Structural and Mutational Analysis of the PhoQ Histidine Kinase Catalytic Domain

INSIGHT INTO THE REACTION MECHANISM*

Received for publication, June 29, 2001, and in revised form, August 6, 2001
Published, JBC Papers in Press, August 7, 2001, DOI 10.1074/jbc.M106080200

PhoQ is a transmembrane histidine kinase belonging to the family of two-component signal transducing systems common in prokaryotes and lower eukaryotes. In response to changes in environmental Mg$^{2+}$ concentration, PhoQ regulates the level of phosphorylated PhoP, its cognate transcriptional response-regulator. The PhoQ cytoplasmic region comprises two independently folding domains: the histidine-containing phosphotransfer domain and the ATP-binding kinase domain. We have determined the structure of the kinase domain of Escherichia coli PhoQ complexed with the non-hydrolyzable ATP analog adenosine 5'-(β,γ-imino)triphosphate and Mg$^{2+}$. Nucleotide binding appears to be accompanied by conformational changes in the loop that surrounds the ATP analog (ATP-lid) and has implications for interactions with the substrate phosphotransfer domain. The high resolution (1.6 Å) structure reveals a detailed view of the nucleotide-binding site, allowing us to identify potential catalytic residues. Mutagenic analyses of these residues provide new insights into the catalytic mechanism of histidine phosphorylation in the histidine kinase family. Comparison with the active site of the related GHL ATPase family reveals differences that are proposed to account for the distinct functions of these proteins.

Two-component signaling systems are used ubiquitously by prokaryotes and also by a number of lower eukaryotes to sense and respond to various environmental conditions. These systems consist of a histidine kinase that acts as the sensor of environmental stimuli and a response regulator that mediates the cellular response, generally at the level of transcriptional control (1). As with many signaling pathways, protein phosphorylation is used as a means to transmit information; however, unlike the majority of phosphoproteins found in higher eukaryotes, in which tyrosine, serine, or threonine serve as the substrate for phosphorylation, histidine kinases autophosphorylate a histidine residue from which the phosphoryl group is subsequently transferred to a conserved aspartate residue in the response regulator. The catalytic mechanism is reasonably well understood for aspartyl phosphorylation, while far less is known about the autokinase reaction. This lack of information is due in part to the relative scarcity of detailed structural information for the histidine kinases. Recently, structural information has become available for the CheA (2, 3) and EnvZ (4) histidine kinases. These structures reveal that the catalytic ATP-binding domain is an autonomously folding α/β-sandwich that shares structural homology with a family of ATPases that include Hsp90, DNA gyrase B, and MutL (5). Although these structures provide some insight into function, they have not allowed the assignment of catalytic residues. Here we describe the 1.6-Å resolution crystal structure of the catalytic domain of the PhoQ histidine kinase complexed with an AMPPNP$^3$ nucleotide. PhoQ is a transmembrane histidine kinase that is involved in Mg$^{2+}$ homeostasis and/or pathogenesis of a number of Gram-negative bacteria (for review see Ref. 6). PhoQ responds to limiting concentrations of extracellular Mg$^{2+}$ by increasing the net phosphorylation of the PhoP transcriptional response-regulator. The cytoplasmic portion of PhoQ consists of independently folded phosphotransfer and ATP-binding catalytic domains. The high degree of resolution reveals a detailed view of the nucleotide-binding site that provides new insights into catalytic mechanism.

EXPERIMENTAL PROCEDURES

Plasmids—The plasmids used for expression of the PhoQ cytoplasmic domain (residues 219–486, which contains the phosphotransfer and catalytic domains) and the PhoQ catalytic domain (residues 331–486) were constructed as follows. For the cytoplasmic domain, DNA corresponding to codons 219–486 was PCR-amplified from plasmid pLPQ2 (7), which contains a full-length copy of the phoQ gene, cut with NdeI and EcoRI (synthetic sites were introduced in the PCR primers), and cloned into the NdeI-EcoRI backbone of pAEQDv (8). The resulting plasmid, pAEQD4TR, has an initiating ATG codon (at the synthetic NdeI site) fused to codon 219 of phoQ and all of the remainder of the gene including the native translational termination codon. In this plasmid, expression of the PhoQ cytoplasmic domain is controlled by a plasmid-borne T7 3010 promoter and ribosome binding site. Variants of pAEQD4TR were constructed in which the Lys-392, Arg-434, or Arg-439 codons were independently replaced with alanine codons. These mutant genes were constructed by encoding the mutant codon in a PCR primer that also contained a nearby unique restriction site, PCR-am-

* This work was supported in part by a postdoctoral fellowship from the Spanish Science Ministry (to A. M.) and by Grants AI41566 (to C. D. W.) and GM34102 (to W. A. H.) from the National Institutes of Health. 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 1ID0) 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 should be addressed. Tel.: 212-305-0251; Fax: 212-305-1468; E-mail: cdw15@columbia.edu.

