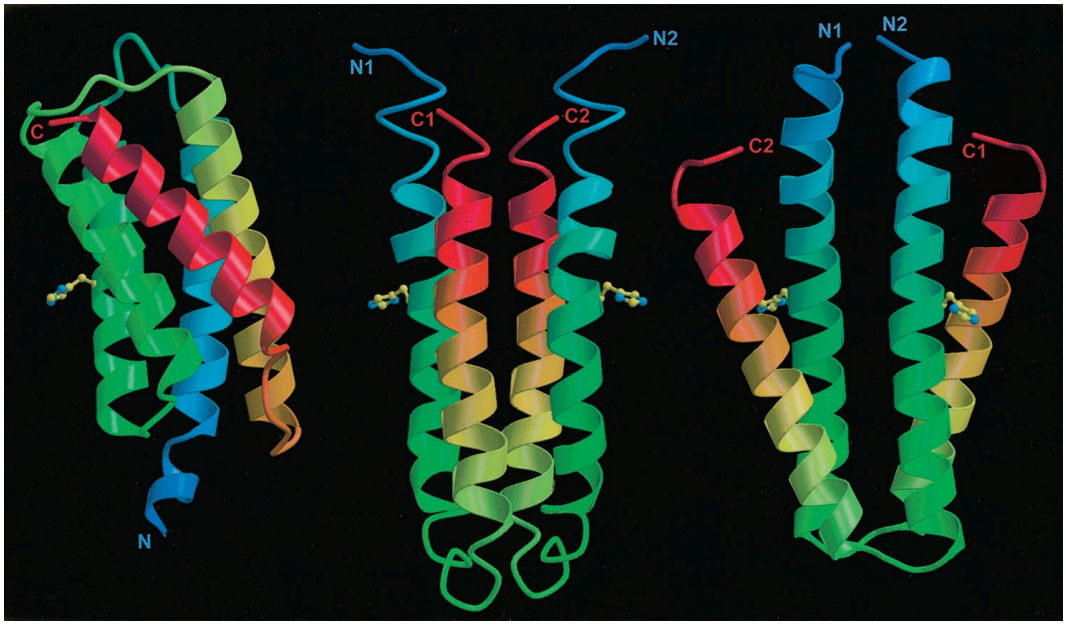

FIG. 8. **Phosphoaccepting domains from CheA, EnvZ, and Spo0B.** *Left*, *S. typhimurium* CheA-P1; *Middle*, *E. coli* EnvZ dimerization domain (PDB entry code 1JOY); *Right*, *B. subtilis* Spo0B dimerization domain (1IXM). The color of each polypeptide chain varies from *dark blue* (N termini) to *red* (C termini). The four-helix bundles of EnvZ and Spo0B result from dimerization. The N and C termini of these subunits are labeled. The structures are oriented such that one active site histidine residue (ball-and-stick) and the helix to which it belongs have similar positions and orientations. The figure was produced with BOBSCRIPT (59) and rendered in RASTER3D (60).

lium (47–49) in addition to its physiologically relevant targets, CheY and CheB.

Rates of phosphotransfer from the isolated phospho-P1 domain to CheY are considerably reduced compared with rates of transfer from the intact phospho-CheA protein. This seems to be primarily because of the fact that phospho-P1 has a very low affinity for CheY (50). Thus it has not been possible to detect any saturation in the rate of phosphotransfer even at concentrations of CheY up to 40 μM (50) so that rates determined for the P1-CheY phosphotransfer reaction are linearly dependent on the concentrations of both P1 and CheY. Phosphotransfer from intact CheA to CheY is saturable with an apparent K_m value for CheY of ~7 μM (51). The k_{cat} estimates obtained with CheA may not reflect saturating rates of transfer from phospho-P1 to CheY, however, because CheY binds to CheA independently of P1 at a second domain, P2 (52, 53). Thus, the response regulator specificity of CheA for CheY may be determined largely by the CheA P2 domain, which follows in sequence the P1 domain and is responsible for binding unphosphorylated CheY.

Although in chemotaxis the ability of P1 to transfer phosphoryl groups between response regulators may be a physiologically important activity (54), the primary function of P1 is to mediate the transfer of phosphoryl groups from ATP. Is it possible that other HPt domains have a similar activity? The best characterized HPt-containing histidine kinase besides CheA is the *E. coli* ArcB protein, and *in vitro* studies have shown clearly that the ArcB histidine kinase can directly phosphorylate the ArcB HPt domain in the absence of a response regulator domain (55, 56). This result raises the possibility for a direct kinase-dependent HPt phosphorylation mechanism in ArcB that may be similar to that which has been characterized in CheA.

The geometry and amino acid side chain chemistries around the phosphoaccepting histidine are conserved in most HPt domains that have been identified. Alanine substitution mutagenesis of these conserved residues in $ArcB_C$ (41) and Ypd1 (40) had little effect on the ability of these proteins to transfer phosphoryl groups to and from response regulator proteins. This is not entirely surprising, because the response regulators can phosphorylate themselves using small molecule phosphodonors such as acetylphosphate, phosphoramidate, and phosphoimidazol (57, 58). The corresponding alanine substitution mutations in P1, K51A and E70A, also have little effect on phosphoryl transfer reactions to the response regulator CheY. In contrast, these mutants were no longer subject to ATP-dependent phosphorylation by the CheA kinase core (Fig. 6). These results indicate that the conserved active site in HPt domains may be primarily designed to mediate ATP-dependent phosphotransfer events associated with histidine kinase activity. Phosphotransfer between response regulators may be a secondary reaction that stems from the solvent-exposed nature of the phospho-histidine group as well as the complementarity between the surface around this site and the surface around the response regulator aspartate.

Acknowledgments—We thank Mikhail Levit and Fred Hughson for their help with the initial characterization and crystallization of P1, the scientific staff of European Molecular Biology Laboratory at Deutsches Elektronen Synchrotron (Hamburg, Germany) and European Synchrotron Radiation Facility (Grenoble, France) for excellent data collection facilities, and Ana González for stimulating discussions about multiwavelength anomalous dispersion data collection.

