The Structure of an Ancient Conserved Domain Establishes a Structural Basis for Stable Histidine Phosphorylation and Identifies a New Family of Adenosine-specific Kinases

J. Shaun Lott, Blake Paget, Jodie M. Johnston, Louis T. J. Delbaere, Jill A. Sigrell-Simon, Mark J. Banfield, and Edward N. Baker

From the Laboratory of Structural Biology, Centre for Molecular Biodiscovery and School of Biological Sciences, University of Auckland, 3A Symonds St., Private Bag 92-019, Auckland, New Zealand, the Department of Biochemistry, University of Saskatchewan, Saskatoon S7N 5E5, Canada, the Department of Research and Development, GE Healthcare, Björkagan 30, SE-751 84 Uppsala, Sweden, and the Institute for Cell and Molecular Biosciences, University of Newcastle upon Tyne, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom

Phosphorylation of both small molecules and proteins plays a central role in many biological processes. In proteins, phosphorylation most commonly targets the oxygen atoms of Ser, Thr, and Tyr. In contrast, stably phosphorylated His residues are rarely found, due to the lability of the N–P bond, and histidine phosphorylation features most often in transient processes. Here we present the crystal structure of a protein of previously unknown function, which proves to contain a stably phosphorylated histidine residue. The protein is the product of open reading frame 2307, from the hyperthermophilic archaeon Pyrobaculum aerophilum, and is representative of a highly conserved protein family found in archaea and bacteria. The crystal structure of PAE2307, solved at 1.45-Å resolution \( (R = 0.208, \text{R}_{\text{free}} = 0.227) \), forms a remarkably tightly associated hexamer. The phosphorylated histidine at the proposed active site, pHis \(^{85}\), occupies a cavity that is at the interface between two subunits and contains a number of fully conserved residues. Stable phosphorylation is attributed to favorable hydrogen bonding of the phosphoryl group and a salt bridge with pHis \(^{85}\) that provides electronic stabilization. In silico modeling suggested that the protein may function as an adenosine kinase, a conclusion that is supported by in vitro assays of adenosine binding, using fluorescence spectroscopy, and crystallographic visualization of an adenosine complex of PAE2307 at 2.25-Å resolution.

Phosphorylation of appropriate functional groups provides one of the key mechanistic devices in biology. Thus, phosphoryl transfer reactions involving molecules such as ATP and GTP, and the phosphorylation state of small molecule substrates, are critical to many enzyme reactions. Likewise, the phosphorylation and dephosphorylation of amino acid side chains in proteins forms the basis of many signaling and regulatory processes. The most common targets are the -OH groups of amino acid side chains such as serine, threonine, and tyrosine and small molecules such as sugars. In part this reflects the ubiquitous occurrence of such groups, but it also depends on the relative stability of the O–P bond to hydrolysis, which gives long-lived phosphorylation.

In contrast, the phosphorylation of nitrogen atoms, with the formation of N–P bonds, is not often observed (1). In proteins, the phosphorylation of histidine residues produces a phosphoramidate bond that has a large standard free energy of hydrolysis. This makes phosphohistidines the most unstable of any known phosphoamino acid (2) and favors the utilization of histidine in rapid processes involving transient phosphorylation. Examples include the use of phosphohistidines as enzyme intermediates, as in the mechanisms of succinyl-CoA synthetase (3) and nucleoside diphosphate kinase (4), or for rapid signaling processes such as the two-component systems of bacteria (5, 6). In the latter, a sensor is connected to a regulator through histidine phosphorylation and a subsequent phospho-transfer to an aspartate residue.

Another well characterized system, the bacterial phosphoenolpyruvate:sugar phosphotransferase system (PTS), consists of proteins that carry out four successive phosphoryl transfers until the final phosphorylation of the sugar, concomitant with transport into the cell across the bacterial cell membrane (7). In the PTS, Enzyme I (EI) first autophosphorylates, using phosphoenolpyruvate as a substrate, and then transfers the phosphoryl group to the histidine-containing protein. Subsequently the phosphoryl group is transferred to the sugar-specific Enzymes II (EIIA and EIIB), which can consist of multisubunit enzymes or separate proteins. Each of four proteins, EI, histidine-containing protein, EIIA, and EIIB, is transiently phosphorylated at a histidine residue, except certain EIIBs.

The abbreviations used are: PTS, phosphoenolpyruvate:sugar phosphotransferase system; PAE2307, protein product of open reading frame 2307 from the P. aerophilum genome; ORF, open reading frame; pHis, phosphorylated histidine; EI and EII, Enzymes I and II; r.m.s., root mean square.
which instead can be phosphorylated at a cysteine residue. During the phosphotransfer in the PTS, the phosphoryl group, commencing with the Ne2 atom on EI, alternates between being bound on the Ne2 atom of a histidine residue of one component and the Ne1 atom of a histidine of the next component (8).

The characteristic instability of phosphorylated histidines means that only a few examples have been defined structurally. The structure of the phosphohistidine form of histidine-containing protein has been determined by NMR spectroscopy (9), but none of the phosphohistidine forms of the PTS proteins have been elucidated by x-ray crystallography. The structures of phosphohistidine intermediates have been determined crystallographically for three enzymes, succinyl-CoA synthetase (3), nucleoside diphosphate kinase (4), and a cofactor-dependent phosphoglycerate mutase (10). From these, it is clear that phosphohistidine residues can be stabilized in certain environments. Indeed, in histone H4, a phosphohistidine (at residue 75) has been shown to have a half-life of 12 days at room temperature and pH 7.6 (11).