1 The abbreviations used are: AMPPNP, adenosine 5'-(β,γ-imino)triphosphate; PCR, polymerase chain reaction; r.m.s., root mean square; MAD, multiwavelength anomalous diffraction; MIR, multiple isomorphous replacement.
plifying a portion of the gene, and cloning the mutated PCR product into the pAED4TR plasmid. The PhoQ catalytic domain was cloned into pAED4 as described above for the cytoplasmic domain, except a promoter that fuses an initiating ATG codon to and amplifies DNA from codon 391 of *phoQ* was used. In the resulting plasmid, pAED4QKD, expression of the C-terminal PhoQ catalytic domain was controlled with a 1 clear solvent boundary but not secondary structural elements. The phases derived from MAD data were used to locate the heavy atom positions in difference Fourier maps. MAD and MIR data were combined, and atom parameters were refined using isomorphous and anomalous differences, with the maximum likelihood method incorporated in MLPHARE (14). The resultant phases to 2.75 Å were further improved by solvent flattening, and the resolution was extended to 1.6 Å using DM (14). The consequent map was interpretable showing several secondary structural elements.

The initial model, containing 120 of the 156 residues, was built using the ARP/wARP program (15). Combining the partial structure phases with experimental phases allowed manual tracing of the remaining residues with the program O (16). Three N-terminal and the five C-terminal residues are missing from the final structure. The model was subjected to interactive cycles of manual rebuilding and conjugated gradient minimization, simulated annealing, and individual B-factor refinement using the programs O and CNS (17). Both anisotropic B-factor and bulk solvent corrections were applied. A molecule of AMP-PNP, one Mg cation, and 189 water molecules were located during the refinement. Eleven residues present double conformation (residues 345, 348, 349, 351, 370, 371, 407, 414, 425, 465, and 471) in the final model. Stereochemistry checks indicate that the refined model is in quite good agreement for expectations for models within this resolution range (18). Statistics for the final model are given in Table I. Atomic coordinates for PhoQ-KD have been deposited with the Protein Data Bank, with accession code 1D10 (19).

*Structure and Sequence Comparisons*—Structures were superimposed based on α-carbon atoms alone, and only atom pairs identified as equivalent were used for r.m.s. deviation calculations. The least-squares superimpositions were calculated using LSQMAN (20). The root mean square deviations were used as the "Brute" option in the program with a cut-off distance criterion of 3.5 Å and a minimum fragment length of three consecutive residues. The coordinates were taken from the Protein Data Bank with entry codes: CheK-KD, 1b3q (2); EnvZ-KD, 1bdx (4); GyrB, 1ei1 (21); MutL, 1b62 (22); and Hsp90, 1bqy (23). Sequence alignments among these proteins were based on structural superimposition. The multiple alignments of each protein with their respective families were taken from Pfam data (18). Structures were compared with access number: histidine kinase family, PF00512 PF; GyrB family, PF00204; MutL family, PF01119; and Hsp90 family, PF00183.

**RESULTS AND DISCUSSION**

**Overall Fold**—The crystal structure of the C-terminal PhoQ catalytic domain (residues 331–486) complexed with the non-hydrolyzable ATP analog, AMP-PNP, was solved by combining experimental phases from MIR with phases from MAD, and this structure was refined at 1.6-Å resolution (Table I). The structure as shown in Fig. 1A is a two-layer α/β sandwich fold composed of a flat, mixed, five-stranded β sheet and three α helices (from left to right: βB, βD, βG, βF, and βE, and α1, α2, and α3, respectively), with dimensions 35 × 50 × 35 Å (Fig. 1). There is a deep cavity in one end where the AMPPNP molecule is localized, and the opposite end is closed by a small antiparallel β-sheet formed by the βA and βC strands. In the final model, segments at the C terminus (residues 331–335) and the C terminus (residues 481–486) are disordered and extend into the solvent. The region comprising residues 433–439 has high B-factors (51.4 Å² on average) compared with the average for the protein (23.4 Å²), indicating an intrinsic flexibility. Nevertheless, the electron density is sufficiently well defined to trace the region (Fig. 1C). This flexible loop is part of a long polypeptide segment that extends away from the rest of the molecule and has been termed the ATP-lid in the homologous EnvZ and CheA proteins (5). The proximity of this region to the bound AMP-PNP molecule and the flexibility indicated by high B-factors suggest important roles in ATP binding and interac-
sections with the phosphorylation site in the histidine phosphotransfer domain (see below).

Comparison with the Histidine Kinase Domains of CheA and EnvZ—The detailed three-dimensional structure of the homologous kinase domains of CheA and EnvZ has been reported previously (2, 4). The structure of PhoQ-KD is shown with the crystal structure of the Thermotoga maritima CheA kinase domain (2.6 Å resolution) and with the NMR structure of the E. coli EnvZ kinase domain (complexed with AMPPNP) in Fig. 2B and an alignment of the amino acid sequences based on the superimposition of the structural elements is shown in Fig. 2A. CheA-KD, which is longer than the two other kinase domain (189 residues compared with 156 and 160 residues, respectively, in PhoQ-KD and EnvZ-KD), possesses two extra α helices, one between α2 and βD, and another before the ATP-lid loop (Fig. 2). Roughly 70% of the superimposed residues are identical for all three proteins and these primarily cluster in the N, G1, F, and G2 boxes that have been classically defined in alignments of the histidine kinase superfamily (25) and in the recently defined G3 box (5). The remaining identical residues are mostly hydrophobic in character and are distributed between all of the structural elements where they generally participate in forming the core of the molecule (Fig. 2A). Surprisingly, the r.m.s. deviation between the backbone α-carbon atoms of the PhoQ-KD and CheA-KD models is lower (1.62 Å) than the r.m.s. deviation between the backbone traces of the more closely related (by sequence and structural organization) PhoQ-KD and EnvZ-KD proteins (1.84 Å). Although we cannot rule out the possibility that the greater deviation from the EnvZ-KD is genuine, it seems more likely that this may result of ambiguities in the overlaid NMR structures. Consequently, detailed structural comparisons were carried out with the CheA-KD structure. Besides the two additional helices in CheA-KD, the principal structural difference between them is the conformations of the high flexibly loop of the ATP-lid (Figs. 2B and 3). ATP-lid mobility is facilitated by the presence of three conserved glycine residues (441, 443, and 445 in PhoQ) of the G2 box and one following the F box (Gly-432 in PhoQ; Figs. 2A and 3). The loop is anchored at its N terminus by the conserved phenylalanine residue (Phe-429 in PhoQ) for which the F box is named and at its C terminus by a conserved hydrophobic residue (Leu-446 in PhoQ and Met-507 in CheA) that begins the α3 helix (Fig. 3). These two residues interact with each other as part of a larger hydrophobic patch in which Ile-428, Leu-446, and Ile-460 form a small pocket where Phe-429 is inserted (Fig. 3). The superimposition of CheA-KD shows a similar hydrophobic patch composed of Phe-487, Leu-486, Met-507, and Met-521 (Fig. 3). The chemical similarity of these residues and the similar organization in the EnvZ-KD model (Phe-387, Leu-386, Ile-408, and Leu-420; Fig. 2A) suggest that this hydrophobic cluster is a general structural feature of histidine kinases. In the nucleotide-free CheA-KD model, the loop is extended toward the solvent in an “open” conformation (Fig. 3). In contrast, the PhoQ loop is in close contact with the main β-sheet, forming a “closed” conformation. The largest distance between equivalent residues in the ATP lid of PhoQ and CheA is 10 Å (Gly-441 in PhoQ and Gly-502 in CheA; Fig. 3). A 30° rotation of the loop about the hydrophobic patch anchor would produce an open conformation in PhoQ that would be structurally similar to that seen in CheA-KD. Three loop residues (Arg-434, Arg-439, and Gln-442) make extensive contacts with the phosphate groups and the chelated Mg2+ ion of the bound AMPPNP molecule in the PhoQ-KD structure, suggesting that nucleotide binding may induce the closed conformation. Apparently, the interactions with the AMPPNP phosphates play the principal role in the loop reorganization since the hydrophobic patch superimposes with minimal differences between the two structures (Fig. 3). Mutagenesis studies of the proposed hinge of the ATP-lid in EnvZ have shown that this region is essential for kinase activity (26).

