

Crystallographic and Modeling Study of the Human PPA1 (Inorganic Pyrophosphatase 1): a Potential Anti-cancer Drug Target

Haiying Niu¹, Jiang Zhu², Quanxin Qu¹, Xia Zhou², Xiaolan Huang², and Zhihua Du²

¹Tianjin First Central Hospital

²Southern Illinois University Carbondale

August 12, 2020

Abstract

Inorganic pyrophosphatases (PPases) catalyze the hydrolysis of pyrophosphate to phosphates. PPases play essential roles in growth and development, and are found in all kingdoms of life. Human possess two PPases, PPA1 and PPA2. PPA1 is present in all tissues, acting largely as a housekeeping enzyme. Besides pyrophosphate hydrolysis, PPA1 can also directly dephosphorylate phosphorylated JNK1. Upregulated expression of PPA1 has been linked to many human malignant tumors. PPA1 knockdown induces apoptosis and decreases proliferation. PPA1 is emerging as a potential prognostic biomarker and target for anti-cancer drug development. In spite of the biological and physiopathological importance of PPA1, there is no detailed study on the structure and catalytic mechanisms of mammalian origin PPases. Here we report the crystal structure of human PPA1 at a resolution of 2.4 Å. We also carried out modeling studies of PPA1 in complex with JNK1 derived phosphor-peptides. The monomeric protein fold of PPA1 is similar to those found in other family I PPases. PPA1 forms a dimeric structure that should be conserved in animal and fungal PPases. Analysis of the PPA1 structure and comparison with available structures of PPases from lower organisms suggest that PPA1 has a largely pre-organized and relatively rigid active site for pyrophosphate hydrolysis. Results from the modeling study indicate the active site of PPA1 has the potential to accommodate double-phosphorylated peptides derived from JNK1. In short, results from the study provides new insights into the mechanisms of human PPA1 and basis for structure-based anti-cancer drug developments using PPA1 as the target.

Abbreviations List:

HRV-3C: Human Rhinovirus 3C Protease **JNK1:** c-Jun N-terminal kinases **1PPA1:** Inorganic Pyrophosphatase 1 **PPA2:** Inorganic Pyrophosphatase 2

Introduction

Inorganic pyrophosphatases (PPases) catalyze the hydrolysis of one molecule of pyrophosphate (PPi) to two molecules of phosphates (Pi). PPases are ubiquitous enzymes present in all cell types. PPases include membrane-bound V-PPases (vacuolar H⁺-translocating PPases) and soluble PPases. V-PPases use pyrophosphate (instead of ATP) as an energy source to transport H⁺ or Na⁺ across membranes in plants and some bacteria, archaea, and parasitic protists [1, 2]. Soluble PPases are divided into three families that differ in their sequences and structures [3-5]. PPases from families II and III are only found in prokaryotes, whereas Family I PPases exist in all kingdoms of life. Family I PPases are single-domain, Mg²⁺ dependent enzymes that exist as hexamers in prokaryotes and dimers in eukaryotes.

PPi is a metabolic product that is generated by many biosynthetic reactions for the synthesis of various important biomolecules, including proteins, nucleic acids, polysaccharides, and fatty acids [6]. The conversion of PPi to Pi by soluble PPases is an essential cellular process. On the one hand, the reaction is critical for

maintaining cellular PPi homeostasis; on the other hand, the reaction provides a thermodynamic pull for the PPi-generating biosynthetic reactions because enzymatic hydrolysis of PPi is highly exergonic.

Soluble PPases are essential enzymes for growth and development. In *E. coli*, the *ppa* gene for PPase is essential for cell growth [7]. In fermenting yeast, PPase defects lead to cell cycle arrest and autophagic cell death [8]. In *C. elegans*, the PPase enzyme PYP-1 is required for larval development and intestinal function [9, 10]. In mammals, the PPase enzyme PPA1 has been reported to have the potential to regulate neurite growth in mouse neuroblastoma cells [11], to induce type I collagen synthesis and stimulate calcification by osteoblasts [12]. Increased expression and activity of cytosolic PPA1 in the liver of rat and mouse has been correlated with aging [13, 14].

Humans possess two Family I PPases, PPA1 (289 aa) and PPA2 (305 aa) [15]. PPA2 is localized to the mitochondrion. The two human PPases share ~70% homology. PPA1 is a ubiquitous enzyme found in all tissues. It is found that PPA1 acts largely as a housekeeping gene [15], presumably due to its essential role in energy metabolism. More recently, the role of PPA1 in tumorigenesis has emerged [16-19]. Upregulated expression of PPA1 has been linked to many human malignant tumors including colorectal cancer [20-22], diffuse large B-cell lymphoma (DLBCL) [23], lung adenocarcinoma [17, 24], prostate cancer [25], hepatocellular carcinoma [19], breast cancer [26], gastric cancer [27, 28], intrahepatic cholangiocarcinoma (ICC) [29], and ovarian cancer [24, 30, 31]. In laryngeal squamous cell carcinoma (LSCC), metastases tumors showed an intense increase of PPA1 compared to primary tumors [32]. These data indicate that PPA1 protein is a potential prognostic biomarker for certain cancers.

The exact mechanism for the involvement of PPA1 overexpression in tumorigenesis is not fully understood. Dysregulated metabolism is an important hallmark of cancer cells [33]. Metabolic alterations caused by PPA1 overexpression may provide the extra energy and biomaterials for the cancer cells to maintain their high proliferation rate. In addition, it is found that PPA1 can directly dephosphorylate phosphorylated JNK1 (c-Jun N-terminal kinase) [22]. Functioning as a JNK phosphatase, PPA1 may regulate the activities of downstream effectors (including p53, β -catenin, Bcl-2, and Caspase-3) that control cell proliferation and apoptosis. In consistence with this speculated mechanism, we previously found that PPA1 overexpression in epithelial ovarian carcinoma (EOC) promoted the dephosphorylation and translocation of β -catenin [31]. In another study on EOC [30], it was found that PPA1 overexpression caused the decrease in p53 and increase in β -catenin expression level. Conversely, PPA1 knockdown led to the increase of p53 and decrease of β -catenin level.

While PPA1 overexpression enhances migration, invasion and proliferation of cancer cells [27], PPA1 knockdown (by RNAi mediated silencing) is shown to induce apoptosis and decrease proliferation in various cancer cells [22-24, 30, 31]. Results from these studies suggest that strategies for PPA1 down-regulation may have therapeutic potential for the treatment of certain cancers. PPA1 down-regulation might be achieved by mRNA knockdown through RNAi or RNA-targeting CRISPR systems. Alternatively, the PPA1 protein might represent a valuable target for the development of specific small-molecule inhibitors as anti-cancer drugs. Efforts had been made to target PPases from pathogenic organisms for small-molecule drug development. Anti-PPase small molecules were shown to have promising antibiotic activity against drug resistant strains of *Staphylococcus aureus* [34]. Small molecule inhibitors of the VSP1 (vacuolar soluble pyrophosphatase 1) protein from *Trypanosoma brucei*, the causative agent of African trypanosomiasis (HAT), provided a 40% protection from death in a mouse model of *T. brucei* infection [35]. Inhibition of PPase activity could also be achieved by targeting allosteric sites, as demonstrated by the discovery of allosteric and selective inhibitors of PPase from *Mycobacterium tuberculosis* [36].

