First Structure of a Eukaryotic Phosphohistidine Phosphatase*

Received for publication, August 30, 2006, and in revised form, September 14, 2006

Robert D. Busam, Ann-Gerd Thorsell, Alex Flores, Martin Hammarström, Camilla Persson, and B. Martin Hallberg1

From the Structural Genomics Consortium, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, SE-17177 Stockholm, Sweden

Phosphatases are a diverse group of enzymes that regulate numerous cellular processes. Much of what is known relates to the tyrosine, threonine, and serine phosphatases, whereas the histidine phosphatases have not been studied as much. The structure of phosphohistidine phosphatase (PHPT1), the first identified eukaryotic-protein histidine phosphatase, has been determined to a resolution of 1.9 Å using multiple-wavelength anomalous dispersion methods. This enzyme can dephosphorylate a variety of proteins (e.g. ATP-citrate lyase and the β-subunit of G proteins). A putative active site has been identified by its electrostatic character, ion binding, and conserved protein residues. Histidine 53 is proposed to play a major role in histidine dephosphorylation based on these observations and previous mutational studies. Models of peptide binding are discussed to suggest possible mechanisms for substrate recognition.

Reversible phosphorylation of residues is crucial in a variety of signaling pathways. Most of our understanding regarding these signaling events in eukaryotes comes from tyrosine, serine/threonine kinases, and phosphatases (1). Less well characterized is histidine phosphorylation-dependent signaling in eukaryotes. A little more than thirty years ago, Histone H4, the first vertebrate protein with a phosphorylated histidine residue, was identified (2). Since then there has been a measured increase in knowledge of mammalian histidine kinases (3). Unfortunately, very little information regarding eukaryotic histidine phosphatases has been available during this same period. This nescience is interesting because histidine phosphorylation is quite prevalent in the cell and likely accounts for ~6% of all phosphorylation in eukaryotes (4). Thus far, only one protein (the β-subunit of heterotrimeric G proteins) in vertebrates has been identified as undergoing reversible histidine phosphorylation where both the kinase (NDPK B) and phosphatase (PHPT1)2 are known (for a recent review, see Ref. 5). However, more information regarding histidine phosphatases is slowly beginning to emerge. To date, the only other structure of a histidine phosphatase is Escherichia coli SixA (6). Under certain anaerobic respiratory conditions, SixA is involved in down-regulation of the E. coli ArcB-to-ArcA phosphorelay system. SixA shows structural homology to the well studied family of arginine-histidine-glycine (RHG) phosphatases (6) but no sequence homology to PHPT1.

Mammalian phosphohistidine phosphatase (PHPT1) was first identified and characterized as a 14-kDa protein in 2002 (7, 8). The enzyme can dephosphorylate the phosphohistidine-containing peptide succinyl-Ala-His(P)-Pro-Phe-nitroanilide, E. coli cheA, rabbit ATP-citrate lyase, and the rat β-subunit of G proteins (7–10). PHPT1 has been suggested to be highly involved in neuronal function. Unlike most phosphatases it does not require divalent cations for activity. Individual point mutations of conserved histidine and arginine residues determined that Arg45, His53, and His102 may play a role in the reaction mechanism as a result of eliminated or reduced phosphatase activity when mutated to alanines (11). PHPT1 is expressed in a variety of vertebrates but not in fungi or bacteria. The PHPT1 DNA sequence shows similarity to testes-specific proteins in Drosophila, occlus, janus-a, and janus-b, according to the Conserved Domain Data Base (12). These proteins may be phosphatases, but little information is available on their function. In this work, we report the first crystal structure of a eukaryotic phosphohistidine phosphatase: PHPT1. The structure enables us to define a substrate-binding pocket and to model possible phosphatase-substrate docking modes. We thereby provide the structural basis for further biochemical, biophysical, and genetic studies in the rapidly developing field of eukaryotic phosphohistidine signaling.

EXPERIMENTAL PROCEDURES

Cloning, Protein Expression, and Purification—Residues 5–125 of the human PHPT1 gene (gi: 19353099) were cloned by ligation-independent cloning into a PET-28 based expression vector incorporating a TEV-cleavable N-terminal His-tag fusion (pNIC-Bsa4). After transformation and liquid culture growth using standard methods, recombinant expression of PHPT1 was induced at 291 K by addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside to either Terrific broth for expression or minimal medium for selenomethionine incorporation. PHPT1 was purified using IMAC on a 1 ml HiTrap chelating HP column followed by gel filtration on a Superdex 75 column (columns from GE Healthcare, Uppsala, Sweden).