The alternate disposition of the β-hairpin between the βF and βG strands represents a second, but minor structural difference between the PhoQ-KD and CheA-KD structures (Fig. 3)). This β-hairpin corresponds to the recently defined G3 motif (5), which includes a conserved glycine residue (Gly-469 in PhoQ) that immediately precedes the βG strand (Fig. 2A). Structurally, the β-hairpin lines the back of the nucleotide-binding site and provides solvent-mediated contacts with the AMPPNP ring. The β-hairpin is placed in the PhoQ-KD structure relative to the CheA-KD structure (Fig. 3), and this
displacement exposes two hydrophobic residues (Met-466 and Leu-467 in PhoQ) to the solvent. Together with Phe-397 in the \( \beta \) strand, these residues form a solvent-exposed hydrophobic region. Perhaps ATP binding induces a shift in the position of the \( \beta \)-hairpin that, in turn, allows the hydrophobic patch to interact with the substrate histidine phospho-transfer domain, which is absent in this structure. These hydrophobic residues are not, however, conserved among histidine kinases, suggesting that the solvent exposure of the hydrophobic patch in PhoQ may simply be due to the space need of the ATP-ring pocket rather than a mechanistic function. Additional experiments will be necessary to address these possibilities.

The Nucleotide Binding Site—The ATP-binding pocket of PhoQ-KD involves not only absolutely conserved residues from the N, G1, F, G2, and G3 boxes, but also partially conserved residues from these motifs and from the ATP-lid loop (Figs. 2A and 4). In particular, residues from the G1, F, and G3 boxes provide the principal contacts with the adenosine moiety, whereas the N and G2 boxes contact both the adenosine moiety and Mg\(^{2+}\) phosphates. The ATP-lid interacts exclusively with the phosphates and the divalent cation. The hydrogen bond between the N6 amino group of AMPPNP and the carboxyl side chain of the conserved Asp-415 in the G1 box is the single direct protein interaction with the adenine ring (Fig. 4). Water mol-

**FIG. 1.** Protein fold of the *E. coli* PhoQ-KD. *A*, ribbon diagram of the overall protein fold completed with AMPPNP-Mg\(^{2+}\). The \( \alpha \) helices are colored gold and labeled \( \alpha 1-\alpha 3 \), \( \beta \) strands are in blue and labeled \( \beta A-\beta G \), and loops are shown in white. The bound AMPPNP molecule is shown in ball-and-stick representation, and the Mg\(^{2+}\) ion is represented as a cyan sphere. *B*, stereo C trace. Every tenth amino acid residue is indicated as a sphere and labeled with its residue number. The high B-factor segment is shown with dashed lines. The AMPPNP-Mg\(^{2+}\) molecule is drawn in a gray ball-and-stick representation. The orientation is the same as in panel *A*. *C*, simulated annealing 2*F*-*F* omit map (residues 433–442 removed) for the flexible loop contoured at 1 \( \sigma \) (cyan) and 2 \( \sigma \) (orange). The AMPPNP molecule is shown without density. Carbon, nitrogen, oxygen, and phosphate are drawn in yellow, blue, red, and green, respectively. Water molecules are omitted for clarity.
Nucleotide-bound PhoQ Kinase Domain Structure

Fig. 2. Comparison of histidine kinases catalytic domains. A, structure-based sequence alignment of histidine kinase ATP-binding domains. α helices are shown as yellow filled boxes, 3₁₀ helices as yellow intermediate shading, and β strands as blue arrows. The identical or similar residues in all sequences are highlighted in red or green, respectively. The solvent accessibility of the PhoQ-KD is indicated for each residue by an open circle if the fraction solvent accessibility is >0.4, a half-filled circle if it is 0.1–0.4, and a filled circle if it is <0.1. Residues that interact with AMPPNP-Mg²⁺ in PhoQ-KD are indicated by triangles. Blue triangles represent interactions with the adenine ring, red with the phosphates, and green with the Mg²⁺ ion. Classical N, F, G1, G2, and the recently defined G3 boxes, as well as ATP lid are indicated as labeled. B, ribbon diagrams of the structures of the sequences aligned in A. From left to right, PhoQ-KD, EnvZ-KD, and CheA-KD. AMPPNP molecules are shown in ball-and-stick representation for PhoQ-KD and EnvZ-KD. Orientation and structural elements are colored as in Fig. 1A with the exception of the ATP-lid, which is shown here in purple.