In spite of the biological and physiopathological importance of human PPA1, there is no detailed study on the structure and catalytic mechanisms of this enzyme. Extensive structural and biochemical studies had been reported on Family I PPases from *Escherichia coli* (E-PPase) [37-39] and *Saccharomyces cerevisiae* (Y-PPase) [40-45]. X-ray crystallographic studies had also been carried out on soluble PPases from other species, including *Thermus thermophilus* (T-PPase) [46], *Pyrococcus horikoshii* (Pho-PPase) [47], human parasites *Plasmodium falciparum* (Pf-PPase) and *Toxoplasma gondii* (Tg-PPase) [48], *Mycobacterium tuberculosis*

(Mt-PPase) [36, 49], *Trypanosoma brucei brucei* (also called vacuolar soluble protein TbbVSP1) [50], and *S. japonicum* (Sj-PPase, not published).

In this paper, we present the crystal structure of human PPA1. We also carried out modeling studies of PPA1 in complex with JNK derived peptides containing phosphotyrosine and phosphothreonine residues.

Materials and Methods

Molecular cloning, protein sample preparation and crystallization

A cDNA plasmid that contains a DNA fragment encoding the full-length human PPA1 (GenBank NC_000010.11) was purchased from DNASU Plasmid Repository. The protein product of this cDNA corresponds to the human PPA1 protein with GenBank accession No. NP_066952. The DNA encoding the full length PPA1 (residues 1-289) was amplified by PCR from the plasmid with the forward primer 5'-CAAGGACCGAGCAGCCCTCA atgagcggttcagcaccg-3' and the reverse primer 5'-ACCACGGGGAACCAACCCTTA CTAGTTTTTCTGGTGATGGAAC-3'. Each of these primers contains a 5'-end specific sequence (21 nucleotides, underlined) for ligation-independent cloning (LIC). The PCR product was purified by agarose gel electrophoresis and processed by T4 DNA polymerase in the presence of 2.5 mM dATP to generate 5'-overhangs at both ends. To find an appropriate way to express the PPA1 protein in *E. coli* in a soluble form, we had tested a series of in-house developed LIC cloning vectors containing different fusion tags, including the MBP, GST, TRX, SUMO, HALO, NusA, and Fc tags, etc. The cloning vector with a HALO tag [51] gave the best result. The cloning vector encodes the Halo tag, His tag, and an eight amino-acid recognition sequence for human rhinovirus (HRV) 3C protease in front of a specific LIC sequence. The vector was linearized by SmaI digestion and subsequently processed by T4 DNA polymerase in the presence of 2.5 mM dTTP to generate 5'-overhangs at both ends. The processed PCR insert and vector were mixed and incubated at room temperature for 20 minutes. The mixture was used to transform NEB-turbo*Escherichia coli* competent cells. The recombinant plasmid was transformed into NiCo21(DE3) *E. coli* cells (New England Biolabs). Cell culture was grown in LB media until it reached an OD₆₀₀ of 0.6-0.8. Isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to the culture (at 0.4 mM final concentration) to induce protein expression. After induction, the temperature was lowered to 288 K and the culture was allowed to grow overnight before harvest.

The PPA1 protein was expressed in *E. coli* as a fusion protein with an N-terminal Halo tag and His tag, and a HRV 3C protease recognition for cleavage motif. The fusion protein was purified by NTA affinity resin. The fusion tags were then cleaved by HRV 3C protease. After removal of the tags, the target protein contains an artificial sequence GPSSP at the N-terminus as a cloning artifact. The cleaved tags were separated from the target protein by a reverse IMAC (immobilized metal-affinity chromatography) process with NTA resin. The purified target proteins were concentrated to ~10 mg/ml in a buffer containing 25 mM Tris (pH 7.5), 200 mM NaCl, and 1 mM DTT.

Crystallization trials were carried out using 96-well-format plates. Ten sets of in-house prepared screening solutions (each containing 96 different conditions with various PEGs as the precipitants) were used in the trials. Multiple crystallization conditions were identified within a few days. After optimizing the crystallization conditions, diffracting crystals of the PPA1 were obtained at 295 K by hanging-drop vapor diffusion against 50 μ l well solution using 96-well format crystallization plates. The crystallization drops contained 2 μ l protein sample mixed with 2 μ l well solution. The well solution contains 20% PEG 3350, 100 mM Bis-Tris propane (pH 7.5), 200 mM KBr, and 10% (v/v) glycerol. Crystals were flash-cooled in liquid nitrogen.

Data collection, data processing, and structure determination

Data collection was carried out at beamline 21ID-D of LS-CAT at the Advanced Photon Source (Argonne National Laboratory). Data were processed, integrated, and scaled with the programs Mosflm and Scala in the CCP4 package [52]. Phasing was achieved by using the molecular replacement method, which was carried out by using the program Phaser [53] that is implemented in the PHENIX package [54]. The monomeric structure of the *E. coli* PPase (PDB code 4um4) was used as a search model. This structure is chosen as a

search model because All plausible space groups of the P222 family were tried. A single solution was found in the space group P212121. Interactive model building was carried out with Coot [55]. Structure refinement was carried out by PHENIX. Structure determination statistics are shown in Table 1. The program UCSF Chimera [56] was used to generate all of the molecular graphic figures. Figures for electrostatic surface rendering were generated by Chimera using the Coulombic Surface Coloring tool with the default settings.

Sequence alignments of the eukaryotic soluble PPases with known structures were carried out by Clustal Omega [57]. Molecular interactions at the PPA1 homodimerization interface were depicted by using the program LIGPLOT [58]. Atomic coordinates and diffraction data have been deposited in the Protein Data Bank with accession code 6C45.

Structural modelling of PPA1 with various putative substrates

The determined crystal structure of human PPA1 does not contain substrate at the putative active site. To gain insights into the possible catalytic mechanism of human PPA1, modelling studies were carried out to place various putative substrates at the active site. At first, a pyrophosphate molecule was docked into the active site by using AutoDock Vina [59]. Chimera was used to prepare the PPA1 receptor structure for docking and to visualize the docking results. To explore whether the PPA1 active site is capable of accommodating peptide substrates containing phosphorylated tyrosine and/or phosphorylated threonine residues, a series of phosphotyrosine and/or phosphothreonine containing peptides were modelled into the PPA1 active site by using Chimera (including energy minimization of the bound peptide at the active site). The phosphotyrosine and/or phosphothreonine containing peptides were initially derived from the phosphorylated MAP kinase Erk2 (PDB code 2erk) [60]. Two Erk2-derived tetrapeptides were used: 184-EY^pVA (superscript p indicates phosphotyrosine) and 181-FLT^pE (superscript p indicates phosphothreonine), to model a Y^p and a T^p at the active site respectively. To prepare peptides corresponding to the JNK1 sequence for modelling studies, the 181-FLT^pEY^pV sequence in the Erk2 structure were mutated to JNK1 sequence 181-MMT^pPY^pV by Chimera. A pentapeptide 182-MT^pPY^pV was used to model the Y^p at the active site. A hexapeptide 181-MMT^pPY^pV was used to model the T^p at the active site.

Results and Discussion

The overall structure of human PPA1 is similar to other Family I PPases

The crystal structure of human PPA1 was determined at a resolution of 2.39 Å using the molecular displacement (MR) method. A monomeric structure of E-PPase was chosen as the search model to minimize bias on the structure determination. The sequence of human PPA1 is much longer than that of E-PPase (289 residues vs 176 residues). The two sequences are only marginally homologous (residues 44-194 in human PPA1 matches residues 16-143 with 27% identity, 41% similarity, and 18% gap). It is also known that prokaryotic and eukaryotic PPases assume different oligomerization states in the crystal structures. The fact that the use of a monomeric E-PPase structure as a search model for MR readily leads to a clear solution of the structure of human PPA1 indicates that the PPA1 and E-PPase share highly homologous core sstructures.