REFERENCES

1. Armitage, J. P. (1999) *Adv. Microb. Physiol.* **41,** 229–289
2. Aizawa, S. I., Harwood, C. S., and Kadner, R. J. (2000) *J. Bacteriol.* **182,** 1459–1471
3. Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000) *Annu. Rev. Biochem.* **69,** 183–215
4. Bilwes, A. M., Alex, L. A., Crane, B. R., and Simon, M. I. (1999) *Cell* **96,** 131–141
5. Dutta, R., Qin, L., and Inouye, M. (1999) *Mol. Microbiol.* **34,** 633–640
6. Grebe, T. W., and Stock, J. B. (1999) *Adv. Microb. Physiol.* **41,** 139–227
7. Koretke, K. K., Lupas, A. N., Warren, P. V., Rosenberg, M., and Brown, J. R. (2000) *Mol. Biol. Evol.* **17,** 1956–1970
8. Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S. K., Zhu, Y., Ishima, R., Liu, D., Tong, K. I., Kurokawa, H., Qian, H., Inouye, M., and Ikura, M. (1999) *Nat. Struct. Biol.* **6,** 729–734
9. Hess, J. F., Oosawa, K., Kaplan, N., and Simon, M. I. (1988) *Cell* **53,** 79–87
10. Zhou, H., Lowry, D. F., Swanson, R. V., Simon, M. I., and Dahlquist, F. W. (1995) *Biochemistry* **34,** 13858–13870
11. Stock, J. B., Stock, A. M., and Mottonen, J. M. (1990) *Nature* **344,** 395–400
12. Sanders, D. A., Gillece-Castro, B. L., Stock, A. M., Burlingame, A. L., and Koshland, D. E. (1989) *J. Biol. Chem.* **264,** 21770–21778
13. Kato, M., Mizuno, T., Shimizu, T., and Hakoshima, T. (1997) *Cell* **88,** 717–723
14. Xu, Q., and West, A. H. (1999) *J. Mol. Biol.* **292,** 1039–1050
15. Song, H. K., Lee, J. Y., Lee, M. G., Moon, J., Min, K., Yang, J. K., and Suh, S. W. (1999) *J. Mol. Biol.* **293,** 753–761
16. Matsubara, M., Kitaoka, S. I., Takeda, S. I., and Mizuno, T. (2000) *Genes Cells* **5,** 555–569
17. Levit, M., Liu, Y., Surette, M., and Stock, J. (1996) *J. Biol. Chem.* **271,** 32057–32063
18. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) *EMBO J.* **9,** 1665–1672
19. Jancarik, J., and Kim, S. H. (1991) *J. Appl. Crystallogr.* **24,** 409–411
20. Leslie, A. G. W. (1987) in *Proceedings of the Daresbury Study Weekend* (Helliwell, J. R., Machin, P. A., and Papiz, M. Z., eds) pp. 39–50, Science and Engineering Research Council, Daresbury Laboratory, Warrington, United Kingdom
21. Collaborative Computational Project (1994) *Acta Crystallogr. Sect. D* **50,** 760–763
22. Terwilliger, T. C., Kim, S.-H., and Eisenbeg, D. (1987) *Acta Crystallogr. Sect. A* **43,** 1–5
23. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D* **55,** 849–861
24. Evans, G., and Pettifer, R. F. (2001) *J. Appl. Crystallogr.* **34,** 82–86
25. de La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol.* **276,** 472–494
26. Boissy, G., de La Fortelle, E., Kahn, R., Huet, J. C., Bricogne, G., Pernollet, J. C., and Brunie, S. (1996) *Structure* **4,** 1429–1439
27. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr. Sect. D* **52,** 30–42
28. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
29. Navaza, J. (1994) *Acta Crystallogr. Sect. A* **50,** 157–163
30. Brünger, 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* **54,** 905–921
31. Brünger, A. T. (1992) *Nature* **355,** 472–475
32. Read, R. J. (1986) *Acta Crystallogr. Sect. A* **42,** 140–149
33. Roussel, A., and Cambillau, C. (1989) in *Silicon Graphics Geometry Partner Directory* (Graphics, S., ed) pp. 71–78, Silicon Graphics, Mountain View, CA
34. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
35. Hutchinson, E. G., and Thornton, J. M. (1996) *Protein Sci.* **5,** 212–220
36. Chou, K. C., Maggiora, G. M., Nemethy, G., and Scheraga, H. A. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 4295–4299
37. Chou, K. C., Maggiora, G. M., and Scheraga, H. A. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 7315–7319
38. Zhou, H., and Dahlquist, F. W. (1997) *Biochemistry* **36,** 699–710
39. Gallivan, J. P., and Dougherty, D. A. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 9459–9464
40. Janiak-Spens, F., and West, A. H. (2000) *Mol. Microbiol.* **37,** 136–144
41. Matsushika, A., and Mizuno, T. (1998) *Biosci. Biotechnol. Biochem.* **62,** 2236–2238
42. McEvoy, M. M., and Dahlquist, F. W. (1997) *Curr. Opin. Struct. Biol.* **7,** 793–797
43. Dumas, C., Lascu, I., Morera, S., Glaser, P., Fourme, R., Wallet, V., Lacombe, M. L., Veron, M., and Janin, J. (1992) *EMBO J.* **11,** 3203–3208
44. Fraser, M. E., James, M. N., Bridger, W. A., and Wolodko, W. T. (2000) *J. Mol. Biol.* **299,** 1325–1339
45. Varughese, K. I., Madhusudan, Zhou, X. Z., Whiteley, J. M., and Hoch, J. A. (1998) *Mol. Cell* **2,** 485–493
46. Zapf, J., Sen, U., Madhusudan, Hoch, J. A., and Varughese, K. I. (2000) *Structure Fold. Des.* **8,** 851–862
47. Ninfa, A. J., Ninfa, E. G., Lupas, A. N., Stock, A., Magasanik, B., and Stock, J. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 5492–5496
48. Olmedo, G., Ninfa, E. G., Stock, J., and Youngman, P. (1990) *J. Mol. Biol.* **215,** 359–372
49. Thomason, P. A., Traynor, D., Stock, J. B., and Kay, R. R. (1999) *J. Biol. Chem.* **274,** 27379–27384
50. Stewart, R. C., Jahreis, K., and Parkinson, J. S. (2000) *Biochemistry* **39,** 13157–13165
51. Stewart, R. C. (1997) *Biochemistry* **36,** 2030–2040
52. Welch, M., Chinardet, N., Mourey, L., Birck, C., and Samama, J. P. (1998) *Nat. Struct. Biol.* **5,** 25–29
53. McEvoy, M. M., Hausrath, A. C., Randolph, G. B., Remington, S. J., and Dahlquist, F. W. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 7333–7338
54. Sourjik, V., and Schmitt, R. (1998) *Biochemistry* **37,** 2327–2335
55. Ishige, K., Nagasawa, S., Tokishita, S., and Mizuno, T. (1994) *EMBO J.* **13,**

5195–5202
56. Tsuzuki, M., Ishige, K., and Mizuno, T. (1995) *Mol. Microbiol.* **18,** 953–962
57. Lukat, G. S., McCleary, W. R., Stock, A. M., and Stock, J. B. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 718–722
58. Silversmith, R. E., Appleby, J. L., and Bourret, R. B. (1997) *Biochemistry* **36,** 14965–14974
59. Esnouf, R. M. (1997) *J. Mol. Graph. Model* **15,** 132–134
60. Meritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524
61. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637
62. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics* **15,** 305–308
63. Sasaki, S. (1989) *KEK Report* **88-14,** 1–136