Fig. 3. ATP-lid movement in PhoQ-KD nucleotide-bound structure. Superimposition of Ca traces from PhoQ-KD (blue) and CheA-KD (yellow) shows the ATP-lid displacement toward the main β-sheet in PhoQ (closed conformation). The AMPPNP molecule is shown in magenta. Glycine Ca atoms are shown as spheres, and side chains of the hydrophobic patch residues are shown as sticks. Selected residues are labeled for each protein in the color corresponding to the backbone trace. β hairpin movement and the conserved Gly in the hairpin are also indicated.

elements mediate additional hydrogen bonds between the protein and the adenine nitrogen atoms (Fig. 4). Specifically, the N6 amine group interacts with the main-chain carbonyl group of Val-386 through water W3. The endocyclic N1 atom of the adenine ring makes a bidentate water-mediated hydrogen bond to the main chain nitrogen atom of conserved Gly-419 and the side chain of Asp-415. Additionally, the adenine N7 atom forms a water-mediated hydrogen bond with the side chain oxygen of conserved Asn-389, which makes an additional water-mediated interaction with the adenine N1 atom. Additional elements responsible for adenine base binding are Tyr-393, which makes an aromatic stacking interaction on one side, and Ile-420, which makes van der Waals contacts with the other face of the adenine ring (Fig. 4). Tyr-393 is held in place by hydrophobic interactions between the aromatic ring and the aliphatic portion of the Lys-392 side chain, which lies parallel to the tyrosine residue (Fig. 4). The ribose moiety of the AMPPNP molecule presents weak interactions with the protein. Its 2’- and 3’-hydroxyl groups are more solvent-exposed than the rest of the sugar moiety with the O3’ atom forming a hydrogen bond with the hydroxyl group of Tyr-393 (Fig. 4).

Although residues in the two conserved boxes (N and G2) interact with the triphosphate moiety of AMPPNP, many of these contacts are provided by amino acids that are only partially conserved among histidine kinases. The γ-phosphate is hydrogen-bonded to the side chains of Gln-442 and Arg-439 of the ATP-lid and with Lys-392 and Tyr-393 of the N box (Figs. 2A and 4). The side chain of Arg-434 forms a hydrogen bond with the β-phosphate and the α-phosphate interacts with the side chains of Asn-385 and Asn-389, for which the N box is named, and with the peptide nitrogen atom of Leu-446 in the G2 box (Figs. 2A and 4). Of these residues, only Asn-389 is absolutely conserved among histidine kinases, whereas the others are only partially conserved. Apparently, the molecular details of the nucleotide-binding site vary for different histidine kinases. The divalent cation is coordinated by the three phosphate groups of the nucleotide via three non-bridging oxygen atoms (Fig. 4). The remaining three octahedral coordination sites of the cation are occupied by the carboxamide oxygens of Asn-385 and Gln-442, and by a water molecule. Surprisingly, the coordination distances between the metal cation and its ligands (about 2.45 Å) are more appropriate for a bound Mn²⁺ than a Mg²⁺ ion (27). The high concentration of Mg²⁺ present in crystallization solution (~150 mM as acetate salt) and the fact that Mg²⁺ is the most common divalent cation required for enzymatic activity by histidine kinases make it most likely the case that Mg²⁺ is the ion present in the structure. However,
Mn$^{2+}$ has been reported as a preferred cation in some bacterial (28) and plant (29) histidine kinases. Future enzymatic assays will be necessary to elucidate the cation preferences of PhoQ.

The nucleotide has a compact conformation, with a C3'-endo sugar pucker and the γ-torsion angle is in the sc conformation (as defined in Ref. 30). Although this “closed” conformation is less common than the “extended” conformation (30), it is present in the structurally related MutL, GyrB, and Hsp90 protein-nucleotide complexes (21, 22, 31). The AMPPNP triphosphate moiety is “curled” due to its α, β, γ-tridentate coordination to Mg$^{2+}$, which brings the γ-phosphate near the ribose ring. The α, β, γ-tridentate Mg$^{2+}$ coordination is also unusual in a metal-nucleotide-protein complex, but is also present in the MutL and GyrB structures. Additional protein structures showing this tridentate coordination are phosphoglycerate kinase (32), pyruvate kinase (33), cyclin-dependent kinase 2 (34), chaperonin GroEL (35), and a protein of unknown function from Methanococcus jannaschii (36). In PhoQ-KD, this closed and curled conformation places the γ-phosphate facing inward toward the α2 helix, a position where phosphate transfer would be obstructed (Fig. 4). Consequently, a γ-phosphate movement (possibly a rotation) must occur to allow histidine phosphorylation (see below).

**Kinetic Analysis of Mutant Variants**—Until now, under-
standing of the catalytic mechanism for the autokinase reaction of histidine kinases has been hindered by the lack of highly conserved residues that might have an enzymatic role and by a lack of structural definition. Disorder in the binding site of the EnvZ catalytic domain structure precluded such analysis. Although parts of the ATP-lid also are flexible in this PhoQ complex, the side chains from three residues that interact directly with Mg\(^{2+}\)/H\(_{11001}\) nucleotide are the best defined parts of this flexible segment (Fig. 1C). The side-chain B-factor values of these residues (31.2, 34.3, and 33.1 Å\(^2\)) for Arg-434, Arg-439, and Gln-442, respectively, are comparable with the average for the side chains overall (25.8 Å\(^2\)). Therefore, the detail observed in the nucleotide binding site of PhoQ allows us to identify potential catalytic residues. As shown in Fig. 4, three basic residues are placed at the phosphotransfer site; Lys-382 and Arg-439 interact at the \(\gamma\)-phosphate side, and Arg-434 interacts with the \(\beta\)-phosphate group. To examine how these groups may contribute to cleavage and/or to transition state stabilization, each residue was individually substituted by alanine in the context of the entire cytoplasmic fragment (including the histidine substrate-containing phosphotransfer domain) and the kinetic parameters were measured. Results of the mutational analysis are shown in Table II.

An alanine substitution for Lys-392 produces a strong effect in \(K_m\) (roughly 40-fold increase) and a weaker effect in \(k_{cat}\) (about 10-fold reduction), which suggests roles in both nucleotide binding and catalysis. Binding may come both from hydrophobic interactions between the aliphatic portion of the lysine side chain and Tyr-393, which appears to buttress the aromatic residue for optimal stacking with the adenine base, and possibly also from amino group interactions with the \(\gamma\)-phosphate group. Electrostatic neutralization of the transition state provided by the e-a-
TABLE II

| Protein  | $K_m$ (μM) | $k_{cat}$ (min$^{-1}$) | $k_{cat}/K_m$ (s$^{-1}$) |
|----------|------------|------------------------|------------------------|
| Wild type| 20.93      | 1.70 $\times$ 10$^{-3}$ | 81.9                   |
| Lys$^{925}$ → Ala 916.7 (40.49) | 2.93 $\times$ 10$^{-3}$ | 0.32                   |
| Arg$^{434}$ → Ala 40.37 (10.06) | 3.30 $\times$ 10$^{-3}$ | 0.82                   |
| Arg$^{439}$ → Ala 58.63 (28.19) | 3.13 $\times$ 10$^{-3}$ | 53.4                   |

Nucleotide-bound PhoQ Kinase Domain Structure

Autokinase assays were performed at 22 °C in 20 mM Tris-Cl (pH 8.0), + 50 mM KCl, + 1 mM MgCl$_2$, and the kinetic parameters were derived as described under “Experimental Procedures.” Mean values were calculated from at least three individual experiments.

**Implications**

**Variation of Catalytic Mechanism among Histidine Kinases**—Although histidine kinases have characteristic amino acid sequences, they nevertheless show considerable variability even within their more conserved N, F, and G boxes; 11 distinct subgroups have been defined (39). Having identified residues that are involved in catalysis in PhoQ, we are now able to investigate histidine kinase variability in the context of catalytic mechanism. As described above, Tyr-393 in PhoQ stacks with the adenine base of the bound nucleotide and Lys-392 is proposed to help position the tyrosine for more optimal nucleotide binding and to stabilize the transition state. Arg-434 plays a critical role in catalysis and is less important in binding. EnvZ is identical to PhoQ-AMPPNP complex, the γ-phosphate group faces an orthogonal direction to that found in the MutL structure (Fig. 5B). The orientation of the γ-phosphate in the MutL structure may be representative of the orientation of the γ-phosphate in PhoQ after it undergoes a proposed rotation (see above) to allow exposure to the histidine substrate. An aspartate residue and structural water molecules in all cases recognize the adenine ring. However, a GHL-conserved threonine residue is bound to one of these water molecules in the ATPases. The differences between the PhoQ-KD active site and that of the three ATPases are most striking when it comes to the nature of the residues interacting with metal-triphosphate (Fig. 5). The absolutely conserved catalytic glutamate residue that serves as general base for water activation in ATP hydrolysis in the ATPases (Glu-29 in MutL) is replaced in histidine kinases by a conserved asparagine (Asn-385 in the PhoQ-KD structure, Fig. 5). The equivalent of Asn-389 in PhoQ is absolutely conserved in all members of the GHKL superfamily but interacts with the divalent cation in the ATPase structures. Surprisingly, Asn-389 does not coordinate the Mg$^{2+}$ cation in PhoQ-KD; instead, Asn-385 and Glu-442 are the two protein residues that interact with Mg$^{2+}$ in the PhoQ-KD structure. The space occupied by Glu-442 in PhoQ is taken up by a conserved lysine residue in GHKL family (Lys-307 in MutL, Fig. 5) that comes from a separate domain contiguous to the nucleotide-binding domain and forms hydrogen bonds with the γ-phosphate (Fig. 5B). This lysine residue has been implicated in playing a key role in transition state stabilization in MutL and GyrB (22, 38). Neither the cytoplasmic fragment nor the isolated catalytic domain of PhoQ possess detectable ATPase activity (data not shown). It would be interesting to see if a basic substitution at residue 442 in conjunction with a glutamate residue in position 385 would confer ATPase activity on PhoQ.

**Structural Similarities with GHL ATPases and Mechanistic Implications**—The kinase domain of the histidine kinases belongs to a larger family of proteins, termed the GHKL family, whose members include GyrB, Hsp90, histidine kinases, and MutL (5). Unlike the histidine kinases, which catalyze the phosphotransfer from ATP to a histidine substrate, the other members of the family are ATPases that use the energy of hydrolysis to mediate the movement of protein subunits that perform various functions. A structurally based sequence alignment of PhoQ-KD, MutL, DNA gyrase B, and Hsp90 is shown in Fig. 5A. PhoQ-KD presents the four conserved motifs characteristic of the superfamily (5), three of which correspond to the classical N, G1, and G2 histidine kinase conserved motifs (Fig. 2A). Topologically, the five β strands and the three α helices that define the structural core of PhoQ-KD superimpose with homologous elements in the other members, although the N-terminal α1 helix and βB strand elements are in the C-terminal sequence of MutL, GyrB, and Hsp90 rather than the N-terminal sequence as in PhoQ-KD (Fig. 5A). Pairwise sequence comparison between PhoQ-KD and the other three proteins gives identities between 14 and 23% (GyrB) for these alignments, which at the high end approaches that observed in comparisons among histidine kinase family members.

Fig. 5B shows an overlay of the PhoQ-KD and MutL structures with a superimposition of the nucleotide molecules. The nucleotides have similar conformations except at the γ-phosphate. In the PhoQ-KD-AMPPNP complex, the γ-phosphate group faces an orthogonal direction to that found in the MutL structure (Fig. 5B). The orientation of the γ-phosphate in the MutL structure may be representative of the orientation of the γ-phosphate in PhoQ after it undergoes a proposed rotation (see above) to allow exposure to the histidine substrate. An aspartate residue and structural water molecules in all cases recognize the adenine ring. However, a GHL-conserved threonine residue is bound to one of these water molecules in the ATPases. The differences between the PhoQ-KD active site and that of the three ATPases are most striking when it comes to the nature of the residues interacting with metal-triphosphate (Fig. 5). The absolutely conserved catalytic glutamate residue that serves as general base for water activation in ATP hydrolysis in the ATPases (Glu-29 in MutL) is replaced in histidine kinases by a conserved asparagine (Asn-385 in the PhoQ-KD structure, Fig. 5). The equivalent of Asn-389 in PhoQ is absolutely conserved in all members of the GHKL superfamily but interacts with the divalent cation in the ATPase structures. Surprisingly, Asn-389 does not coordinate the Mg$^{2+}$ cation in PhoQ-KD; instead, Asn-385 and Glu-442 are the two protein residues that interact with Mg$^{2+}$ in the PhoQ-KD structure. The space occupied by Glu-442 in PhoQ is taken up by a conserved lysine residue in GHKL family (Lys-307 in MutL, Fig. 5) that comes from a separate domain contiguous to the nucleotide-binding domain and forms hydrogen bonds with the γ-phosphate (Fig. 5B). This lysine residue has been implicated in playing a key role in transition state stabilization in MutL and GyrB (22, 38). Neither the cytoplasmic fragment nor the isolated catalytic domain of PhoQ possess detectable ATPase activity (data not shown). It would be interesting to see if a basic substitution at residue 442 in conjunction with a glutamate residue in position 385 would confer ATPase activity on PhoQ.

**Variation of Catalytic Mechanism among Histidine Kinases**—Although histidine kinases have characteristic amino acid sequences, they nevertheless show considerable variability even within their more conserved N, F, and G boxes; 11 distinct subgroups have been defined (39). Having identified residues that are involved in catalysis in PhoQ, we are now able to investigate histidine kinase variability in the context of catalytic mechanism. As described above, Tyr-393 in PhoQ stacks with the adenine base of the bound nucleotide and Lys-392 is proposed to help position the tyrosine for more optimal nucleotide binding and to stabilize the transition state. Arg-434 plays a critical role in catalysis and is less important in binding. EnvZ is identical to PhoQ at these positions. Analysis of 467 histidine kinases reveals two major classes with respect to these three catalytic residues.

The predominant class of histidine kinases is like PhoQ in having a basic/aromatic pair at positions 392/393. Fifty-eight percent of the histidine kinases have either lysine or arginine in the position corresponding to Lys-392 in PhoQ, and members of this class nearly always have an aromatic or a histidine residue (86 and 13%, respectively) in place of Tyr-393. Presum-
ably, the members of this class utilize a binding arrangement similar to that observed in the PhoQ-KD structure. Kinases in this class typically either have a basic residue (41%) as in PhoQ or a glutamine (47%) residue at the position corresponding to Arg-434. Clearly, those kinases that lack a basic residue at this position must utilize a different basic residue or mechanism to substitute for the role Arg-434 plays in catalysis in PhoQ. The high frequency with which glutamine is found at this position suggests that it plays an important role in the function of those proteins.

Histidine kinases of a second major class (34% of all) have either aspartate or glutamate (22%) or asparagine or glutamine (12%) in the position corresponding to Lys-392 in PhoQ, and in this class the residue corresponding to Tyr-393 is usually an alanine (45%) or a histidine (30%). This second major class usually has three residues displaying the motif of Thr/Ser (68/25%)–Thr/Gly/Ser (72/13%)–Lys/Arg (66/28%) aligned with residues 434–436 of PhoQ. CheA belongs to this second class (Fig. 2A). The sequence preferences of the second class are strikingly similar to those of the GHL ATPases, which have Asp and Ala/Glu, respectively, at the positions corresponding to Lys-392 and Tyr-393, and a Thr/Gly/Ser/Gly/Thr/Lys triad in place of Arg-434 (Fig. 5A). The lysine residue is in all cases hydrogen-bonded to the nucleotide β-phosphate (Lys-79 in MutL) and in MutL Thr-77 and Ser-78 interact with the ribose moiety with the α-phosphate and Ser-78 (Fig. 5B). Presumably, similar interactions occur in the non-PhoQ-like class of histidine kinases and the binding energy contributed by these interactions may compensate for the lack of the tyrosine-adenine stacking interaction observed in the PhoQ-KD structure.

Orthodox histidine kinase proteins with a typical domain organization, as exemplified by PhoQ and EnvZ, have a transmembrane N-terminal sensor domain linked through the histidine phosphotransfer domain to the kinase domain. Atypical histidine kinase proteins, such as CheA, have the histidine domain remote from the kinase domain and separate from the sensor domain. There also exist hybrid kinases in which an additional pair of phosphoacceptor and phosphodonor sites intervenes, either continguously on the kinase protein or as separate proteins, between the kinase and the response regulator to form a phosphoryl relay system (1). PhoQ happens both to be a typical orthodox kinase and to have both the predominant catalytic configuration, whereas CheA is an atypical kinase having the secondary class of catalytic configuration. These associations are not general, however; members of the different classes of catalytic configuration are distributed variously among the different types of domain organization.

The crystal structure of T. maritima CheA-KD complexed with a variety of ATP analogs was published just before submission of this article (3). Together, the present work and the results of Bilwes et al. provide both corroboratory and complementary views of the histidine kinase active site. The repositioning of the ATP-lid to a closed conformation in response to nucleotide binding seen in the case of CheA-KD is similar to that deduced here from the PhoQ-KD structure. There do appear to be significant differences in the geometry of nucleotide binding, however, as expected from the analysis above which places PhoQ and CheA in different catalytic classes. A detailed comparison between these structures when the molecular coordinates of CheA-KD become available in the PDB will likely provide important insights into the catalytic mechanisms utilized by these two proteins. Mutational studies to evaluate the roles of candidate catalytic residues in CheA will also be of interest.
Structural and Mutational Analysis of the PhoQ Histidine Kinase Catalytic Domain: INSIGHT INTO THE REACTION MECHANISM
Alberto Marina, Christina Mott, Anna Auyzenberg, Wayne A. Hendrickson and Carey D. Waldburger

*J. Biol. Chem.* 2001, 276:41182-41190.
doi: 10.1074/jbc.M106080200 originally published online August 7, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M106080200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 38 references, 7 of which can be accessed free at http://www.jbc.org/content/276/44/41182.full.html#ref-list-1