2 The abbreviations used are: PHPT, phosphohistidine phosphatase; Bistris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; r.m.s.d., root mean square deviation; ACL, ATP-citrate lyase.
Crystallization, Data Collection, and Structure Solution—Initial crystal screening with selenomethionine-derivatized protein using commercial screens gave hits in several salt conditions. After optimization, large rhombohedral crystals could be obtained by mixing protein solution (15 mg/ml) with an equal volume of reservoir (2.0 M ammonium sulfate, 0.1 M Bis-tris propane, pH 5.5). Crystals were swept into a reservoir with 15% butanediol and then dunked in liquid nitrogen. The data were indexed in space group R3 with unit cell parameters 

\[ a = b = 228 \AA, c = 29.4 \AA. \]

Multiple-wavelength anomalous dispersion data (Table 1) were collected around the selenium K-edge at Beamline 14.1 (BESSY II, Berlin, Germany). XDS/XSCALE (14) was used to process the data and put the three data sets on an approximately similar scale. XPAR (Bruker AXS) were used to prepare FA-values for substructure solution in SHELXD (15). The resulting map could be partially auto-traced using ARP/wARP (17). Unfortunately, only one of the four protomers in the asymmetric unit was well ordered and numerous attempts with different refinement protocols failed to give acceptable residuals. Since twinning is common in space group R3, twinned refinement was attempted with different operators but did not improve residuals or the quality of difference maps.

A search for alternative crystal forms was initialized with native protein (23 mg/ml). Both protein with and without intact His-tag was used in this screening process. A new crystal form was obtained in 2.0 M sodium formate, 0.1 M Bis-tris, pH 7.0. After optimization, the best diffracting crystals were grown in 1.9 M sodium formate, 0.1 M Bis-tris, pH 6.5, at room temperature in 2 days by the hanging drop vapor diffusion method. Diffraction data were collected to 1.9 Å (\( \lambda = 1.033 \AA \)) on a PHPT1 crystal at Beamline ID29 (European Synchrotron Radiation Facility, Grenoble, France). The data were indexed in space group R3 with unit cell parameters \( a = b = 112.5 \AA, c = 29.4 \AA. \)

In the new crystal form, there is only one PHPT1 molecule in the asymmetric unit. Diffraction data were processed using XDS/XSCALE (14). The large cell model was used as the starting model to obtain a molecular replacement solution using MOLREP (18). The resulting model was improved through several rounds of model building and refinement in Coot (19) and REFMAC5 (18). Progress of the refinement was monitored by the free \( R \) factor for 5% of the data (20). The final model consisted of 100 amino acid residues, 17 formate ions, and 59 waters. The N-terminal His-tag and linker (17 residues), residues 31–38, and the last three C-terminal amino acids are not visible in the structure. The final residuals for this model were \( r = 17.1\% \) and \( R_{\text{free}} = 22.5\% \).

Structural Similarity Searches—DALI (21) was used to determine the closest structural neighbors of PHPT1. The top five structures based on statistical significance (Z-score cutoff of 3.3) were aligned and compared with PHPT1.

Modeling of Ligands—The peptides succinyl-Ala-His(P)-Pro-Phe-p-nitroanilide, Met-Gly-His(P)-Ala-Gly-Ala-Ile, and Tyr-Ser-His(P)-Asp-Asn-Ile-Ile-Cys-Gly were modeled into the putative active site using the Molegro virtual docker (22). The sequences with phosphorylated histidines MGHAGAI and YSHDNIICG were chosen based on the conserved sequence among succinyl-CoA synthetases and its structure (23) and a loop between two \( \beta \)-strands in the \( \beta \)-subunit of the G protein transducin structure (24), respectively. In the case of the \( \beta \)-subunit of the G protein transducin peptide, the histidine was modeled as phosphorylated on the N\( \varepsilon \) position. Typical docking runs consisted of docking a single ligand with the PHPT1 protein with no solvent molecules. The search algorithm typically used default values for 5–10 runs. In some cases the population size was increased from 50 to 100. Multiple poses were returned for each run and those with phosphohistidines near His\( ^{53} \) were considered potential ligands.