The polypeptide chain of human PPA1 adopts a single domain globular fold consisting of seventeen β-strands, four α-helices, and four ₃₁₀-helices (Figures 1 and 2). The core of the globular fold is composed of a 5-stranded antiparallel β-barrel (strands 5, 11, 13, 14 and 15) and a 5-stranded antiparallel β-sheet (strands 1, 2, 3, 6 and 7) (Figure 2A). The β-barrel and β-sheet are packed closely together. Other parts of the molecules, including the four α-helices, four ₃₁₀-helices, and five short β-strands (strands 4, 8, 9, 10, 12, 16, and 17) surround the β-core structure. A parallel 2-stranded β-sheet formed by the two short strands β10 and β17 helps to anchor the C-terminus of the polypeptide to the protein fold.

Among the soluble PPases with known structures, human PPA1 share the highest sequence homology with the PPases from *S. japonicum* and *S. cerevisiae* (about 70% similarity, see Figure 1). The structures of the

PPases from these three species are similar. Superimposition of the monomeric human PPA1 structure to the monomeric structure of PPase from *S. japonicum* (PDB code 4QLZ) and *S. cerevisiae* (PDB code 2IHP) gave a RMSD of 0.98 Å (based on 3032 common atoms) and 0.97 Å (based on 2981 common atoms) respectively. Figure 2B shows a superimposition of the monomeric structures of human PPA1 and Y-PPase. The two structures superimpose well in most parts of the molecules, including the active site. Some differences exist in the α 4-helix, β 8, β 9, β 10, β 17 regions, as well as some connecting loops. In comparison with the structure of E-PPase (PDB code 4UM4), the β -barrel, β 6, β 7, α 1-helix, α 2-helix, and the N-terminal portion of α 3-helix of human PPA1 have their counterparts in the E-PPase structure (RMSD=2.60 Å over 640 common atoms). These portions of the structure are common in all known structures of soluble Family I PPases.

Human PPA1 and other three eukaryotic PPases (Sj-PPase, Y-PPase, and Pf-PPase) have a C-terminal extension (after the α 3-helix) sequence compared to other Family I PPases (Figure 1). This C-terminal extension sequence assumes a similar structure in human PPA1, Sj-PPase, and Y-PPase (Figure 2B), and is involved in the homodimerization of the PPases (see below). The C-terminal extension in Pf-PPase assumes a different structure that is not involved in the homodimerization of Pf-PPase [48].

Human PPA1 forms a dimeric structure that is conserved in a subset of Family I PPases

In the crystal, human PPA1 exists as a homodimer. The two protomers in the homodimer assume a relative orientation that is analogous to an identical twin standing arm in arm, facing opposite directions (Figures 2C and 2D). Similar homodimers are observed in the structures of PPases from *S. japonicum* and *S. cerevisiae* (Sj-PPase and Y-PPase, PDB codes 4QLZ and 2IHP). Formation of the human PPA1 homodimer buries 1860 Å² of solvent accessible surface area (SASA) from the two protomers, comparable to those in the Sj-PPase and Y-PPase homodimers (1830 Å² and 2030 Å² respectively). The three crystals have different space groups (P21 21 21, P32, and P1 21 1). The presence of similar homodimers in the structures of the three PPases crystallized in different space groups and the large buried SASA of dimerization indicate that formation of the homodimer is unlikely due to crystal packing artifact.

The homodimerization of human PPA1 is driven by multiple factors including shape complementarity of the dimerization interface, hydrophobic contacts, intermolecular hydrogen bonds, and electrostatic interactions. A large number of residues, including Arg52, Trp53, Asn83, Phe85, Pro86, Lys88, Ser127, Val129, Asp165, Lys179, Pro180, Gly181, Tyr182, Ala185, Asp281, Lys282, Trp283, His285, and His286 are involved in homodimerization (Figures 3 and 4). These residues are scattered in six different regions within the primary sequence, with the three regions around Phe85 (strand β 10 and the following loop), around Pro180 (linker between α 1- α 2 helices and n-terminus of α 2-helix), and around Trp283 (3_{10} helix η 4, strand β 17, and the following loop) having most of the residues (Figure 1). Importantly, residues within the C-terminal extension play an important role in dimerization by not only directly participating in the interfacial interactions but also holding the β 10 region in position.

A large portion of the interface molecular surface is defined by the sidechains of the hydrophobic residues and aliphatic parts of the sidechains of other residues (white to orange red surface areas in Figure 4). Hydrophobic contacts in the dimerization interface are extensive (Figure 3). Some examples include the contacts mediated by Phe85, Pro86, and Trp283 (Figure 4), which are conserved among the three eukaryotic PPases (human PPA1, Sj-PPase, and Y-PPase, see Figure 1).

There are several hydrogen bonds at the dimerization interface. Except the one between the backbone oxygen of Asp281 and sidechain of Arg52, all other hydrogen bonds are mediated by structured water molecules (Figure 3). Two electrostatic interactions are observed, between the sidechains of the Asp165-Lys282 and Asp281-Arg52 pairs. The Asp281-Arg52 pair is conserved in human PPA1, Sj-PPase, and Y-PPase (Figure 1).

Soluble Family I PPases are known to exist in different oligomeric states. Prokaryotic PPases form hexamers under physiological conditions [37-39, 46, 47, 49]. All but one known eukaryotic PPases structures form

dimers [40-45, 48]. The exception is Tb_bVSP1 , which forms a tetramer (dimer of dimer) [50]. The crystal structure of human PPA1 reveals a dimerization mode that is conserved in the S_j-PPase and Y-PPase. The C-terminal extensions in these PPases are critically involved in dimerization. Other eukaryotic PPases either do not have a C-terminal extension (T_g-PPase and Tb_bVSP1) or has a C-terminal extension in different configurations (P_f-PPase). The dimerization modes in T_g-PPase, Tb_bVSP1 , and P_f-PPase are different from each other, and different from the conserved mode among human PPA1, S_j-PPase, and Y-PPase [50]. The available crystal structures of eukaryotic PPases show that diverse modes of dimerization exist in eukaryotic soluble Family I PPases. Previously, it was proposed (based on sequence conservation) that soluble Family I PPases from animal and fungi might share a similar mode of dimerization [50]. The crystal structure of human PPA1 provides a strong piece of evidence to support this proposal.

Although it is now generally believed that soluble Family I PPases exist in a multimeric state, the functional roles of multimerization and its structural diversity on PPase function, if any, are not known. In the case of human PPA1 (and the homologous S_j-PPase and Y-PPase), dimerization places the two active sites (one from each protomer) on opposite molecular surface of the dimer, far away and isolated from each other (Figures 2C and 2D). Catalytic reactions at the two active sites should be independent to each other. Residues involved in dimerization are also far away from the active site in the monomeric structure, dimerization should only have allosteric effect, if any, on the active site. Of course, even if dimerization is not required for the phosphatase function of human PPA1 under physiological condition, it may still be relevant to other PPA1 function(s) that is not known currently.

A very outstanding feature of the human PPA1 dimeric structure is the presence of a large cleft at the dimerization interface (Figure 2D). The size of the cleft can easily accommodate a 4-turn α -helix. Whether this cleft represents a functional site of human PPA1 is not known and deserves further studies. From the perspective of structure-based drug development using human PPA1 as a target, the dimerization interface cleft may serve as a useful allosteric target site. In the case of Mt-PPase, inhibitors bind in a non-conserved interface between monomers of the hexameric structure were identified, which block the hydrolysis reaction in an uncompetitive and allosteric manner [36].

Human PPA1 has a largely pre-organized active site

Extensive studies had been carried out on the active site structures and catalytic mechanisms of Y-PPase [40-45]. These previous knowledges are critical for analyzing the active site of human PPA1 in the current structure.