Here we describe the crystal structure of a functionally uncharacterized protein, PAE2307, which has been discovered to contain a stably phosphorylated histidine residue. This protein was selected as a target in a pilot structural genomics enterprise, focused on open reading frames (ORFs) from the hyperthermophilic archaeon Pyrobaculum aerophilum (12), which were annotated as unknowns and for which no functional or structural predictions could be made based on amino acid sequence. PAE2307 is a representative of a conserved family of proteins found in both archaeal and bacterial species, which has been described as an “ancient conserved domain” (13) and presumably has an evolutionarily distant origin, prior to the divergence of the bacteria and the archaea. The high level of conservation in this family suggests some important but previously uncharacterized biological function (14, 15). The discovery of a phosphorylated histidine strongly suggests that PAE2307 is involved in a phosphotransfer reaction common to both bacteria and archaea and enables us to analyze the structural features that contribute to stable histidine phosphorylation. Additionally, bioinformatic evidence coupled with weak structural similarity to a thermostable DNA polymerase suggests that nucleotides or nucleosides are the most likely substrate for the phosphorylation reaction thought to be catalyzed by PAE2307. In silico modeling identified a binding site for adenosine that places the C5’-hydroxyl group of ribose adjacent to the phosphorylated histidine, suggesting that the protein may function as a nucleoside kinase. This model is supported by in vitro measurements using fluorescence spectroscopy and the visualization by x-ray crystallography of the adenosine-bound form of the protein.

**EXPERIMENTAL PROCEDURES**

**ORF Selection**—ORFs that were annotated as “unknown” in the preliminary annotation of the *P. aerophilum* genome (12) were selected as an initial target set. These ORFs were then screened for putative trans-membrane helices using DAS (16) and TMHMM (17). PSI-BLAST (18) searches using the ORFs that lacked predicted transmembrane helices were manually inspected, and those with any significant similarity to characterized genes were removed. Finally, the remaining ORFs were passed through a fold prediction algorithm (19), and those that gave a Z-score of >5 were removed. The remaining ORFs were considered ‘true’ unknowns and were selected for further study.

**Protein Expression, Purification, and Crystallization**—PAE2307 was overexpressed in *Escherichia coli*, using the expression vector pET28, and was purified by immobilized metal ion affinity chromatography and size-exclusion chromatography as described (20). The expression vector adds an N-terminal polyhistidine tag, which was used for purification, but was not cleaved, and therefore remained on the protein during subsequent studies. Native PAE2307 crystals were initially obtained in two crystal forms, one of which was suitable for x-ray analysis (20). This crystal form was grown from an unbuffered solution of 50 mM KH₂PO₄ containing 20% polyethylene glycol 8000, and proved to be tetragonal, space group I4,22, with unit cell dimensions a = b = 120.0, c = 156.5 Å and three copies of the protein monomer in the asymmetric unit. The crystals diffracted to 1.45-Å resolution, and they were used for the native structure determination but were not reproducible. For subsequent ligand soaking experiments, a third crystal form (Type III) was obtained by mixing protein solution (6.8 mg/ml in 20 mM HEPES, 150 mM NaCl, pH 8.0) in a 1:1 ratio with mother liquor (21% (w/v) methoxypolyethylene glycol 5000, 0.2 M N-(1,1-dimethyl-2-hydroxyethyl)-3-amino-2-hydroxypropanesulfonic acid/KOH, pH 8.3) and incubating in a sitting-drop experiment at 18 °C. Thick, plate-like crystals appeared after 3 days and reached their maximum size after 14 days. These crystals proved to be orthorhombic, space group P₂₁2₁2₂, a = 76.9, b = 109.8, c = 112.5 Å, with six molecules in the asymmetric unit.

**Data Collection, Structure Determination, and Refinement**—The structure of native apoPAE2307 was solved by single-wavelength anomalous diffraction phasing from a platinum derivative, as described previously (20). Manual building using O (21) into a map calculated from experimental phases to 2.1 Å yielded a partial model, which was refined against a 1.45-Å resolution native data set, collected at 110 K at the Stanford Synchrotron Radiation Laboratory (beamline 9-2, Area Detector Systems Corp. Quantum 4 detector, wavelength 0.9792 Å). The data statistics are in Table 1. A single round of simulated annealing using CNS (22) reduced the *R*<sub>free</sub> from 53.3% to 43.4%, and subsequent automatic model building using ARP/wARP (23) resulted in an almost complete amino acid chain for all protomers. Two further rounds of model building using O, interspersed with simulated annealing, energy minimization, and individual atomic B-factor refinement using CNS, resulted in a final model comprising residues 5–167 of monomers A and B, and residues 6–167 of monomer C, with *R*<sub>crys</sub> = 20.8% and *R*<sub>free</sub> = 22.1%. No density was found for the N-terminal His tag on any of the three molecules in the asymmetric unit. Full refinement statistics are shown in Table 1. The Ramachandran plot produced by PROCHECK (24) showed that 90.8% of all residues fall in the most favored regions, and only two residues per monomer, Ala<sup>166</sup> and Phe<sup>28</sup>, are in disallowed regions; both are adjacent to the active site, and both have unambiguous electron density.
The complex with adenosine was obtained by soaking a Type III native crystal in a series of cryoprotectant solutions, which consisted of mother liquor containing 1%, 5%, 10%, 15%, and 20% glycerol, and which were all saturated with adenosine. Adenosine was poorly soluble in the mother liquor, but its concentration was estimated to be at least 250 μM. The crystal was then flash-cooled in liquid nitrogen for data collection. Data were collected at 110 K at the European Synchrotron Radiation Facility, Grenoble (beamline ID-29, wavelength 0.9793 Å). The data were processed with MOSFLM and SCALA from the CCP4 program suite (25), giving a full data set to 2.25 Å (Table 1). The structure of the adenosine complex was solved by molecular replacement with MOLREP (26), using the 1.45-Å native structure as search model. Refinement was with REFMAC5 (27), with manual rebuilding achieved by using COOT (28). Clear electron density for a bound adenosine molecule was found in the putative active sites of each of the six molecules in the asymmetric unit; these were modeled in and refined with full occupancy. The final structure comprises residues 5–167 of molecules A, B, D, and E, and 6–167 of molecules C and F; six adenosine molecules; two phosphate ions; and 407 water molecules. Final values of \( R_{\text{cryst}} \) and \( R_{\text{free}} \) were 18.2% and 24.1%, respectively (see Table 1 for full details).

**Crystallographic Verification of His<sup>85</sup> Phosphorylation in ApoPAE2307**—The first electron density maps of the I4<sub>22</sub> crystal form strongly suggested that His<sup>85</sup> was phosphorylated in each of the three protomers. The initial refinement, in which this residue was modeled as an unmodified histidine, returned B-factors for His<sup>85</sup> and the associated Asp<sup>139</sup> that were consistent and similar: 10–14 Å<sup>2</sup> and 13–15 Å<sup>2</sup>, respectively, for protomer A, 12–15 Å<sup>2</sup> and 11–13 Å<sup>2</sup> for protomer B, and 11–13 Å<sup>2</sup> and 14–15 Å<sup>2</sup> for protomer C. Difference maps after this refinement showed unequivocal electron density for phosphoryl groups attached to each histidine, at levels of 4e, 2e, and 4e, respectively, for monomers A, B, and C, and with P::Ne2 (His) distances of 1.7–1.8 Å. Phosphoryl groups were then added to each His<sup>85</sup> side chain, at full occupancy, and a further round of refinement was carried out. The B-factors for the three phosphoryl groups were all in the range 20–23 Å<sup>2</sup>, slightly higher than the values for the non-hydrogen atoms of the side chains, which remained at 12–17 Å<sup>2</sup>. However, difference maps showed some negative density around the phosphorus atoms, so a round of grouped occupancy refinement was performed for the histidine phosphate groups, which lowered the occupancy of the phosphate groups to 0.74 for protomer A, 0.60 for protomer B, and 0.67 for protomer C. These values were retained in further rounds of refinement.

**In Silico Docking Experiments**—The structures of potential ligands at the highest resolution available were obtained from the HIC-Up data base (29) and were energy-minimized and converted to MOL2 format using the PRODRG server (30). *In silico* ligand binding was performed using the program GOLD (31).

**Fluorescence Spectroscopy**—Spectroscopic experiments were performed on an Hitachi F-4500 fluorescence spectrophotometer, with excitation and emission slits widths of 5 nm, and a scan speed of 240 nm/min. Excitation and emission wavelengths of 280 nm and 347 nm, respectively, were used for all samples, using a quartz cuvette with 0.5-cm path length. Analysis by mass spectrometry gave a mass of 20,674 Da, compared with 20,675 Da expected for the full-length His-tagged protein minus its N-terminal Met residue. This showed that the protein used in these assays was homogeneous and not phosphorylated. Protein (5 μM) and ligands (0.5–500 μM) were each dissolved in a buffer solution containing 20 mM HEPES and 150 mM NaCl at pH 8.0. The data were corrected for inner filter effects using a correction factor as described by Eftink (32). The true protein fluorescence intensity was determined by subtracting the buffer fluorescence intensity from the sample fluorescence intensity, and multiplying by the correction factor (C) for that particular sample, where \( C = 10^{((\Delta A_{\text{em}})/2 + (\Delta A_{\text{em,0}}))} \), \( \Delta A_{\text{em}} \) is the absorbance change due to substrate at the excitation wavelength, and \( \Delta A_{\text{em,0}} \) is the absorbance change due to substrate at the emission wavelength. Ligand binding data were fitted using SigmaPlot v9.01 (Systat Software Inc.) using a one-site saturation model: \( \Delta F = ([\text{Ligand}]/(K_d + [\text{Ligand}])) \), where \( \Delta F \) is the normal-
ized change in fluorescence compared with a solution of 5 μM protein alone, ΔF_{max} is the maximum normalized change in fluorescence at saturation, K_d is the dissociation constant, and [Ligand] is the concentration of ligand.

RESULTS

Protein Sequence Analysis—A PSI-BLAST query of the NCBI non-redundant protein sequence data base (release date: December 15, 2005), with the PAE2307 sequence, shows convergence after three iterations (using an E-value cutoff of 0.05) to produce a set of protein sequences derived from 42 distinct species. The proteins form a discrete family of well conserved sequences, all of which are at least 35% identical to each other in amino acid sequence. The members of this protein family are drawn equally from archaeal species (18 examples) and bacterial species (24 examples). They cover a broad phylogenetic range, including representatives from the crenarchaeota, euryarchaeota, and Proteobacteria, Thermotogae, Actinobacteria, Deinococci, and cyanobacteria. However, the family contains no eukaryotic members. A multiple sequence alignment of a phylogenetically diverse selection of these protein sequences is shown in Fig. 1. The multiple sequence alignment reveals 24 residues that are absolutely conserved, including the phosphorylated histidine and several other residues that surround the putative active site. This protein family is classified in the Pfam and Interpro databases as protein of unknown function DUF355 (Pfam accession number PF04008; Interpro accession number IPR007153) and is annotated with the comment that “The high level of conservation in this family suggests some as yet unknown important biological function.” The genomic context of the PAE2307 gene is uninformative, because it does not sit in an obvious operon and is flanked in the genome by poorly characterized ORFs, which are annotated as a putative resistance protein and putative transcriptional regulator. There is also no clear synteny, as the context of orthologs in other species is variable.

Overall Fold of PAE2307—The PAE2307 monomer (Fig. 2a) is folded into a single domain with an extended C-terminal arm, culminating in an α-helix, residues 148–162, that plays a major role in oligomerization. The monomer folds around a central five-stranded antiparallel β-sheet, comprising (in order) strands β6-β2-β5-β3-β4 (Fig. 2b). On one face of this sheet are
packed two α-helices, α1 (residues 28–42) and α3 (residues 99–105). The other face is covered by a third helix α2 (residues 67–82) and a second, three-stranded, antiparallel β-sheet, comprising strands β1, β7, and β8. The phosphorylated histidine, pHis85, described later, is found at the start of β5, the central strand of the main β-sheet, in a cavity bounded by loops from two adjacent monomers. A striking feature of the structure, possibly related to protein thermostability, given that it is derived from a hyperthermophilic organism, is the large number of charged side chains (7 Asp, 13 Glu, 13 Lys, 10 Arg in 167 residues; 25% overall), which give rise to many salt bridges (Table 2). Eleven intramolecular salt bridges are found, and a
Further eight help to stabilize the quaternary structure (see below). Also notable are two internal cysteine residues that are close enough to form a disulfide bond (inter-sulfur distance = 3.8 Å) but in the present structure do not do so.

Quaternary Structure—The protein monomers are arranged in a tightly associated hexamer (Fig. 2c), best described as a dimer of trimers. Dynamic light scattering analysis showed that PAE2307 has a particle size of 123 ± 10 kDa in solution, consistent with a hexameric species. The trimer is formed by the three monomers of the crystal asymmetric unit, with the hexamer being completed by 2-fold crystallographic symmetry. At the center of the trimer, the outer strands (β8; residues 131–140) of the small β-sheet from each of the three monomers face each other around the 3-fold axis. Taking molecule A as the reference, residues 131–135 run antiparallel to 136–140 of molecule B and residues 136–140 run antiparallel to 131–135 of molecule C. Many hydrophobic side chains pack between these three small β-sheets to give stability to the trimer.

Residues 139–144 of molecule A continue on to make extensive inter-subunit contacts, passing through a cleft made by the β3-β4 and the β7-β8 loops from subunit C, with a salt bridge (Arg59–Glu130) completing the encirclement of the peptide as it passes between them. Residues 144–147 from molecule A then run parallel to residues 59–63 of molecule C, making a number of hydrogen bonds, both main chain and side chain; Arg62 and Glu147, both invariant in all the sequences in Fig. 1, form a salt bridge and make other hydrogen bonds that contribute to this very intimate subunit association.

Finally, the C-terminal helix has a striking amino acid composition with no fewer than 11 charged residues out of 16, and extends as an arm to make extensive interactions with the adjacent monomer in the trimer (A with C, C with B, and B with A). Three arginine residues on the inner face of the helix, Arg152, Arg155, and Arg160, play a central role, making three salt bridges, and additional hydrogen bonds, with residues from β1 and β4 of the adjacent subunit, and three intrahelix salt bridges. Three further salt bridges involving Glu119 and Lys166 from this C-terminal arm link it with the β1-β2 loop of the adjacent subunit. These intermolecular salt bridges are listed in Table 1 and clearly play a large part in the trimer stability.

A phosphate ion is found within the narrow tunnel that runs along the 3-fold axis of the trimer, bound by the Arg111 side chain of each protomer in the trimer. The P–O2 bond lies along the 3-fold axis such that O2 interacts with the NH2 groups of three Arg111 side chains. In addition, Ne of each Arg111 side chain in the trimer interacts with one of the other phosphate oxygens.

The trimer can be described as a domed assembly with a rather flat surface as its base. The flat face is formed by the three sets of helices α1 and α3 that back against the central β-sheets of the three monomers; the hexamer then forms by the packing of these two flat faces against each other. Overall, each monomer buries 1830 Å² of surface area (18.6% of its total) in contact with one adjacent monomer in the trimer, and 1910 Å² of surface area (19.4% of its total) in contact with the other adjacent monomer. The total of 3740 Å² of surface area in each monomer is buried in intra-trimer subunit contacts, from a total surface area of 9838 Å² per monomer, indicates an extremely tightly associated trimer. A further 1015 Å² of surface area per monomer is buried in inter-trimer subunit contacts, which have a majority (60%) hydrophobic character. The center of the hexamer is an open, water-filled chamber, with a noticeably positively charged surface.

Fold Similarity—Searching the Protein Data Bank with the PAE2307 monomer using DALI (33) or SSM (34) gives two obvious matches: the conserved hypothetical proteins TT1634 from Thermus thermophilus (PDB code: 1VGG) and TA1353 from Thermoplasma acidophilum (PDB code: 1RLH), which are both orthologs of PAE2307. The structure of the T. thermophilus protein is extremely similar to PAE2307. It is also a hexamer, and the two structures superimpose with a root-mean-square (r.m.s.) difference in Cα atomic positions of only 0.78 Å. However, its active site histidine is not phosphorylated. The T. acidophilum homologue lacks the C-terminal α4 helix, and the crystal structure contains only a monomer in the asymmetric unit, but forms trimers of the same sort as PAE2307 via 3-fold crystallographic symmetry. The T. acidophilum protein and PAE2307 superimpose with an r.m.s. difference in Cα positions of 1.24 Å.

Besides these obvious cases, the structural similarity to known protein structures detected by fold comparison is limited to a partial match of the central β-sheet and the α1 and α3 helices. Many proteins display a topologically similar packing of two helices against a 3- or 4-stranded antiparallel sheet (a ferredoxin-like folding arrangement) in a variety of different contexts, but the packing of helices on either side of a five-stranded antiparallel sheet appears to be novel. The most similar such arrangement is found in the N-terminal domain of a B-type DNA polymerase from Thermococcus gorgonarius (LTGO (35)). Interestingly, this domain has an adjacent three-stranded antiparallel β-sheet, albeit with different topological connections, which is involved in inter-domain interactions within the polymerase (Fig. 3). The inference that can be drawn from this similarity is limited as the functional importance of this domain for the polymerase is unknown.

Active Site—The phosphorylated histidine residue, pHs85, is located in a cavity at the interface between two subunits. This is the presumed active site, giving 6 active sites per hexamer. The cavity is formed by residues 26–30 and 53–58 of one subunit (the β2-α1 and β3-β4 loops, respectively) and residues 115–119 (the β6-β7 loop) and 138–142 of the adjacent subunit. At the center of this cavity, His85 is phosphorylated on its Ne2...
atom, with its N61 atom forming a salt bridge with Asp159 from the adjacent subunit. One end of the cavity is distinctly hydrophobic, bounded by Ile97 at the end, Phe28 and Trp95 with their aromatic rings 7.5 Å apart forming opposing sides, and Tyr165 over the top. Also flanking the cavity are Asn20 and His27, which could play a role in substrate binding.

Each of the oxygen atoms of the phosphoryl group on pHis85 is hydrogen-bonded, with good geometry. The O2P atom is at the N terminus of a 3_10 helix with a hydrogen bond to the peptide NH of Phe28, O1P hydrogen bonds to the peptide NH of Ala55, and O3P bonds to Oy of Ser56 and N82 of Asn118 of the adjacent subunit. The stabilizing environment of the phosphoryl group is illustrated in Fig. 4.

Many of the groups in and around the putative active site are fully or mostly conserved in all homologous sequences (Fig. 1). The phosphorylated His85, together with Phe28, Phe77, Asn118, Asp139, Gly140, and Tyr165 are all invariant, residue 95 is always aromatic, residue 97 is always either Ile or Val, and Asn20 and His27 are only replaced by similar hydrogen-bonding residues. Between them, these residues account for many of the conserved residues in the sequence alignment in Fig. 1, with the other residues presumably conserved for conformational reasons (Gly and Pro) or because they make multiple stabilizing hydrogen bonds or salt bridges.

**Functional Hypothesis and in Silico Ligand Binding**—The presence of a stably phosphorylated histidine strongly suggested that PAE2307 has a biochemical role that involves phosphate transfer. The putative substrate for this phosphorylation is not clear, but two lines of evidence imply a role in DNA or nucleotide metabolism. First, there is the weak structural similarity to a B-type DNA polymerase described above. Secondly, when analyzed using a phylogenetic profiling technique based on the comparison of orthologs from 81 microbial genomes,5 PAE2307 is found to cluster with DNA-binding proteins from *Methanopyrus kandleri*, *Sulfolobus solfataricus*, and *Thermoplasma acidophilum*, a gyrase from *S. solfataricus*, and adenylate and predicted nucleotide kinases from *M. kandleri*.

To test the feasibility of a role for PAE2307 in nucleotide metabolism, in silico docking studies were carried out with a number of nucleosides and nucleotides, using the program GOLD (31). The phosphate of group of pHis85 was taken as the center of the active site, and a radius of 20 Å was used to define the possible binding surface, which encompassed the whole inter-subunit cleft. Phosphorylated nucleotides bound in a number of configurations, with the binding being primarily dictated by the interaction of their phosphate groups with the side chains of Arg62 and Lys58, and the base ring making few contacts with the protein. However, nucleosides reproducibly bound in the hydrophobic pocket adjacent to the active site, with GOLD fitness scores (~80) comparable to those obtained for tight protein-ligand complexes (31). The predicted binding mode of adenosine (Fig. 5a) had the adenine base stacked between the conserved side chains of Phe28 and Trp95 and hydrogen-bonded to the phenolic oxygen of Tyr165. The
involvement of Tyr\textsuperscript{165} in this site emphasizes the importance of the many interactions that hold the C-terminal helix H\textsubscript{9251} (residues 148–162) in place. The predicted positioning of the ring places the ribose moiety adjacent to the phosphate of pHis\textsuperscript{85}, with the ribose O\textsubscript{5} –C\textsubscript{5} bond gauche to the C\textsubscript{4} –O\textsubscript{4} bond and trans to the C\textsubscript{4} –C\textsubscript{3} bond, correctly positioning the O\textsubscript{5} atom to accept the phosphoryl group from the pHis\textsuperscript{85}. This binding mode suggested that PAE2307 may function as a nucleoside kinase. The nature of the phosphate donor could not be predicted, although it would presumably bind on the opposite side of pHis\textsuperscript{85}, in the other half of the binding cleft, in such a way as to be able to donate a phosphate group.

In Vitro Ligand Binding—To test the hypothesis that PAE2307 may function as a nucleoside kinase, in vitro ligand binding experiments were carried out to establish whether nucleosides were able to bind to the protein. The presence of Trp\textsuperscript{95} in the predicted nucleoside binding site enabled intrinsic protein fluorescence to be used to monitor ligand binding. Addition of various nucleosides to the protein in solution resulted in the reduction of observed fluorescence and allowed the measurement of binding curves, as shown in Fig. 6. Adenosine showed relatively tight binding in comparison to the other nucleosides tested (Fig. 6\textsubscript{a}), with a calculated $K_D$ of 15 \textmu{}M compared with 515, 223, and 302 \textmu{}M for guanosine, thymidine, and deoxyuridine, respectively. Therefore, PAE2307 appears to be specific in binding adenosine over other nucleosides. To test whether the phosphorylation state of the nucleoside was important in binding, the three phosphorylated forms of adenosine, AMP, ADP, and ATP, were tested for their ability to bind (Fig. 6\textsubscript{b}). AMP bound only slightly less tightly than adenosine alone, with a calculated $K_D$ of 26 \textmu{}M. However, the di- and triphosphates bound much more weakly, with $K_D$ values of 740 and 741 \textmu{}M, respectively. Interestingly, adenine base alone bound to the protein more tightly than either ADP or ATP, with an estimated $K_D$ of 198 \textmu{}M. This preference for binding non- or mono-phosphorylated forms of adenosine over the higher
Histidine Phosphorylation in a Novel Adenosine Kinase

phosphorylated forms supports the hypothesis that PAE2307 acts as a kinase, with adenosine and/or AMP as preferred substrate.

Structure of Adenosine Complex—To confirm that adenosine binds as predicted by in silico modeling, the crystal structure of an adenosine complex of PAE2307 was determined and refined at a resolution of 2.25 Å (Table 1). Electron density maps showed unambiguous positive difference peaks at the predicted adenosine binding sites, as shown in Fig. 5b. Adenosine modeled into this difference density refined well, with atomic B-factors that were very similar to surrounding residues (25–30 Å²), and no residual difference electron density, thus validating the predictions made from in silico modeling. A comparison of the predicted and observed conformations of adenosine is shown in Fig. 5c. The binding mode is largely as predicted, but there are several small changes in the active site, the most notable of which is that the side chain of Trp95 has rotated about Cβ–Cy to enable better stacking with the ring of the adenine base. The conformation of the observed ligand is slightly different to that predicted by modeling, with a hydrogen bond made from the N1 rather than the N6 nitrogen of the adenine base to the phenolic oxygen of Tyr165, and a rotation around the N9–C1 bond, which enables hydrogen bonds to be made from the O2' and O3'–hydroxyl groups on the ribose ring to the side chain of Asn220 and the backbone nitrogen of Ala177, respectively. One additional change is that, because this form of the protein is not phosphorylated, a hydrogen bond is able to form between the Ne2 nitrogen of His85 and the O5' of the adenosine ribose. Overall, the in silico modeling had accurately placed the adenosine ligand, with a root-mean-square difference in atomic positions of only 1.85 Å between the predicted and observed conformations.

DISCUSSION

The structure of PAE2307 demonstrates the powerful role that structural biology can play in providing functional hypotheses for proteins of previously unknown function. The unexpected discovery of a phosphorylated histidine in PAE2307, in a solvent-exposed cavity that contains a number of fully or mostly conserved residues, leaves little doubt that this is the active site and that the function of this protein and its homologs is in phosphoryl transfer. Given the rarity of stably phosphorylated histidine residues, it also allows us to examine the factors that stabilize such modifications and how these may be related to function.

There are currently three other examples of crystal structures of phosphohistidine-containing proteins, nucleoside diphosphate kinase (4), succinyl-CoA synthetase (3), and cofactor-dependent phosphoglycerate mutase (10). A common factor in each of these structures is that the non-phosphorylated imidazole nitrogen of the active histidine side chain is hydrogen-bonded to an oxygen atom of the protein. In nucleoside diphosphate kinase, N6 of the histidine is phosphorylated, while Ne2 interacts with a carboxylate oxygen from a glutamate side chain. In succinyl-CoA synthetase, Ne2 is phosphorylated while N61 also interacts with a carboxylate oxygen of a glutamate side chain. In cofactor-dependent phosphoglycerate mutase, Ne2 is phosphorylated while the N61 hydrogen bonds to a peptide carbonyl oxygen. Histidine phosphorylation involves donation of a pair of electrons from the imidazole nitrogen to form the N–P bond and thus generates a positive charge on the imidazole ring. The ion pair interactions in nucleoside diphosphate kinase and succinyl-CoA synthetase, in which the phosphohistidine ring nitrogen interacts with negatively charged carboxylate oxygen, thus likely stabilize the phosphohistidine group. In the cofactor-dependent phosphoglycerate mutase structure, the interaction between the phosphohistidine and the carbonyl oxygen atom is only a hydrogen bond and would have minimal ionic character.

We conclude that histidine phosphorylation is stabilized by appropriate ion pair hydrogen bonds involving the non-phosphorylated imidazole nitrogen. This is consistent with the full occupancy for the phosphoryl groups in nucleoside diphosphate kinase and succinyl-CoA synthetase, compared with the partial occupancy (0.28) in cofactor-dependent phosphoglycer-
ate mutase where there is a hydrogen bond but no ionic inter-
action. In agreement with this, the salt bridge in PAE2307
between pHis$^{85}$ Nδ1 and the carboxylate oxygen of Asp$^{39}$ sta-
tilizes the phosphoryl group on Ne2 and is expressed in an
occupancy of $\sim 0.7$ in each of the three crystallographically
independent protomers.

The in vivo substrate for the phosphorylation reaction pre-
dicted to be catalyzed by PAE2307 is not known, but the phy-
logenetic linkage of orthologous proteins to DNA-binding pro-
teins and nucleotide kinases, and a weak structural similarity
to a domain (of unknown function) from a thermostable DNA
polymerase, gave the first suggestion that a nucleotide or nucle-
oside may be the substrate. The favorable adenosine-binding
mode predicted by in silico ligand binding analyses led to our
functional hypothesis that the in vivo biochemical function of
this highly conserved protein family is as an adenosine-specific
nucleoside kinase. This hypothesis is strongly supported by our
experimental binding studies. The in vitro binding studies dem-
onstrated that adenosine, but not other nucleosides, bound
tightly to the protein ($K_D = 15 \mu M$) and that di- or triphospho-
ylated adenosine nucleotides bind much less well than aden-
osine or AMP. The crystallographic analysis of adenosine bind-
ing confirmed its specific and stereochemically favorable
binding mode, with the base snugly accommodated between
invariant or highly conserved residues and the O5’ oxygen of
the adenosine ribose in a position to receive a phosphate trans-
ferred from the phosphorylated His$^{85}$.

A priori, it is possible to hypothesize that AMP may be the
phosphate donor rather than the substrate of the proposed
kinase activity. If this were the case, AMP would lose a phos-
phate to become adenosine in the course of the reaction. How-
ever, as adenosine binds more tightly than AMP ($K_D$ of 15 $\mu M$
compared with 26 $\mu M$ for AMP), this scenario would lead to a
dead-end adenosine-protein complex. If, on the other hand,
AMP or adenosine is the acceptor substrate and phosphoryla-
tion leads to formation of ADP or ATP, the di- or triphospho-
ylated product would dissociate from the protein, as both have
a much lower binding affinity (30- to 50-fold weaker). We there-
fore propose that AMP or adenosine is the acceptor in a
phosphoryl transfer from another (unknown) donor.

Adenosine kinases have been characterized from a num-
ber of species, with structures known for the human (36) and
Toxoplasma gondii (37) enzymes. These belong to a wider
family of carbohydrate kinases called the ribokinase family
after the archetypal member (38). They share a common fold
comprising a large $\alpha/\beta$ domain and a smaller, more variable,
domain, and have an ordered, associative mechanism, in
which the phosphate is transferred directly from the donor
ATP to the substrate, without formation of a phosphoen-
zyme intermediate (37, 39). A basic group, usually Asp, acti-
vates the hydroxyl of the acceptor molecule.

PAE2307 has a different fold and appears to be represen-
tative of a novel family of nucleoside kinases. Although it is
possible that the active site histidine, His$^{85}$, could act as a
base to remove the C5’ hydroxyl proton, the observed stable
phosphorylation of His$^{85}$ and its favorable orientation for
phosphoryl transfer to the hydroxyl group strongly suggest
that its mechanism involves a phosphohistidine intermedi-
ate. Unlike the enzymes of the ribokinase family, however,
PAE2307 does not undergo any conformational change out-
side of the active site as a result of adenosine binding; the
native and adenosine-bound structures are essentially
identical.

Although the phosphate donor in the reaction catalyzed by
PAE2307 is not yet identified, its likely binding site can be pro-
posed from an examination of the structure. The cavity in the
protein surface in which the active site histidine is found is
‘L’-shaped, with His$^{85}$ sitting at the corner of the ‘L.’ The aden-
osine molecule binds into the short arm of the ‘L,’ and it is
possible that the phosphate donor will bind into the cavity
along the long arm of the ‘L,’ in such a way as to be able to
transfer phosphate to adenosine via His$^{85}$. Several strongly or
absolutely conserved residues line this region, including Glu$^{54}$,
Ser$^{56}$, and Gly$^{57}$ on one side and Ile$^{29}$ and Arg$^{155}$ on the other,
with Glu$^{54}$ and Arg$^{62}$ forming the cavity floor. However,
memories to attempts to model the binding of potential phosphate donor
molecules to this region in silico were not successful.

PAE2307 exhibits a strong preference for adenine-contain-
ing ligands ($K_D$ of 15 $\mu M$ for adenosine compared with 515, 223,
and 302 $\mu M$ for guanosine, thymidine, and deoxyuridine,
respectively). The reason for this specificity is apparent upon
inspection of the adenosine binding site, as the C2 carbon of
the adenine ring packs closely against the side chain of Ile$^{97}$. The C2
carbon of the base ring is modified in guanine by the addition
of an amino group, and in thymine and cytosine by a carboxyl
oxygen, and either of these modifications would cause a clear
steric clash, disfavoring productive binding. Ile$^{97}$ is itself part of
a strongly conserved sequence motif ($^{6}$P(i/L)N(i/V)L$^{158}$) in helix $\alpha3$, which forms the back of the binding site, indicating
that the observed binding specificity is likely to be shared by the
whole protein family.

In summary, the structure of PAE2307 identifies it as a likely
nucleoside kinase, with a strong preference for adenosine over
other nucleosides. The nature of the donor molecule for the
phosphorylation reaction is unknown, but the reaction appears
to proceed via a phosphohistidine intermediate, in contrast
to known nucleoside kinases. The biological purpose of this
phosphorylation reaction is unclear, but the strong conservation
of sequence within the protein family and its wide phylogenetic
distribution among the archaea and bacteria implies that it is
likely to have an important role.

Acknowledgments—X-ray diffraction data from the native crystal
were collected by Peter Haebel and from the adenosine-soaked crystal
by Fasseli Coulibaly and Richard Bunker. Mass spectrometry data
were obtained by Kayal Rajasekaran and Martin Middelitch.

REFERENCES

1. Klump, S., and Krieglstein, J. (2002) Eur. J. Biochem. 269, 1067–1071
2. Matthews, H. R. (1995) Pharmacol. Ther. 67, 323–350
3. Fraser, M. E., James, M. N., Bridger, W. A., and Wolodko, W. T. (2000) J.
Mol. Biol. 299, 1325–1339
4. Morera, S., Chiaudi, M., LeBras, G., Lascu, I., and Janin, J. (1995) Biochem-
istry 34, 11062–11070
5. Alex, L. A., and Simon, M. I. (1994) Trends Genet. 10, 133–138
6. West, A. H., and Stock, A. M. (2001) Trends Biochem. Sci. 26, 369–376
7. Postma, P. W., Lengeler, J. W., and Jacobson, G. R. (1993) Microbiol. Rev.
Histidine Phosphorylation in a Novel Adenosine Kinase

57, 543–594
8. Schauder, S., Nunn, R. S., Lanz, R., Erni, B., and Schirmer, T. (1998) J. Mol. Biol. 276, 591–602
9. van Nuland, N. A., Boelens, R., Scheek, R. M., and Robillard, G. T. (1995) J. Mol. Biol. 246, 180–193
10. Bond, C. S., White, M. F., and Hunter, W. N. (2001) J. Biol. Chem. 276, 3247–3253
11. Kim, Y., Pesis, K. H., and Matthews, H. R. (1995) Biochim. Biophys. Acta 1268, 221–228
12. Fitz-Gibbon, S. T., Ladner, H., Kim, U. J., Stetter, K. O., Simon, M. I., and Miller, J. H. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 984–989
13. Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D., and Koonin, E. V. (2001) Nucleic Acids Res. 29, 22–28
14. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) Nucleic Acids Res. 32, D138–141
15. Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrett, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P., Bucher, P., Copley, R. R., Courcelle, E., Das, U., Durbin, R., Falquet, L., Fleischmann, W., Griffiths-Jones, S., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lopez, R., Letunic, I., Lonsdale, D., Silventoinen, V., Orchard, S. E., Pagni, M., Peyruc, D., Ponting, C. P., Selengut, J. D., Servant, F., Sigrist, C. J., Vaughan, R., and Zdobnov, E. M. (2003) Nucleic Acids Res. 31, 315–318
16. Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. (1997) Protein Eng. 10, 673–676
17. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) J. Mol. Biol. 305, 567–580
18. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
19. Fischer, D., and Eisenberg, D. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 11929–11934
20. Lott, J. S., Banfield, M. J., Sigrell, J., and Baker, E. N. (2003) Acta Crystallogr. Sect. D 59, 2242–2246
21. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) Acta Crystallogr. Sect. A 47, 110–119
22. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewsiki, 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
23. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Nat. Struct. Biol. 6, 458–463
24. Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
25. CCP4 (Collaborative Computational Project No. 4) (1994) Acta Crystallogr. Sect. D 50, 760–763
26. Vagin, A., and Teplyakov, A. (1997) J. Appl. Crystallogr. 30, 1022–1025
27. Murshudov, G. N., Dodson, E. J., and Vagin, A. A. (1997) Acta Crystallogr. Sect. D 53, 240–255
28. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. Sect. D 60, 2126–2132
29. Kleywegt, G. J., and Jones, T. A. (1998) Acta Crystallogr. Sect. D 54, 1119–1131
30. Schuettelkopf, A. W., and van Aalten, D. M. F. (2004) Acta Crystallogr. Sect. D 60, 1355–1363
31. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Proteins 52, 609–623
32. Eftink, M. R. (1997) Methods Enzymol. 278, 221–257
33. Holm, L., and Sander, C. (1995) Trends Biochem. Sci. 20, 478–480
34. Krissinel, E., and Henrick, K. (2004) Acta Crystallogr. Sect. D 60, 2256–2268
35. Hopfner, K.-P., Eichinger, A., Engh, R. A., Huber, R., Laue, F., Ankenbauer, W., and Angerer, B. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 3600–3605
36. Mathews, I. I., Ealick, E., and Ealick, S. E. (1998) Biochemistry 37, 15607–15620
37. Schumacher, M. A., Scott, D. M., Mathews, I. I., Ealick, S. E., Roos, D. S., Ullman, B., and Brennan, R. G. (2000) J. Mol. Biol. 296, 549–567
38. Sigrell, J. A., Cameron, A. D., Jones, T. A., and Mowbray, S. L. (1998) Structure 6, 183–193
39. Dyguda, E., Szczczel, B., and Sokalski, W. A. (2004) Int. J. Mol. Sci. 5, 141–153
40. Gouet, P., Robert, X., and Courcelle, E. (2003) Nucleic Acids Res. 31, 3320–3323