Protein Structure Accession Numbers—The atomic coordinates and structure factors have been deposited in the Research

---

TABLE 1
Data collection and refinement statistics

| Data collection | Native R3, small cell | Peak | Se-Met R3 remote | Inflection |
|-----------------|----------------------|------|-----------------|-----------|
| Wavelength (Å)  | 1.033                | 0.9798 | 0.95373 | 0.97997 |
| Resolution (Å)  | 30–1.9 (2.0–1.9)     | 30–2.0 (2.1–2.0) | 30–2.0 (2.1–2.0) | 30–2.0 (2.1–2.0) |
| Completeness (%)| 97.5 (88.8)          | 96.9 (88.2) | 98.1 (99.0) | 96.4 (85.7) |
| Envelope (%)(i)| 20.62 (6.47)         | 11.61 (3.85) | 11.71 (3.72) | 9.7 (3.28) |
| Refinement      | 5.6 (22.6)           | 6.3 (29.4) | 5.6 (23.1) | 6.1 (26.8) |
| Resolution (Å)  | 0.97997              | 21.3–1.9 (2.0–1.9) | 75,450/208,920 | 76,343/214,276 |
| Reflections (%)(unique/observed) | 17.1 (19.4) | 22.5 (24.4) & | 74,707/165,905 |
| r.m.s.d.        | Bond lengths (Å)     | 0.020 |                 |           |
| Bond angles (°) | 1.8                  | 24.4 |                 |           |
| Ramachandran statistics (%) | 98.5, 1.5 | | | |
| Number of atoms | Protein              | 985  |                 |           |
|                  | Water                | 59   |                 |           |
|                  | Formate              | 51   |                 |           |

\[ a = b = 112.5 \AA, c = 29.4 \AA. \]

---

K. Cowtan, unpublished results.
RESULTS AND DISCUSSION

Structural Model—PHPT1 is a 125-amino acid human cytosolic protein that folds into a compact elbow-shaped molecule of a mixed $\alpha/\beta$ fold with novel topology (Fig. 1, A and B). The molecule is defined by six $\beta$-strands flanked by two $\alpha$-helices. A single helix and three strands lie approximately perpendicular to each other near the base of the molecule. There are two central anti-parallel $\beta$-strands ($\beta_2$ and $\beta_4$) that extend the full length of the protein. On either side of these strands are two sets of parallel $\beta$-strands ($\beta_1$, $\beta_3$ and $\beta_5$, $\beta_6$). The core of the protein is defined by this $\beta$-sheet and the N-terminal histidine tag extends from it. Electron density for the tag was partly not interpretable so it was not modeled, but it is likely required for crystallization because of the numerous crystal contacts that is formed with this extended strand. On the surface there is a positively charged pocket that consists of conserved residues (Fig. 2, A and B). This area is a potential site for phosphatase activity because of its positive surface charge. The cavity volume is $\sim 30 \, \text{Å}^3$ and found at the base of the molecule (Fig. 1C). Two individual amino acid substitutions to alanine resulted in the loss of phosphatase activity (11). These mutations, His$^{53} \rightarrow$ Ala and His$^{102} \rightarrow$ Ala, map to this region of the enzyme. Additionally, many identical and conserved residues of orthologous proteins define this basic patch and surrounding atoms (Fig. 1D). Formate and sulfate ions are found in this pocket in both the small and large cell structures, respectively (Fig. 2, A and B). Both ions bind in similar locations near His$^{53}$ in the positive patch but form a binding network distinct from one another. This arrangement suggests there may be multiple ways to bind substrates within the active site based on their size and charge. In total, 17 formate ions solvate PHPT1. Some of these ions are bound to surfaces distinct from the active site.

A DALI (21) search revealed that the closest structural neighbors of PHPT1 in the Protein Data Bank are of no apparent relation. Polyphosphate kinase ($Z$ score = 4.3, r.m.s.d. 3.1 Å over 74 Ca atoms), PurS ($Z$ score = 3.4, r.m.s.d. 2.4 Å over 50 atoms), and plasma gelsolin ($Z$ score = 3.4, r.m.s.d. 2.8 Å over 63 atoms) were the most similar. Weak structural homology to SixA (6) is found in the central $\beta$-sheet with 23 Ca atoms superimposable within a distance of 3.8 Å.