The human PPA1 structure does not contain the substrate at the putative active site. To gain insights into how the speculative active site residues (inferred from structure-based sequence alignment with Y-PPase) orient in relative to the pyrophosphate substrate, the substrate was modeled into the human PPA1 structure by superimposition of the structure with a substrate-bound, fluoride-inhibited Y-PPase structure (PDB code 1RE6A) [45]. The overall structures superimposed well with a small RMSD of 0.98 Å (Figure 2B). The portions of structures that define the active site show only very slight difference in backbone and sidechain conformations (Supplementary Figure S1). The superimposed coordinates of the pyrophosphate were merged with the PPA1 coordinates to generate the structure of PPA1 with a pyrophosphate at the active site.

Figure 5A shows the active site structure of human PPA1 with a modeled pyrophosphate substrate. The 14 human PPA1 active site residues (matching the established active site residues in Y-PPase) are Glu49, Lys57, Glu59, Arg79, Tyr94, Gly95, Asp116, Asp118, Asp121, Asp148, Asp153, Lys155, Tyr193, Lys194 (Y-PPase residue numbers are one less than the shown human PPA1 residue numbers). Most of these residues locate in the β -barrel and the adjacent β_6 -strand, which are well conserved and defined in all structures of soluble Family I PPases. Tyr193 and Lys194 make the transition from the 3_{10} -helix η_2 to the α_2 -helix. The active site residues can be divided into two groups based on relative spatial relationship to the pyrophosphate substrate. The first group consists of all of the negatively charged residues and Gly95. These residues cluster on one side of the active site, close to the P2 phosphorus atom. The second group includes all of

the positively charged residues and the two tyrosines. These residues largely locate on the other side of the active site, close to the P1 phosphorus atom (Figure 5A). Studies on Y-PPase reveal that the active site carboxylates are responsible for coordinating with 4 metal ions and activating a water molecule as the reaction nucleophile, while the positively charged sidechains stabilize the transition state and leaving group [40-45].

Due to the fact that most of the active site residues are located in the rigid core of the structure, their positions (locations of the backbone atoms) should have little changes upon substrate binding and during the reaction. Of course, conformations of the sidechains (especially the longer sidechains of lysine and arginine residues) could be changed. Comparison of the apo-structure of human PPA1 and the substrate-bound, fluoride-inhibited Y-PPase structure shows that the active site residues superimpose well with some conformation differences in the sidechains of Lys57, Arg79, and Lys194 (Supplementary Figure S1). Low B-factor values for residues at the active site also suggest relatively lower flexibility of these residues (Figure 5B). These data and analysis indicates that the active site of human PPA1 is largely pre-organized, which would minimize the need for conformational reorganization during catalysis.

The active site of human PPA1 has the potential to accommodate double-phosphorylated peptides from JNK1

While it is well established that PPases catalyze the hydrolysis of pyrophosphate, several recent studies reveal that PPA1 may also function as a protein phosphatase [11, 22, 61]. It was found that human PPA1 could directly dephosphorylate phosphorylated JNK1 in both phosphor-peptide and phosphor-protein levels, while no catalytic activity towards pERK or p-p38 was detected [22]. PPA1-silencing significantly down-regulated colon cancer cell proliferation. This antiproliferation effect is impaired by JNK inhibitor, indicating that the role of PPA1 in colon cancer is at least partially related to regulation of JNK activity. Evidence for the function of PPA1 as a pJNK phosphatase was also obtained in studies on neuronal differentiation in mouse, rat, and chick embryo [11, 61]. It was shown that PPA1 knockdown or overexpression led to increased or decreased JNK phosphorylation level, while no alteration of JNK phosphorylation level was detected after treatment with a catalytically inactive PPA1 mutant. PPA1 may play a role in neuronal differentiation via JNK dephosphorylation. JNK is a member of the mitogen-activated protein (MAP) kinase family. JNK regulates the activity of numerous downstream molecules, including c-Jun, p53, and Bcl2, by phosphorylation. JNK is activated by a dual phosphorylation of Thr183 and Tyr185 within a 180-FMMTPYVV motif. Dephosphorylation of JNK by protein phosphatases inactivates the enzyme [62].

So far, structural and mechanistic studies on soluble Family I PPases only concerned inorganic pyrophosphate as the substrate. Given the emerging role of PPA1 as a JNK phosphatase, we carried out modeling studies to investigate whether the known pyrophosphate binding active site could also accommodate phosphor-peptides.

Four structural models were constructed, each with a short phosphor-peptide bound at the known pyrophosphate binding active site. The phosphorus atom of the phosphotyrosine or phosphothreonine residue within the phosphor-peptide is located at the same position as the P2 phosphorus atom of a pyrophosphate substrate. The sequences of the four phosphor-peptides match either JNK1 or the MAP kinase Erk2. A JNK1-derived pentapeptide 182-MTpPYpV was used to model the Yp at the active site (Figure 6A). A JNK1-derived hexapeptide 181-MMTpPYpV was used to model the Tp at the active site (Figure 6B). Two Erk2-derived tetrapeptides, 184-EYpVA and 181-FLTpE, were used to model the Yp and the Tp at the active site respectively (Supplementary Figure S2).

All of the phosphor-peptides can be accommodated in the active site pocket, without steric clash between the phosphor-peptide and PPA1. The PPA1 structures with a bound phosphor-peptide showed only very minor conformational changes (during energy minimization in Chimera) for residues on the outer portion of the active site pocket. Superimpositions of the phosphor-peptide bound PPA1 structures with the apo-structure give RMSD values less than 0.2 Å. The (putative) active site of human PPA1 that would bind the pyrophosphate substrate is located at the inner and deep portion of a cavity (Figure 5B). The location of the active site is indicated by the black arrow. The cavity is indicated by the white oval). The outer portion of

the cavity becomes much larger, expanding downward and leftward while getting shallower.

The pyrophosphate binding site is mostly defined by residues from the inner β -core and therefore relatively rigid and pre-organized. The outer portion of the cavity is defined by residues from peripheral structures, including the strands $\beta 8$ and $\beta 9$, the 3_{10} helix $\eta 2$, the loop connecting $\beta 6$ and $\beta 7$, and the loop connecting $\beta 12$ and $\beta 13$ (Figure 5A). Since many of these residues are surface exposed, their sidechains (especially the longer sidechains in Lys63, Lys74, Lys75, Glu149, Glu151, and Lys199) are expected to have large flexibilities. The B-factor surface rendering may also reflect the different flexibilities of atoms in different areas of the cavity (Figure 5B). These properties of the cavity explain how the cavity can accommodate various phosphor-peptides in the modeling studies. While the phosphor-residue is reaching into the inner (and deeper) portion of the cavity, the flanking residues are accommodated by the outer, shallower, and more flexible portion of the cavity.

These modeling results indicate that the active site pocket known for pyrophosphate binding and catalysis has the potential to accommodate various phosphor-peptides with single- or double-site phosphorylation. It should be pointed out that the models only minimized steric crash while keeping good geometry of the peptides and protein. No effort was made to optimize the inter-molecular interactions (such as hydrogen bonds, hydrophobic contacts, electrostatic interaction, etc). When these interactions come into play in reality, the PPA1 active site pocket might have substrate specificity for certain phosphor-peptides. Of note, the six charged residues Lys63, Lys74, Lys75, Glu149, Glu151, and Lys199 are placed in different areas in the pocket (in Figure 5B, Lys 63 at bottom left, Lys74 and Lys75 at top left, Lys199 at top right, Glu149 and Glu151 on the left). The peculiar distribution of charged residues may have implications in peptide substrate specificity. Of course, the exact structural details of peptide substrate recognition by human PPA1 can only be revealed by high resolution structures of PPA1 in complex with phosphor-peptide or phosphor-protein substrates. It is known that human PPA1 can directly dephosphorylate pJNK1 but cannot dephosphorylate pERK or p-p38 [22]. Structures of the PPA1-pJNK1 complexes are needed to reveal the molecular basis of this substrate specificity. The structures will also be very helpful for the development of peptide or peptide-like inhibitors of human PPA1.

Conclusion

Prior to our study, many crystal structures of soluble Family I PPases from lower organisms are available. But no structural study on PPases from higher organisms has been reported. Given the emerging role of human PPA1 in tumorigenesis and the potential value of human PPA1 as a target for anti-cancer drug development, it becomes essential to obtain an atomic resolution structure of human PPA1. Such a structure is presented in the current study. The structure, in conjunction with results from modeling studies and previous knowledge on PPase structures, sheds new insights into the mechanism of PPA1 functions. The structure also provides the basis for structure-based drug discovery by using *in silico* virtual screening of small molecule compound libraries, which is being carried out in our lab.

Data Availability

The atomic coordinates and diffraction data for the structures used to support the findings of this study have been deposited in the Protein Data Bank with accession code 6C45.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The work is supported by grant from The Tianjin First Central Hospital of China (yuanCM201810).

References

1. Baykov AA, Malinen AM, Luoto HH, Lahti R. Pyrophosphate-fueled Na⁺ and H⁺ transport in prokaryotes. *Microbiol Mol Biol Rev.* 2013;77(2):267-76. Epub 2013/05/24. doi: 10.1128/mmbr.00003-13. PubMed PMID: 23699258; PubMed Central PMCID: PMCPMC3668671.
2. Tsai JY, Kellosalo J, Sun YJ, Goldman A. Proton/sodium pumping pyrophosphatases: the last of the primary ion pumps. *Curr Opin Struct Biol.* 2014;27:38-47. Epub 2014/04/29. doi: 10.1016/j.sbi.2014.03.007. PubMed PMID: 24768824.
3. Baykov AA, Cooperman BS, Goldman A, Lahti R. Cytoplasmic inorganic pyrophosphatase. *Prog Mol Subcell Biol.* 1999;23:127-50. Epub 1999/08/17. PubMed PMID: 10448675.
4. Kajander T, Kellosalo J, Goldman A. Inorganic pyrophosphatases: one substrate, three mechanisms. *FEBS Lett.* 2013;587(13):1863-9. Epub 2013/05/21. doi: 10.1016/j.febslet.2013.05.003. PubMed PMID: 23684653.
5. Huang H, Patskovsky Y, Toro R, Farelli JD, Pandya C, Almo SC, et al. Divergence of structure and function in the haloacid dehalogenase enzyme superfamily: *Bacteroides thetaiotaomicron* BT2127 is an inorganic pyrophosphatase. *Biochemistry.* 2011;50(41):8937-49. Epub 2011/09/08. doi: 10.1021/bi201181q. PubMed PMID: 21894910; PubMed Central PMCID: PMCPMC3342813.
6. Kornberg A. Pyrophosphorylases and phosphorylases in biosynthetic reactions. *Advances in enzymology and related subjects of biochemistry.* 1957;18:191-240. PubMed PMID: 13444110.
7. Chen J, Brevet A, Fromant M, Leveque F, Schmitter JM, Blanquet S, et al. Pyrophosphatase is essential for growth of *Escherichia coli*. *Journal of bacteriology.* 1990;172(10):5686-9. PubMed PMID: 2170325; PubMed Central PMCID: PMC526883.
8. Serrano-Bueno G, Hernandez A, Lopez-Lluch G, Perez-Castineira JR, Navas P, Serrano A. Inorganic pyrophosphatase defects lead to cell cycle arrest and autophagic cell death through NAD⁺ depletion in fermenting yeast. *J Biol Chem.* 2013;288(18):13082-92. Epub 2013/03/13. doi: 10.1074/jbc.M112.439349. PubMed PMID: 23479727; PubMed Central PMCID: PMCPMC3642350.
9. Islam MK, Miyoshi T, Yamada M, Tsuji N. Pyrophosphatase of the roundworm *Ascaris suum* plays an essential role in the worm's molting and development. *Infection and immunity.* 2005;73(4):1995-2004. doi: 10.1128/IAI.73.4.1995-2004.2005. PubMed PMID: 15784540; PubMed Central PMCID: PMC1087427.
10. Ko KM, Lee W, Yu JR, Ahnn J. PYP-1, inorganic pyrophosphatase, is required for larval development and intestinal function in *C. elegans*. *FEBS Lett.* 2007;581(28):5445-53. doi: 10.1016/j.febslet.2007.10.047. PubMed PMID: 17981157.
11. Tezuka Y, Okada M, Tada Y, Yamauchi J, Nishigori H, Sanbe A. Regulation of neurite growth by inorganic pyrophosphatase 1 via JNK dephosphorylation. *PLoS One.* 2013;8(4):e61649. doi: 10.1371/journal.pone.0061649. PubMed PMID: 23626709; PubMed Central PMCID: PMC3633968.
12. Polewski MD, Johnson KA, Foster M, Millan JL, Terkeltaub R. Inorganic pyrophosphatase induces type I collagen in osteoblasts. *Bone.* 2010;46(1):81-90. doi: 10.1016/j.bone.2009.08.055. PubMed PMID: 19733704; PubMed Central PMCID: PMC2818162.
13. Panda H, Pandey RS, Debata PR, Supakar PC. Age-dependent differential expression and activity of rat liver cytosolic inorganic pyrophosphatase gene. *Biogerontology.* 2007;8(5):517-25. doi: 10.1007/s10522-007-9094-6. PubMed PMID: 17415680.
14. Kharbhih WJ, Sharma R. Age-dependent increased expression and activity of inorganic pyrophosphatase in the liver of male mice and its further enhancement with short- and long-term dietary restriction. *Biogerontology.* 2014;15(1):81-6. doi: 10.1007/s10522-013-9481-0. PubMed PMID: 24271717.
15. Fairchild TA, Patejunas G. Cloning and expression profile of human inorganic pyrophosphatase. *Biochim Biophys Acta.* 1999;1447(2-3):133-6. PubMed PMID: 10542310.

16. Hamler RL, Zhu K, Buchanan NS, Kreunin P, Kachman MT, Miller FR, et al. A two-dimensional liquid-phase separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification. *Proteomics*. 2004;4(3):562-77. Epub 2004/03/05. doi: 10.1002/pmic.200300606. PubMed PMID: 14997480.
17. Chen G, Gharib TG, Huang CC, Thomas DG, Shedden KA, Taylor JM, et al. Proteomic analysis of lung adenocarcinoma: identification of a highly expressed set of proteins in tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2002;8(7):2298-305. PubMed PMID: 12114434.
18. Kachman MT, Wang H, Schwartz DR, Cho KR, Lubman DM. A 2-D liquid separations/mass mapping method for interlysate comparison of ovarian cancers. *Anal Chem*. 2002;74(8):1779-91. Epub 2002/05/03. doi: 10.1021/ac011159c. PubMed PMID: 11985308.
19. Megger DA, Bracht T, Kohl M, Ahrens M, Naboulsi W, Weber F, et al. Proteomic differences between hepatocellular carcinoma and nontumorous liver tissue investigated by a combined gel-based and label-free quantitative proteomics study. *Mol Cell Proteomics*. 2013;12(7):2006-20. doi: 10.1074/mcp.M113.028027. PubMed PMID: 23462207; PubMed Central PMCID: PMC3708182.
20. Tomonaga T, Matsushita K, Yamaguchi S, Oh-Ishi M, Kodera Y, Maeda T, et al. Identification of altered protein expression and post-translational modifications in primary colorectal cancer by using agarose two-dimensional gel electrophoresis. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2004;10(6):2007-14. PubMed PMID: 15041719.
21. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339(6121):819-23. Epub 2013/01/05. doi: 10.1126/science.1231143. PubMed PMID: 23287718; PubMed Central PMCID: PMC3795411.
22. Wang P, Zhou Y, Mei Q, Zhao J, Huang L, Fu Q. PPA1 regulates tumor malignant potential and clinical outcome of colon adenocarcinoma through JNK pathways. *Oncotarget*. 2017;8(35):58611-24. Epub 2017/09/25. doi: 10.18632/oncotarget.17381. PubMed PMID: 28938583; PubMed Central PMCID: PMC5601679.
23. Li L, Aruna, Luo D, Jin A. Clinical significance and functional validation of inorganic pyrophosphatase in diffuse large B cell lymphoma in humans. *Cytotechnology*. 2018;70(2):641-9. Epub 2017/12/14. doi: 10.1007/s10616-017-0165-5. PubMed PMID: 29234945; PubMed Central PMCID: PMC5851958.
24. Luo D, Wang G, Shen W, Zhao S, Zhou W, Wan L, et al. Clinical significance and functional validation of PPA1 in various tumors. *Cancer Med*. 2016;5(10):2800-12. Epub 2016/10/30. doi: 10.1002/cam4.894. PubMed PMID: 27666431; PubMed Central PMCID: PMC5083733.
25. Lexander H, Palmberg C, Auer G, Hellstrom M, Franzen B, Jornvall H, et al. Proteomic analysis of protein expression in prostate cancer. *Analytical and quantitative cytology and histology / the International Academy of Cytology [and] American Society of Cytology*. 2005;27(5):263-72. PubMed PMID: 16447818.
26. Mishra DR, Chaudhary S, Krishna BM, Mishra SK. Identification of Critical Elements for Regulation of Inorganic Pyrophosphatase (PPA1) in MCF7 Breast Cancer Cells. *PLoS One*. 2015;10(4):e0124864. doi: 10.1371/journal.pone.0124864. PubMed PMID: 25923237; PubMed Central PMCID: PMC4414593.
27. Jeong SH, Ko GH, Cho YH, Lee YJ, Cho BI, Ha WS, et al. Pyrophosphatase overexpression is associated with cell migration, invasion, and poor prognosis in gastric cancer. *Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine*. 2012;33(6):1889-98. doi: 10.1007/s13277-012-0449-5. PubMed PMID: 22797819.
28. Yang Y, Cai J, Yin J, Wang D, Bai Z, Zhang J, et al. Inorganic pyrophosphatase (PPA1) is a negative prognostic marker for human gastric cancer. *Int J Clin Exp Pathol*. 2015;8(10):12482-90. Epub 2016/01/02. PubMed PMID: 26722435; PubMed Central PMCID: PMC4680380.

29. Xu D, Miao Y, Gu X, Wang J, Yu G. Pyrophosphatase 1 expression is associated with future recurrence and overall survival in Chinese patients with intrahepatic cholangiocarcinoma. *Oncol Lett.* 2018;15(5):8095-101. Epub 2018/05/10. doi: 10.3892/ol.2018.8278. PubMed PMID: 29740496; PubMed Central PMCID: PMC5934715.
30. Li H, Xiao N, Li Z, Wang Q. Expression of Inorganic Pyrophosphatase (PPA1) Correlates with Poor Prognosis of Epithelial Ovarian Cancer. *The Tohoku journal of experimental medicine.* 2017;241(2):165-73. doi: 10.1620/tjem.241.165. PubMed PMID: 28202851.
31. Niu H, Zhou W, Xu Y, Yin Z, Shen W, Ye Z, et al. Silencing PPA1 inhibits human epithelial ovarian cancer metastasis by suppressing the Wnt/beta-catenin signaling pathway. *Oncotarget.* 2017;8(44):76266-78. Epub 2017/11/05. doi: 10.18632/oncotarget.19346. PubMed PMID: 29100310; PubMed Central PMCID: PMC5652704.
32. Bodnar M, Luczak M, Bednarek K, Szyberg L, Marszalek A, Grenman R, et al. Proteomic profiling identifies the inorganic pyrophosphatase (PPA1) protein as a potential biomarker of metastasis in laryngeal squamous cell carcinoma. *Amino Acids.* 2016;48(6):1469-76. Epub 2016/03/08. doi: 10.1007/s00726-016-2201-8. PubMed PMID: 26948660; PubMed Central PMCID: PMC54875942.
33. Cerella C, Radogna F, Dicato M, Diederich M. Natural compounds as regulators of the cancer cell metabolism. *Int J Cell Biol.* 2013;2013:639401. Epub 2013/06/14. doi: 10.1155/2013/639401. PubMed PMID: 23762063; PubMed Central PMCID: PMC3670510.
34. Lv W, Banerjee B, Molland KL, Seleem MN, Ghafoor A, Hamed MI, et al. Synthesis of 3-(3-arylpyrrolidin-1-yl)-5-aryl-1,2,4-triazines that have antibacterial activity and also inhibit inorganic pyrophosphatase. *Bioorg Med Chem.* 2014;22(1):406-18. Epub 2013/12/10. doi: 10.1016/j.bmc.2013.11.011. PubMed PMID: 24315189; PubMed Central PMCID: PMC3914758.
35. Kotsikorou E, Song Y, Chan JM, Faelens S, Tovian Z, Broderick E, et al. Bisphosphonate inhibition of the exopolyphosphatase activity of the *Trypanosoma brucei* soluble vacuolar pyrophosphatase. *J Med Chem.* 2005;48(19):6128-39. Epub 2005/09/16. doi: 10.1021/jm058220g. PubMed PMID: 16162013.
36. Pang AH, Garzan A, Larsen MJ, McQuade TJ, Garneau-Tsodikova S, Tsodikov OV. Discovery of Allosteric and Selective Inhibitors of Inorganic Pyrophosphatase from *Mycobacterium tuberculosis*. *ACS Chem Biol.* 2016;11(11):3084-92. Epub 2016/09/14. doi: 10.1021/acscchembio.6b00510. PubMed PMID: 27622287.
37. Salminen T, Kapyla J, Heikinheimo P, Kankare J, Goldman A, Heinonen J, et al. Structure and function analysis of *Escherichia coli* inorganic pyrophosphatase: is a hydroxide ion the key to catalysis? *Biochemistry.* 1995;34(3):782-91. Epub 1995/01/24. doi: 10.1021/bi00003a011. PubMed PMID: 7827037.
38. Samygina VR, Moiseev VM, Rodina EV, Vorobyeva NN, Popov AN, Kurilova SA, et al. Reversible inhibition of *Escherichia coli* inorganic pyrophosphatase by fluoride: trapped catalytic intermediates in cryo-crystallographic studies. *J Mol Biol.* 2007;366(4):1305-17. Epub 2007/01/02. doi: 10.1016/j.jmb.2006.11.082. PubMed PMID: 17196979.
39. Samygina VR, Popov AN, Rodina EV, Vorobyeva NN, Lamzin VS, Polyakov KM, et al. The structures of *Escherichia coli* inorganic pyrophosphatase complexed with Ca(2+) or CaPP(i) at atomic resolution and their mechanistic implications. *J Mol Biol.* 2001;314(3):633-45. Epub 2002/02/16. doi: 10.1006/jmbi.2001.5149. PubMed PMID: 11846572.
40. Heikinheimo P, Lehtonen J, Baykov A, Lahti R, Cooperman BS, Goldman A. The structural basis for pyrophosphatase catalysis. *Structure.* 1996;4(12):1491-508. Epub 1996/12/15. PubMed PMID: 8994974.
41. Oksanen E, Ahonen AK, Tuominen H, Tuominen V, Lahti R, Goldman A, et al. A complete structural description of the catalytic cycle of yeast pyrophosphatase. *Biochemistry.* 2007;46(5):1228-39. Epub 2007/01/31. doi: 10.1021/bi0619977. PubMed PMID: 17260952.

42. Halonen P, Baykov AA, Goldman A, Lahti R, Cooperman BS. Single-turnover kinetics of *Saccharomyces cerevisiae* inorganic pyrophosphatase. *Biochemistry*. 2002;41(40):12025-31. Epub 2002/10/03. doi: 10.1021/bi026018z. PubMed PMID: 12356302.
43. Pohjanjoki P, Fabrichniy IP, Kasho VN, Cooperman BS, Goldman A, Baykov AA, et al. Probing essential water in yeast pyrophosphatase by directed mutagenesis and fluoride inhibition measurements. *J Biol Chem*. 2001;276(1):434-41. Epub 2000/10/14. doi: 10.1074/jbc.M007360200. PubMed PMID: 11031269.
44. Zyryanov AB, Pohjanjoki P, Kasho VN, Shestakov AS, Goldman A, Lahti R, et al. The electrophilic and leaving group phosphates in the catalytic mechanism of yeast pyrophosphatase. *J Biol Chem*. 2001;276(21):17629-34. Epub 2001/03/30. doi: 10.1074/jbc.M100343200. PubMed PMID: 11279052.
45. Heikinheimo P, Tuominen V, Ahonen AK, Teplyakov A, Cooperman BS, Baykov AA, et al. Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase. *Proc Natl Acad Sci U S A*. 2001;98(6):3121-6. Epub 2001/03/15. doi: 10.1073/pnas.061612498. PubMed PMID: 11248042; PubMed Central PMCID: PMCPMC30617.
46. Teplyakov A, Obmolova G, Wilson KS, Ishii K, Kaji H, Samejima T, et al. Crystal structure of inorganic pyrophosphatase from *Thermus thermophilus*. *Protein Sci*. 1994;3(7):1098-107. Epub 1994/07/01. doi: 10.1002/pro.5560030713. PubMed PMID: 7920256; PubMed Central PMCID: PMCPMC2142889.
47. Liu B, Bartlam M, Gao R, Zhou W, Pang H, Liu Y, et al. Crystal structure of the hyperthermophilic inorganic pyrophosphatase from the archaeon *Pyrococcus horikoshii*. *Biophys J*. 2004;86(1 Pt 1):420-7. Epub 2003/12/26. doi: 10.1016/s0006-3495(04)74118-1. PubMed PMID: 14695284; PubMed Central PMCID: PMCPMC1303807.
48. Jamwal A, Yogavel M, Abdin MZ, Jain SK, Sharma A. Structural and Biochemical Characterization of Apicomplexan Inorganic Pyrophosphatases. *Scientific reports*. 2017;7(1):5255. Epub 2017/07/14. doi: 10.1038/s41598-017-05234-y. PubMed PMID: 28701714; PubMed Central PMCID: PMCPMC5507929.
49. Pratt AC, Dewage SW, Pang AH, Biswas T, Barnard-Britson S, Cisneros GA, et al. Structural and computational dissection of the catalytic mechanism of the inorganic pyrophosphatase from *Mycobacterium tuberculosis*. *J Struct Biol*. 2015;192(1):76-87. Epub 2015/08/25. doi: 10.1016/j.jsb.2015.08.010. PubMed PMID: 26296329.
50. Jamwal A, Round AR, Bannwarth L, Venien-Bryan C, Belrhali H, Yogavel M, et al. Structural and Functional Highlights of Vacuolar Soluble Protein 1 from Pathogen *Trypanosoma brucei brucei*. *J Biol Chem*. 2015;290(51):30498-513. Epub 2015/10/24. doi: 10.1074/jbc.M115.674176. PubMed PMID: 26494625; PubMed Central PMCID: PMCPMC4683271.
51. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol*. 2008;3(6):373-82. Epub 2008/06/07. doi: 10.1021/cb800025k. PubMed PMID: 18533659.
52. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Pt 4):271-81. Epub 2011/04/05. doi: 10.1107/s0907444910048675. PubMed PMID: 21460445; PubMed Central PMCID: PMC3069742.
53. Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallogr*. 2004;60(Pt 3):432-8. Epub 2004/03/03. doi: 10.1107/s0907444903028956. PubMed PMID: 14993666.
54. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Echols N, Headd JJ, et al. The Phenix software for automated determination of macromolecular structures. *Methods*. 2011;55(1):94-106. Epub 2011/08/09. doi: 10.1016/j.jymeth.2011.07.005. PubMed PMID: 21821126; PubMed Central PMCID: PMC3193589.

55. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 2010;66(Pt 4):486-501. Epub 2010/04/13. doi: 10.1107/s0907444910007493. PubMed PMID: 20383002; PubMed Central PMCID: PMC2852313.
56. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25(13):1605-12. Epub 2004/07/21. doi: 10.1002/jcc.20084. PubMed PMID: 15264254.
57. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019;47(W1):W636-w41. Epub 2019/04/13. doi: 10.1093/nar/gkz268. PubMed PMID: 30976793; PubMed Central PMCID: PMC6602479.
58. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 1995;8(2):127-34. Epub 1995/02/01. PubMed PMID: 7630882.
59. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31(2):455-61. Epub 2009/06/06. doi: 10.1002/jcc.21334. PubMed PMID: 19499576; PubMed Central PMCID: PMC3041641.
60. Canagarajah BJ, Khokhlatchev A, Cobb MH, Goldsmith EJ. Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell.* 1997;90(5):859-69. Epub 1997/09/23. doi: 10.1016/s0092-8674(00)80351-7. PubMed PMID: 9298898.
61. Tezuka Y, Herai N, Inomata Y, Kagami K, Yamauchi J, Nishigori H, et al. Upregulation of inorganic pyrophosphatase 1 as a JNK phosphatase in hypothyroid embryonic chick cerebellum. *Life Sci.* 2015;128:94-100. Epub 2015/03/10. doi: 10.1016/j.lfs.2015.02.019. PubMed PMID: 25748422.
62. Ip YT, Davis RJ. Signal transduction by the c-Jun N-terminal kinase (JNK)—from inflammation to development. *Current opinion in cell biology.* 1998;10(2):205-19. Epub 1998/04/30. PubMed PMID: 9561845.

Tables

Table 1. Data collection and refinement statistics.

Data collection	Native
Space group	P21 21 21
Unit-cell parameters	
a (Å)	a = 50.27
b (Å)	b = 137.04
c (Å)	c = 216.44
α (°)	90
β (°)	90
γ (°)	90
Resolution (Å)	84.94–2.39
No. of unique reflections	60418
Completeness (%) ^a	99.9 (98.9)
R _{merge} (%) ^b	11.8
$\langle I/\sigma(I) \rangle$	8.1
CC _{1/2}	0.99
Multiplicity	4.5
Anomalous completeness (%)	N/A
Anomalous multiplicity	N/A
Refinement	
Resolution (Å)	84.94–2.39

Data collection	Native
No. of reflections	60418
$R_{\text{work}}/R_{\text{free}}++$ (%) ^c	19.3/25.1
No. of atoms	
Protein	1135
Water	526
B factors (\AA^2)	
Protein	39.19
Water	38.6
R.m.s.d., bonds (\AA)	0.008
R.m.s.d., angles ($^\circ$)	1.248
Ramachandran plot	
Favored (%)	96.6
Outliers (%)	0.7

^aValues in parentheses are for the highest-resolution shell

^b $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $\langle I(hkl) \rangle$ is the average intensity of reflection hkl .

^c $R_{\text{work}} = \frac{\sum_{hkl} (|F_{\text{obs}}| - |F_{\text{calc}}|)}{\sum_{hkl} |F_{\text{obs}}|}$, where F_{obs} and F_{calc} are the observed and calculated structure factors respectively. R_{free} is calculated as for R_{work} but only use a randomly selected subset of data (6%) which were excluded from refinement.

Figures

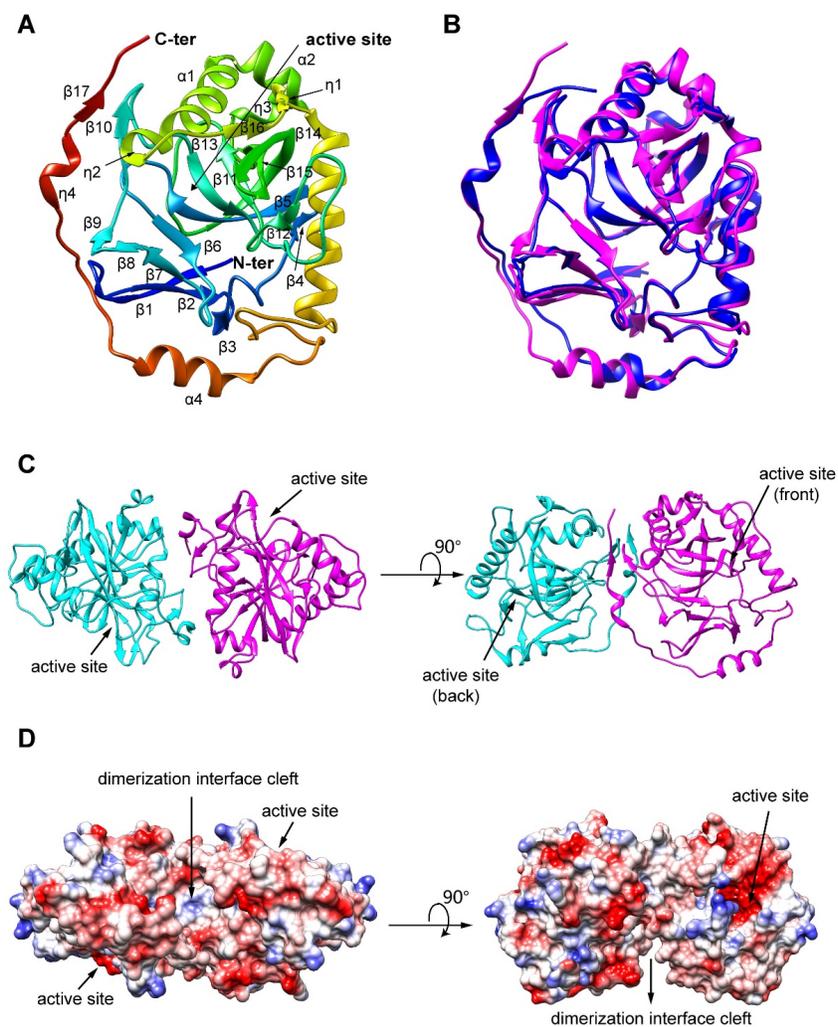


Figure 2. Overall structure of the human PPA1 protein. A) The single domain monomeric structure. B) Superimposition of the monomeric structures of human PPA1 (magenta) and yeast PPase (blue, PDB code 1E6A). C) The dimeric structure of human PPA1 in two different viewing angles. The two monomers are colored cyan and magenta respectively. D) Electrostatic surface representation of the human PPA1 homodimeric structures. The surface areas are colored according to electrostatic potentials, from blue for the most positive, to white, to red for the most negative regions.

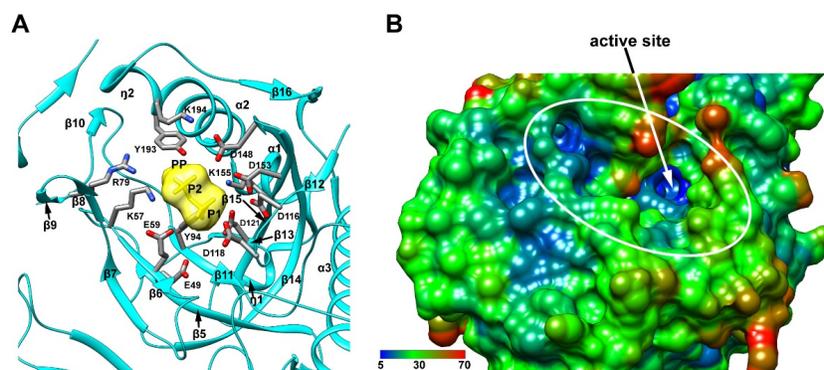


Figure 5. The putative active site of human PPA1. A) The active site with a modeled pyrophosphate (rendered by transparent surface mode in yellow). The side-chains are shown in stick mode (colored by atom type) for residues corresponding to the active site catalytic residues in yeast PPase. B) The active site pocket on the surface of PPA1. The surface is colored according to B factor values of the atoms.

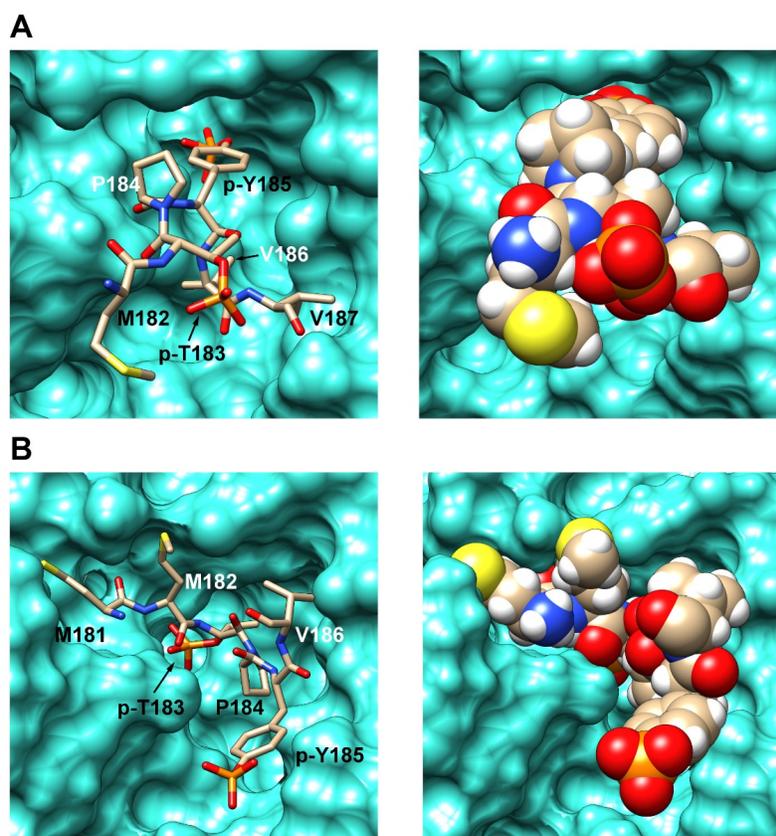


Figure 6. Structural models of the human PPA1 active site with potential peptide substrates containing a phosphotyrosine and a phosphothreonine. A) A phosphotyrosine (p-Y185), within the context of a pentapeptide 182-MT^PPYPV, is bound at the active site. B) A phosphothreonine (p-T183), within the context of a hexapeptide 181-MMT^PPYPV, is bound at the active site. The phosphorylated

peptides are rendered in stick mode and space-filling mode in the two panels of the figures.