Active Site—The residues that define the active site are Lys$^{21}$, Glu$^{51}$, Tyr$^{52}$, His$^{53}$, Arg$^{78}$, Tyr$^{93}$, and Met$^{95}$ (Fig. 2C). Almost all of these residues are identical in orthologous proteins among the animal kingdom. Therefore, these residues likely contribute important structural features to maintain the phosphatase activity of the enzyme. The active site reside His$^{53}$ is located at
the beginning of helix α1; near this residue, two formate ions and two water molecules form an intricate hydrogen bonding network. A total of five formate molecules are found in a pocket near His^53. The pocket is defined by two loops; Glu^51 and Tyr^52 on a loop joining helix α1 and β-strand β4 define one side of the pocket while Tyr^93 and Met^95 on a loop between β5 and α2 define the other side. Furthermore, Arg^78 is found in the middle of the β4 strand and may be involved in coordinating the phosphate group. The Arg^78 → Ala mutant resulted in a 30% decrease in activity (11), suggesting it does play a role in phosphatase substrate recognition or catalysis. There are three conserved glycines immediately upstream of Arg^78; these glycines offer Arg^78 a great deal of conformational flexibility, presumably needed in substrate recognition and catalysis. Furthermore, in a scenario where substrate interacts with Arg^78, the nearby glycine-rich region may flex so that a tightening of the pocket around the peptide substrate occurs. Close to Arg^78, Lys^21 is located in the middle of the β2 strand and is involved in hydrogen bonding to a formate ion. In the pocket, only Arg^78 and Lys^21 are found on β-strands. These residues may act as anchors and assist in coordination of the phosphohistidine to the active site. It is not known whether the N^61 or N^62 of the imidazole ring of His^53 is involved in dephosphorylation and it is unfortunately not clear from the structure in what orientation the imidazole ring should be positioned. From a conformational standpoint, the side chain of His^53 projects radially from its turning backbone with the imidazole pointing downward the bottom of the active site pocket, a substrate N^61 interaction is most likely. Interestingly, His^53 is calculated to have a pK_a of 3.13 using PROPKA (25). Therefore, it is likely that His^53 is not protonated at the N^62 position in the crystallization solution and in the physiological context.

**Substrate Binding**—ATP-citrate lyase (ACL) is a substrate for PHPT1 and it has been suggested that PHPT1 dephosphorylates phospho-His^760 (9). A BLAST (26) search revealed that a conserved domain (SucD, succinyl-CoA α-subunit) is found from residues 649–777 of human ACL. The histidine maps to this domain and sequence analysis show that this residue and adjacent residues (GHAGA) are highly conserved among higher eukaryotes (data not shown). The structure of this subunit contains a phosphorylated histidine at residue His^259 (23). This amino acid corresponds to His^760 in CLUSTALW (27) sequence alignments of the succinyl-CoA α-subunit with ACL. This residue is found on an extended loop of 22 residues between two α-helices near the surface of the protein. Considering the length and makeup of the loop, it is entirely possible that it is flexible and could bind in the pocket of PHPT1 (Fig. 2D).

The β-subunit of G proteins is phosphorylated on histidine residue 266 (28–31) and is dephosphorylated by PHPT1 (10). The structure of the β-subunit (24) shows this residue in a surface loop of eight amino acids connecting two β-strands. This small loop is modeled in the pocket of PHPT1 (Fig. 2E). In these two cases it appears that some rearrangement of these loops would be required for the substrate to bind in the pocket. It is unclear if there are additional sites of binding separate from the active site. The binding pocket of PHPT1 can accommodate a variety of ligands and studies are in place to assess possible inhibitors and binding partners that should allow for the mechanism of action to be discerned.

The structure of PHPT1 is the first glimpse at a eukaryotic histidine phosphatase. The protein is of a novel topology, and the structure reveals a positively charged active site defined by

**FIGURE 2. The putative active site.** A, solid surface representation of the proposed active site with five formate ions. B, solid surface representation of the proposed active site with five sulfate ions. The colors indicate positive (blue) and negative (red) electrostatic potential at the solvent accessible surface. The electrostatic calculations were done with APBS (32). C, PHPT1 active site residues and solvent molecules. The pocket is defined by the labeled residues. Also visible are five formate ions and two waters (red spheres). Schematic coloring is as for Fig. 1A. D, peptide MGHAGAI of succinyl-CoA synthetase bound in one possible orientation in the putative active site of PHPT1. The ligand (magenta) is hydrogen-bonded to Lys^21 and His^53. E, peptide YSHDNICG of the β-subunit of the G protein transducin bound in a possible orientation in the proposed active site of PHPT1. The ligand (magenta) is hydrogen-bonded to His^53, Arg^93, and Tyr^93. Residues labeled in purple and black are for the ligand and PHPT1, respectively. In both cases, the histidine in the ligand is phosphorylated at the N^62 position. The nitrogen (N^61 or N^62) of His^53 that interacted with substrate varied within each set of docking solutions.
conserved residues that provide a suitable environment for binding phosphohistidine-containing substrates. His^{53} is proposed to be involved with the phosphatase activity of this enzyme. Furthermore, Arg^{78} and Lys^{21} may act as anchors to provide a stable scaffold for substrate and phosphohistidine interactions as well as provide a possibility for charge stabilization of transition states in the catalytic reaction. Future challenges include investigation of co-crystal structures with target peptides to understand in detail the structural basis for substrate recognition and catalysis.

Acknowledgments—We acknowledge support during data collection from Martin Fieber-Erdmann and Laurent Terradot at Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY) and the European Synchrotron Radiation Facility, respectively.

REFERENCES

1. Pawson, T., and Scott, J. D. (2005) Trends Biochem. Sci. 30, 286–290
2. Chen, C. C., Smith, D. L., Bruegger, B. B., Halpern, R. M., and Smith, R. A. (1974) Biochemistry 13, 3785–3789
3. Besant, P. G., and Attwood, P. V. (2005) Biochim. Biophys. Acta 1754, 281–290
4. Matthews, H. R. (1995) Pharmacol. Ther. 67, 323–350
5. Klumpp, S., and Krieglstein, J. (2005) Biochim. Biophys. Acta 1754, 291–295
6. Hamada, K., Kato, M., Shimizu, T., Ibara, K., Mizuno, T., and Hakoshima, T. (2005) Genes Cells 10, 1–11
7. Klumpp, S., Hermesmeier, J., Selke, D., Baumeister, R., Kellner, R., and Krieglstein, J. (2002) J. Cereb. Blood Flow Metab. 22, 1420–1424
8. Ek, P., Pettersson, G., Ek, B., Gong, F., Li, J. P., and Zetterqvist, O. (2002) Eur. J. Biochem. 269, 5016–5023
9. Klumpp, S., Bechmann, G., Maurer, A., Selke, D., and Krieglstein, J. (2003) Biochim. Biophys. Res. Commun. 306, 110–115
10. Maurer, A., Wieland, T., Meissl, F., Niroomand, F., Mehringer, R., Krieglstein, J., and Klumpp, S. (2005) Biochem. Biophys. Res. Commun. 334, 1115–1120
11. Ma, R., Kanders, E., Sundh, U. B., Geng, M., Ek, P., Zetterqvist, O., and Li, J. P. (2005) Biochem. Biophys. Res. Commun. 337, 887–891
12. Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Marchler, G. H., Mullokandov, M., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Yamashita, R. A., Yin, J. J., Zhang, D., and Bryant, S. H. (2005) Nucleic Acids Res. 33, D192–D196
13. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) J. Med. Chem. 36, 104–125
14. Kabsch, W. (1993) J. Appl. Crystallogr. 26, 795–800
15. Schneider, T. R., and Sheldrick, G. M. (2002) Acta Crystallogr. Sect. D Biol. Crystallogr. 58, 1772–1779
16. de La Fortelle, E., Bricogne, G., and Carter, C. W., Jr. (1997) Methods Enzymol. 276, 472–494
17. Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S. (2001) Acta Crystallogr. Sect. D Biol. Crystallogr. 57, 1445–1450
18. Collaborative Computational Project, No. 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
19. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D60, 2126–2132
20. Brunger, A. T. (1992) Nature 355, 472–475
21. Holm, L., and Sander, C. (1993) J. Mol. Biol. 233, 123–138
22. Thomsen, R., and Christensen, M. H. (2006) J. Med. Chem. 49, 3315–3321
23. Fraser, M. E., James, M. N., Bridger, W. A., and Wolodko, W. T. (2000) J. Mol. Biol. 299, 1325–1339
24. Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E., and Sigler, P. B. (1996) Nature 379, 369–374
25. Li, H., Robertson, A. D., and Jensen, J. H. (2005) Proteins 61, 704–721
26. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) J. Mol. Biol. 215, 403–410
27. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673–4680
28. Kowluru, A., Seavey, S. E., Rhodes, C. J., and Metz, S. A. (1996) Biochem. J. 313, 97–107
29. Wieland, T., Nurnberg, B., Ulbarri, I., Kaldenberg-Stasch, S., Schultz, G., and Jakobs, K. H. (1993) J. Biol. Chem. 268, 18111–18118
30. Cuello, F., Schulze, R. A., Heemeyer, F., Meyer, H. E., Lutz, S., Jakobs, K. H., Niroomand, F., and Wieland, T. (2003) J. Biol. Chem. 278, 7220–7226
31. Hippe, H., Lutz, S., Cuello, F., Knorr, K., Vogt, A., Jakobs, K. H., Wieland, T., and Niroomand, F. (2003) J. Biol. Chem. 278, 7227–7233
32. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 10037–10041
33. Delano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA