Crystal Structure of Bovine Heart Phosphotyrosyl Phosphatase at 2.2-Å Resolution[†,‡]

Marie Zhang,[§] Robert L. Van Etten,*[,§] and Cynthia V. Stauffacher*[,‖]

Departments of Chemistry and Biological Sciences, Purdue University, West Lafayette, Indiana 47907

Received May 3, 1994; Revised Manuscript Received June 21, 1994[⊗]

ABSTRACT: The first X-ray crystallographic structure of a member of the class of low molecular weight (M_r 18 000) phosphotyrosyl phosphatases is presented. Bovine heart phosphotyrosyl phosphatase (BHPTP) exemplifies this class and is highly homologous (94% sequence identity) to an isoenzyme known as red cell acid phosphatase that is present throughout human tissues. The high-resolution (2.2-Å) crystal structure of BHPTP shows that the enzyme consists of a four-strand central parallel β sheet with α helices packed on both sides in a manner characteristic of a Rossmann fold. A bound phosphate ion defines the active site location in a loop of the first $\beta\alpha\beta$ motif at the C-terminus of the β sheet. The location and enzymatic significance of the residues in the characteristic low molecular weight PTPase active site motif, including the essential arginine (Arg 18) and nucleophilic cysteine (Cys 12), are described. The functional role of a histidine (His 72) suggested previously to be near the active site is defined in the structure, as well as a potential proton donor for the leaving group in the tyrosyl phosphate cleavage. Surface maps of BHPTP define a hydrophobic crevice suitable for phosphotyrosyl peptide binding. Comparison of the BHPTP structure to the related, but structurally distinct enzyme PTP1B is made, illustrating the unique way this smallest of these phosphatases has formed the phosphotyrosine active site.

Many cellular processes are regulated through a series of phosphorylations and dephosphorylations of proteins catalyzed by ubiquitous groups of kinases and phosphatases. A great deal is now known about the kinase families, particularly the serine/threonine kinases, which appear to have a central role in many of the phosphorylation cascades that control, among other pathways, the rhythms of the cell cycle and signal transduction. Phosphorylation of tyrosines appears to effect a fine control in these regulatory mechanisms, often in combination with phosphorylation of serine or threonine on the same substrate (Edelman *et al.*, 1987; Cohen, 1989; Fisher *et al.*, 1991). There are a number of families of phosphotyrosyl phosphatases (PTPases),[1] including receptor-like molecules having an extracellular domain, a membrane-spanning region, and one or two intracellular domains (Charbonneau & Tonks, 1992; Zhang & Dixon, 1994). Soluble PTPases also exist in the cell, both in the cytoplasm and in the nucleus. Many of these enzymes have molecular weights in the 30 000–50 000 range, are similar in size to the catalytic domains of the membrane-bound enzymes, and in many cases share a high level of sequence identity with them. There is a distinct class of low molecular weight PTPases (M_r near 18 000) found to be ubiquitous in mammalian cells (Heinrikson, 1969; Rehkop & Van Etten, 1975; Chernoff & Li, 1985; Oakada *et al.*, 1986; Waheed *et al.*, 1988; Zhang & Van Etten, 1990). Except for a CXXXXXR motif at the catalytic site, the low molecular weight phosphotyrosyl phosphatases show no apparent sequence homology to the rest of the PTPases. Bovine heart PTPase is a prototypical example of the small soluble PTPases that contain a single catalytic domain and have a molecular weight of approximately 18 000.

Among the earliest descriptions of the low molecular weight phosphotyrosyl protein phosphatases from mammalian organs were the studies of the bovine enzyme by Heinrikson (1969) and Baldijao *et al.* (1975). Subsequent purification to homogeneity of the bovine liver, human liver, and human placental enzymes opened the way to more detailed structure–function investigations (Lawrence & Van Etten, 1981; Taga & Van Etten, 1982; Waheed *et al.*, 1988). More recently, cDNA cloning and sequencing results (Wo *et al.*, 1992a,b) have definitively demonstrated that the low molecular weight enzyme which was isolated and purified to homogeneity from human placenta and liver is in fact the same protein as "red cell" acid phosphatase, an enzyme that continues to find major use as a phenotypic genetic marker [cf. Sokal *et al.* (1991)]. Moreover, mRNA blotting experiments showed that the enzyme, far from being unique to red cells, is expressed in all major human organs including brain (Wo *et al.*, 1992a). These proteins are also highly homologous. For example, the primary sequence of the bovine enzyme is 94% identical to one of the two major human isoenzymes.

Although the exact biological function of these small cytoplasmic PTPases is just now being investigated, these enzymes have been shown to readily hydrolyze a number of physiological and nonphysiological substrates. Studies of the homogeneous human placental enzyme confirmed an earlier indication that these enzymes were active phosphotyrosyl protein phosphatases (Chernoff & Li, 1985; Waheed *et al.*, 1988). Initially, the low molecular weight PTPases were characterized as acid phosphatases because of the common use of *p*-nitrophenyl phosphate (pNPP) as a substrate for the estimation of the pH "optimum". Because that substrate has

[†] This work was supported by USDHHS Research Grant GM27003, by USDHHS core grants to the Purdue Cancer Center and AIDS Center, and by a grant from the Lucille P. Markey Foundation for the expansion of structural biology in the Department of Biological Sciences.

[‡] The atomic coordinates for BHPTP structure have been deposited with the Brookhaven Protein Data Bank under the filename 1PNT.

* Authors to whom correspondence may be addressed.

[§] Department of Chemistry.

‖ Department of Biological Sciences.

[⊗] Abstract published in *Advance ACS Abstracts,* August 1, 1994.

[1] Abbreviations: BHPTP, bovine heart phosphotyrosyl phosphatase; EMTS, ethyl mercury thiosalicylate; FOM, figure of merit; HCPTP-A and HCPTP-B, isoenzymes of human low molecular weight phosphotyrosyl phosphatase; MIR, multiple isomorphous replacement; pNPP, *p*-nitrophenyl phosphate; PTPase, phosphotyrosyl phosphatase.

0006-2960/94/0433-11097$04.50/0

an unusually low pK_{a2}, the pH dependence was distorted toward an acidic maximum. In fact, a more careful examination of the pH dependence of the enzyme showed that V_{max} was almost completely flat between pH 4 and pH 7 (Lawrence & Van Etten, 1981). This behavior was subsequently confirmed for hydrolysis by the homogeneous human placental enzyme, employing substrates such as phosphotyrosine and red cell band 3 protein that had been phosphorylated on tyrosines 8 and 22 (Waheed *et al.*, 1988). In addition to its activity against substrates such as band 3 (Boivin & Galand, 1986), the low molecular weight human PTPase also dephosphorylates tyrosine residues of angiotensin and tyrosine kinase p40 (Waheed *et al.*, 1988), while the bovine enzyme dephosphorylates autophosphorylated EGF receptor *in vitro* with a K_m in the nanomolar region (Ramponi *et al.*, 1989).

Analysis of the primary sequence of the low molecular weight protein tyrosyl phosphatases identified a conserved motif characteristic of this class of phosphatases that is related to, but distinct from, the motif known for the high molecular weight cytoplasmic and membrane-bound enzymes. In the high molecular weight PTPases, the invariant signature motif is HCXAGXGR(S/T) (Zhang & Dixon, 1994) and is universally found in the C-terminal third of the molecular sequence. In the low molecular weight PTPases, a similar invariant motif, XCXXXXCRS,[2] is found, but in these enzymes, the motif is located very close to the N-terminus of the protein. The relative placement of the first cysteine and the arginine and serine residues in these two motifs is similar, although the required histidine and the glycine-rich nature of the loop in the high molecular weight species are not preserved in the low molecular weight PTPases. This similarity in the conserved sequence motifs not only identifies these proteins as PTPases but also suggests a common mechanism for the dephosphorylation of tyrosines in these otherwise very different molecules.

Although no three-dimensional structure has been described for any member of this important class of low molecular weight PTPases, some structural relationships may be inferred from the extensive kinetic and mechanistic information that is available. Early studies demonstrated the presence of numerous highly reactive cysteines, at least one of which appeared to be essential for activity (Laidler *et al.*, 1982). Following inactivation with iodoacetate, a labeled CysArg peptide was isolated from the active site of the bovine enzyme (Lawrence & Van Etten, 1981). Later work showed that in fact two cysteines, at positions 12 and 17, were selectively modified by iodoacetate (Camici *et al.*, 1989). These two cysteines are part of the low molecular weight PTPase signature motif and are followed in the sequence by the conserved arginine. Consistent with the presence of arginine at the active site, the bovine enzyme is inactivated by phenylglyoxal and by cyclohexanedione (Zhang *et al.*, 1992).

Inactivation of the enzyme by photooxidation as well as by reaction with diethyl pyrocarbonate raised the possibility of a histidine as well as cysteine in the active site region (Lawrence & Van Etten, 1981). This hypothesis received strong confirmation from recent NMR experiments involving the titration of the enzyme with vanadate ion (Zhou *et al.*, 1993). BHPTP contains only two possibilities for this proposed active site residue, His 66 and His 72. The spectral position of His 72 (but not that of His 66) was significantly shifted upon titration of the enzyme with the strong competitive inhibitor vanadate ion. This ion binds tightly to numerous phosphatases, either because of its resemblance to the trigonal-bipyramidal transition state for displacement reactions catalyzed by these enzymes or because of the ability of the early transition metal oxoanions to readily replace their oxygen atoms by a variety of other oxygen, nitrogen, or sulfur ligands (Van Etten *et al.*, 1974). Although a recent publication claimed that "both His 66 and His 72 participate in the active site of the enzyme" (Chiarugi *et al.*, 1994), such a conclusion is in sharp contrast to spectral and site-directed mutagenesis experiments which show that none of the histidines of this protein are essential for catalytic activity (Davis *et al.*, 1994a).

The low molecular weight PTPases have been shown to catalyze a stereospecific phospho transfer from phenyl [^{16}O,^{17}O,^{18}O]phosphate to propanediol via a double displacement mechanism involving an $S_N2(P)$ reaction, with the formation of a phosphoenzyme intermediate (Saini *et al.*, 1981). Pre-steady-state, burst titration kinetics have been demonstrated (Zhang & Van Etten, 1991a). Consistent with earlier suggestions (Lawrence & Van Etten, 1981), the enzymatic nucleophile has been established to be cysteine by stoichiometric trapping of the phosphoenzyme and identification of the phosphocysteinyl linkage by NMR (Wo *et al.*, 1992b). Site-directed mutagenesis studies on the recombinant, nonfusion protein show that Cys 12 is the nucleophile while Cys 17 can be mutated to Ala with only minor kinetic consequences (Davis *et al.*, 1994b). Besides Cys 12 and Cys 17, there are six other cysteine residues present in this protein, all in the free sulfhydryl form (Laidler *et al.*, 1982). As with Cys 17, none of them are essential for activity, although mutagenesis leads to minor activity differences, suggesting that several of them may be relatively near the active site region (Davis *et al.*, 1994a). The enzyme is inactivated by phenylarsine oxide, suggesting that at least two of the sulfhydryl groups are in close proximity (Zhang *et al.*, 1992).

The bovine enzyme and two human isoenzymes have been cloned and overexpressed in *Escherichia coli* (Wo *et al.*, 1992a,b). The recombinant enzyme differs from the enzyme isolated from tissue only in that the N-terminal alanine is not acetylated. Indeed, expression of the bovine enzyme as a fusion protein that had seven additional amino acids resulted in a protein with high catalytic activity, indicating that the N-terminal region can be somewhat altered without marked consequences to the protein structure (Wo *et al.*, 1992b).

In order to extend our studies of the mechanism of this protein tyrosyl phosphatase, we have crystallized the bovine heart enzyme (Zhang *et al.*, 1994) and here present the X-ray crystallographic structure determination to 2.2 Å. With this structure we can now not only describe the fold of the molecule and how its structure relates to the larger enzymes but can also examine the relationship of the conserved sequence motif to the structure and activity of the enzyme. Due to the presence of a phosphate ion at the active site we can now also define the residues that are important for substrate binding and orientation and for forming the active site configuration. Finally, we are able to suggest detailed roles for the residues identified by the extensive biochemical studies of the mechanism of this molecule as being important to the enzyme activity.

EXPERIMENTAL PROCEDURES

Structure Determination. The structure of BHPTP was determined by the multiple isomorphous replacement (MIR) method (Table 1). All data were collected on a Siemens

[2] The segment corresponding to the invariant motif in all low molecular weight PTPases of known sequence is actually VCLGNICRS. This is represented in the text as XCXXXXCRS for direct comparison of similar elements in the motif to those in the high molecular weight invariant sequence.

Table 1: Structure Determination[a]

data set	resolution (Å)	completeness (%)	redundancy	R_{sym}[b]	R_{ISO}[c]	sites (no.)	phasing power[d]
native	2.2	90	3.1	0.060			
K_2PtCl_4(a)	3.0	98	3.0	0.054	0.105	2	0.6
K_2PtCl_4(b)	3.0	94	1.4	0.033	0.098	1	1.2
$K_2Pt(NO_2)_4$	3.0	96	2.1	0.043	0.080	4	0.6
EMTS	3.0	100	3.1	0.051	0.170	2	2.3

[a] Crystals of BHPTP were obtained as previously described (Zhang *et al.*, 1994). The crystals belong to space group $C2$ with unit cell dimensions of a = 95.3 Å, b = 43.3 Å, c = 41.2 Å, and β = 113.5° with one molecule per asymmetric unit (V_m = 2.17 Å^3/Da). Diffraction data were collected at room temperature on a Siemens analytical multiwire area detector with monochromated Cu Kα X-rays produced from an Elliott GX20 rotating anode X-ray generator operated at 35 kV and 40 mA. A crystal to detector distance of 12 cm was used with 0.15° oscillations per frame for the 180–250° sweeps of data taken from each crystal. Complete diffraction data to 2.2-Å resolution were collected from two crystals of native protein. Data to 3.0 Å were collected from one crystal of each of four heavy atom derivatives. Diffraction data were indexed, integrated, merged, and scaled with the XDS program (Kabsch, 1988). Heavy atom derivatives were produced by transferring single BHPTP crystals into wells containing synthetic mother liquor (15% PEG 4K, 10% 2-propanol, and 5% glycerol, in 100 mM HEPES, pH 7.5) and the heavy atom compound. The K_2PtCl_4 derivative gave two different sets of binding sites due to slightly different soaking conditions, one with two binding sites of nearly equal but low occupancy [K_2PtCl_4 (a)] and the other with a single binding site of high occupancy [K_2PtCl_4 (b)]. Soaking conditions for these two derivative data sets were 10 mM K_2PtCl_4 for 22 h for data set a and 12.5 mM K_2PtCl_4 for 19 h for data set b. Soaking conditions for the initial platinum and mercury derivatives were 20 mM $K_2Pt(NO_2)_4$ for 24 h and for ethylmercury thiosalicylate (EMTS) 0.01 mM for 5 days. Difference Patterson synthesis revealed the two binding sites of EMTS and the single binding site in K_2PtCl_4. Phases were calculated from these two derivatives (FOM = 0.62 for data to 3 Å) and used to locate the binding sites of other heavy atom derivatives by difference Fourier analysis. MIR phase determination and refinement were carried out with MLPHARE from the CCP4 program suite (CCP4, 1979). The figure of merit (FOM) for the refined phases was 0.68 for all reflections to 3.0-Å resolution. With the help of a "boned" electron density map obtained using program Abone and with the graphics program O (Jones *et al.*, 1991), the polypeptide backbone was traced as an initial polyalanine model, although many side chains could already be clearly seen. In this fit 149 of the 157 residues could be located. The electron density for residues 1–4 at the N-terminus was particularly poor. This electron density map was improved by combining the polyalanine model phases with the MIR phases and solvent flattening using the program SFALL from the CCP4 package. The whole polypeptide chain was then fit into the electron density with program O using the main-chain database and side-chain rotamer libraries provided. The initial model had an R factor of 0.453 for all data to 3.0 Å. [b] R_{sym} is the R factor on intensities for multiple observation of symmetry-related reflections. [c] R_{ISO} is the R factor on structure factors for all reflections in common with the native data. [d] Phasing power is the ratio of the mean amplitude of heavy atom structure factor to the rms lack of closure error.

Table 2: Structure Refinement of BHPTP[a]

refinement program	resolution (Å)	total atoms/ waters	rms bond length (Å)	rms bond angle (deg)	% in most favored region	R factor (%)	R_{free} (%)
TNT	20.0–3.0	1255/0	0.011	2.69		22.6	
	20.0–2.2	1255/90	0.011	2.34	92.4	17.2	
XPLOR	8.0–3.0	1255/0	0.010	1.63		21.4	
	6.0–2.2	1255/74	0.007	1.38	90.3	19.9	29.1

[a] The structure of BHPTP was refined at 3.0 Å from the initial model coordinates with one round of simulated annealing. These improved coordinates were then further refined using both the TNT (Tronrud *et al.*, 1987) and XPLOR (Brunger *et al.*, 1987) refinement procedures. Simulated annealing followed by conventional positional and individual B-factor refinement was done in XPLOR, resulting in an R factor of 0.214. TNT refinement produced a very similar model with an R factor of 0.226; the overall phase difference between these two models was 32.83°. At this stage, the resolution was extended from 3.0 to 2.2 Å using the XPLOR model, adding all data in one step. These additional data raised the overall R factor to 0.352. The model was refit in the higher resolution map and then refined independently again with both TNT and XPLOR. Three cycles of manual rebuilding and refinement were done in both refinement procedures. In the TNT refinement increasing numbers of water molecules were added in each cycle. In the final TNT refinement, 90 water molecules as well as the active site PO_4^{3-} ion were included. For the XPLOR refinement, the same 90 water molecules which appeared in the TNT refined map were added; 74 of those waters survived refinement and no additional ones appeared. The Ramachandran plots for TNT and XPLOR refinement all have greater than 90% of the residues in the most favored regions and show no outliers. A comparison of these two refined structures showed that both models were very similar (rms deviation in Cα positions of 0.166 Å) with substantive differences in only two residues at the end of a loop at the opposite end of the molecule from the active site. Mean phase difference between models at 3.0 Å is 32.83° and at 2.2 Å is 33.95°.

analytical multiwire area detector and processed with the program XDS (Kabsch, 1988). Model building was done with the graphics program O (Jones *et al.*, 1991). Initially, several related platinum derivatives were produced by varying the soaking conditions and the strength of the platinum ligands. Although mercury compounds were not considered good choices for derivatives, since the enzyme has eight reactive cysteines, a strong two-site ethylmercury thiosalicylate (EMTS) derivative was found which greatly improved the phases. Tungstate, an active site inhibitor, was also tried as a derivative but in this crystal form could not access the active site and was found at a disordered site near the crystallographic 2-fold. The initial electron density map, based on 3.0-Å MIR phases from four derivatives (Table 1), was clear enough to trace the entire backbone with the exception of the first four residues. One round of phase combination and solvent flattening produced a map which allowed all amino acids to be built into the electron density (initial R factor of 0.453). One round of simulated annealing with XPLOR produced a corrected model which was refined simultaneously with TNT and XPLOR (Table 2). Iterative model rebuildings using $2F_o - F_c$ and $F_o - F_c$ maps were carried out during the refinement. The final refinement in TNT and XPLOR produced very similar structures with rms deviations of bond lengths and angles comparable to other well-refined structures. These refinements also showed that the first four residues at the N-terminus have relatively large temperature factors (>50 Å^2) and are assumed to be very flexible. The final R factors obtained from the TNT and XPLOR refinements are 0.172 and 0.199, respectively. The R_{free} calculation from XPLOR with 431 randomly chosen reflections (6% of total reflections) dropped from an initial value of 0.476 to a final value of 0.291, indicating real improvements in the model. The final model used is from the TNT refinement and possesses overall good stereochemistry, with 92.4% of the residues in the most favored region and no outliers in a Ramachandran plot. A portion of the $2F_o - F_c$ electron density map around the active site of the enzyme is shown in Figure 1.

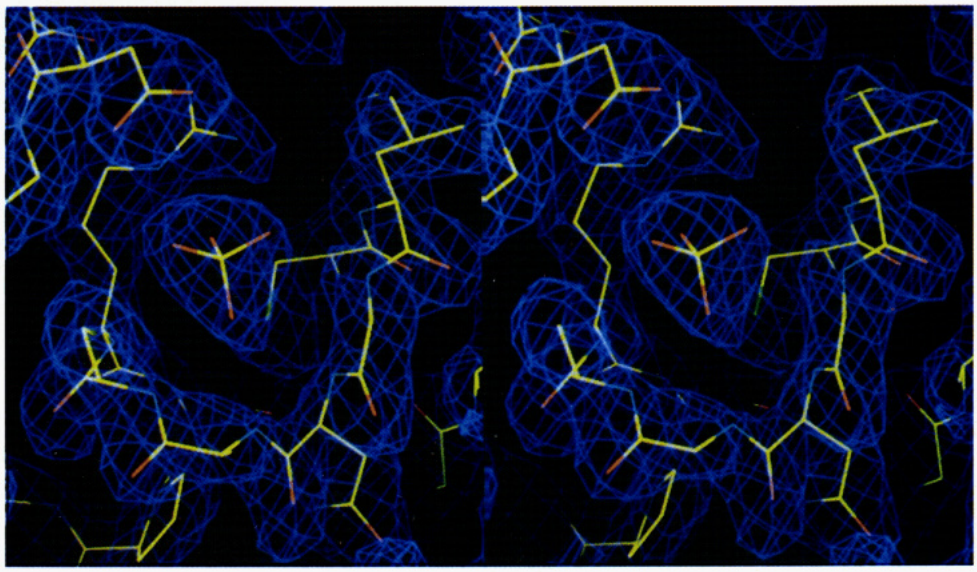

FIGURE 1: Electron density from the refined coordinates superimposed on the structure of BHPTP. A portion of the 2.2-Å $2F_o - F_c$ map (F_c and α_{calc} from TNT refined coordinates) including the active site loop and phosphate ion is shown in the stereoview above. Map contours shown are at 1σ value of electron density. Display graphics are from the program O (Jones *et al.*, 1991).

RESULTS AND DISCUSSION

Structure Description. BHPTP is a compact, oblate ellipsoidal structure of relative dimensions 34 × 42 × 48 Å characterized by a highly twisted central β sheet with α helices packed on both sides (Figure 2). This overall structure has the character of a classic dinucleotide binding or Rossmann fold, with two clear right-handed $\beta\alpha\beta$ motifs contributing to the central four-strand parallel β sheet (Figure 3). One such motif running from residue 6 to residue 45 contains the PTPase characteristic sequence including the cysteine identified as the catalytic nucleophile. This structure, which lies at the center of the molecule (Figure 2a), forms a typical phosphate binding loop at the C-terminus of the β sheet and the N-terminus of $\alpha 1$. The other $\beta\alpha\beta$ motif encompasses residues 87–115 and completes the Rossmann fold and one side of the BHPTP structure. Long loops connect these regions of secondary structure, including an extended loop region between residue 46 and residue 80 which folds over one edge of the β sheet and is tacked down at the opposite end of the molecule by a short helix. A C-terminal loop and a helix (residues 135–157) complete the structure, closing the third side of the molecule. The small helix of the extended loop and the helix of the N-terminal $\beta\alpha\beta$ motif pack against the C-terminal six-turn α helix on one side of the sheet in a two-on-one helix structure with both shorter helices crossing at near identical angles and separated by 10 Å (Figure 2b). Four residues at the N-terminus of the molecule extend from the structure in a highly flexible segment.

The BHPTP molecule contains eight conserved cysteine residues, all of which are in the free sulfhydryl form (Figure 4). Although no disulfide bond is found, some of the cysteines (Cys 145, 148, and 149) form a cluster just beneath the active site in the hydrophobic core of the structure. Cys 90 and Cys 12 appear to be close enough to form a disulfide bond, but their side chains point away from each other, leaving both of their SH groups free. Although partially buried, all the cysteine residues in the structure can easily be reached by cysteine-derivatizing compounds, explaining the earlier chemical modification data (Laidler *et al.*, 1982). Next to the cysteine cluster, a group of aromatic residues (Phe 26, Trp 39, Tyr 87, Tyr 119, Phe 152) are located within 5 Å of each other, forming the central hydrophobic core. These groups of residues appear to serve the general function of stabilizing the overall BHPTP structure.

Structure of the Active Site of BHPTP. The refined structure of BHPTP explains all the electron density in the cell except for a strong spherical density in the active site loop. The stock protein solution used for crystallization contained about 2 mM inorganic phosphate, which is a competitive inhibitor of BHPTP ($K_i = 2$ mM). The strong electron density located at the active site of the BHPTP structure was therefore modeled as an inorganic phosphate ion. The active site of BHPTP was identified not only by the presence of this phosphate ion but also by the location of the nucleophilic cysteine residue (Cys 12) and the essential arginine residue (Arg 18). These residues mark the ends of the characteristic XCXXXXCRS sequence motif which exists near the N-terminus of the molecule in low molecular weight PTPases. In the BHPTP structure, the characteristic motif exists as a loop (residues 12–19) that connects the first β strand of the central sheet with the first turn of α helix in the $\beta\alpha\beta$ motif. The phosphate ion sits at the center of this loop nestled between the side chains of Cys 12 and Arg 18. This active site is located at the C-terminal end of the central sheet in an open cleft and is relatively exposed to the solvent. The flexible residues at the N-terminus extend out of the globular structure at the opposite end of the molecule, explaining why small modifications at the N-terminus do not affect the enzyme activity.

A surface representation of the BHPTP structure (Figure 5) shows the overall nature of the crevice that contains the active site. The conserved sequence loop and the side chain of Arg 18 completely encircle the phosphate ion position, forming a deep hole in the surface of the molecule. Three other loop segments (residues 46–58, 90–95, and 128–132) form a larger ring surrounding the active site which can be seen as the next circular terrace on the molecular surface. The side chains of one tryptophan (Trp 49) and two tyrosines (Tyr 131 and Tyr 132) extend out from either side of the active site, like two claws on a lobster. This rather unusual grouping of these two types of aromatic residues on the surface of the molecule contributes to a large hydrophobic patch on the structure surface. These aromatic claws may also determine the type of specificity motif toward phosphate substrates. They

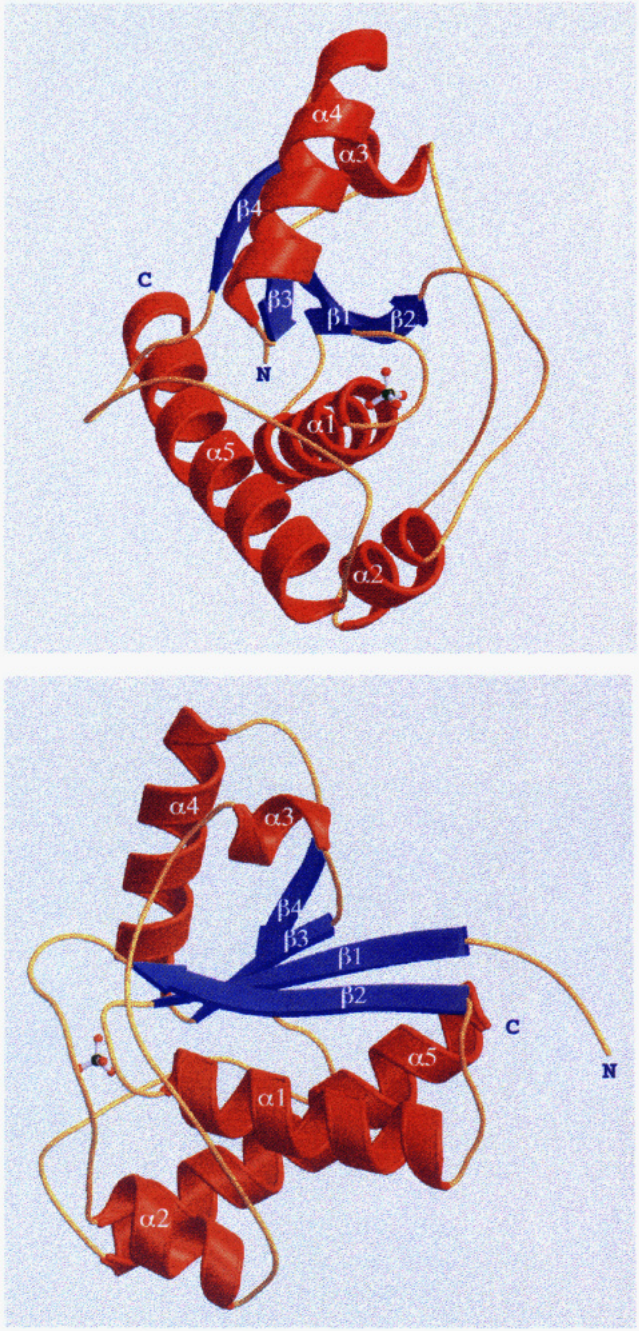

FIGURE 2: Ribbon diagrams of the fold of bovine heart phosphotyrosyl phosphatase (BHPTP). Two views of the molecule separated by a 90° rotation are shown in panel a (top) and panel b (bottom). The two-on-one crossing helical motif lies toward the bottom of these views. The $\beta\alpha\beta$ motif containing the active site loop (residues 6–45) can be seen at the center of (a) and at the front edge of the sheet in (b). The second $\beta\alpha\beta$ motif (residues 85–115) completes the Rossmann fold at the top left of (a) and back top of (b). The extended loop (residues 46–80), containing the variable sequences of the low molecular weight PTPases, folds over one side of the molecule crossing the active site. A final long loop (residues 116–134) and C-terminal helix (residues 135–157) closes the structure. Ribbon diagrams were produced with the programs MOLSCRIPT (Kraulis, 1991) and RASTER3D (E. A. Merritt, and M. Murphy, unpublished results).

provide a site for packing against the tyrosine portion of the phosphorylated tyrosine peptide and might also provide some selectivity against phosphorylated serine and theronine peptides. The open nature of the crevice in which the active site is found may also explain the capacity of this enzyme to dephosphorylate a number of substrates *in vitro*.

Phosphate Binding in the BHPTP Active Site and Role of the Conserved Arginine. The details of the phosphate ion

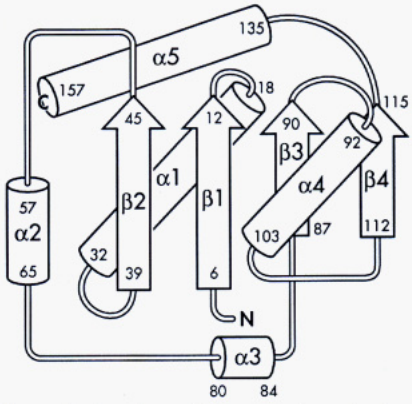

FIGURE 3: Schematic drawing of the secondary structure elements of the BHPTP structure. Sequence numbers for amino acids at the beginning and end of each structural element are indicated. The two $\beta\alpha\beta$ motifs comprising the Rossmann fold form the core of the structure. The binding site for the phosphate ion in BHPTP is formed by the loop in the central $\beta\alpha\beta$ motif of the diagram between Cys 12 at the end of β1 and Arg 18 at the beginning of α1.

binding site can be clearly seen in the BHPTP structure (Figure 6). The conserved motif XCXXXXCRS of the low molecular weight PTPase family forms a loop typical of nucleotide binding proteins (Schultz, 1992). In these structures, the initial $\beta\alpha\beta$ motif binds the phosphate in a loop which starts at the end of the first β strand and runs into the following α helix, whose N-terminal backbone NHs are then well positioned for phosphate binding. In the BHPTP structure, hydrogen bonds from main-chain NHs of residues 13–18 in this loop serve to position the phosphate. At the beginning of the loop, main-chain NHs of Leu 13 and Gly 14 contact the first phosphate oxygen, while Asn 15, Ile 16, and Cys 17 donate main-chain NH hydrogen bonds to position the second oxygen. Arg 18 provides a main-chain and an NϵH hydrogen bond to the third phosphate oxygen and a terminal NH hydrogen bond to the first of these three phosphate oxygens, closing the circle of hydrogen bonds which positions the phosphate in the active site. BHPTP does not have the typical sequence motif GXGXXG for phosphate binding that is observed in many other proteins, including dehydrogenases and the protein kinases (Schultz, 1992), but the present sequence folds into a peptide backbone arrangement which appears to provide equivalent phosphate binding ability. Although there is only a single Gly in the loop, the main chain can assume this configuration due to stabilizing interactions of the active site loop side chains with neighboring conserved residues (Figure 6b and below).

The conserved arginine residue side chain is also stabilized by an interaction with neighboring residues. Arg 18 is held in position by an aspartic acid residue (Asp 92) which interacts with the second terminal NH group on Arg 18 (Figure 6b). The intimate involvement of Arg 18 in phosphate binding in BHPTP is in good agreement with mutagenesis results in which mutation of Arg 18 to Ala completely abolishes the enzyme activity (Davis *et al.*, 1994b). Thus the conserved Arg of the sequence motif is found in a position where it can serve a function in substrate recognition, orientation, and stabilization.

Location and Role of the Active Site Cysteines. The two cysteines of the active site sequence motif (Cys 12 and Cys 17) can now also be assigned functions. Of the eight conserved cysteine residues, early chemical modification studies of BHPTP demonstrated that at least one of them is located at the active site (Zhang *et al.*, 1992). Later experiments with a stoichiometrically trapped, covalent phosphoenzyme intermediate using ^{31}P NMR, in conjunction with cysteine site-

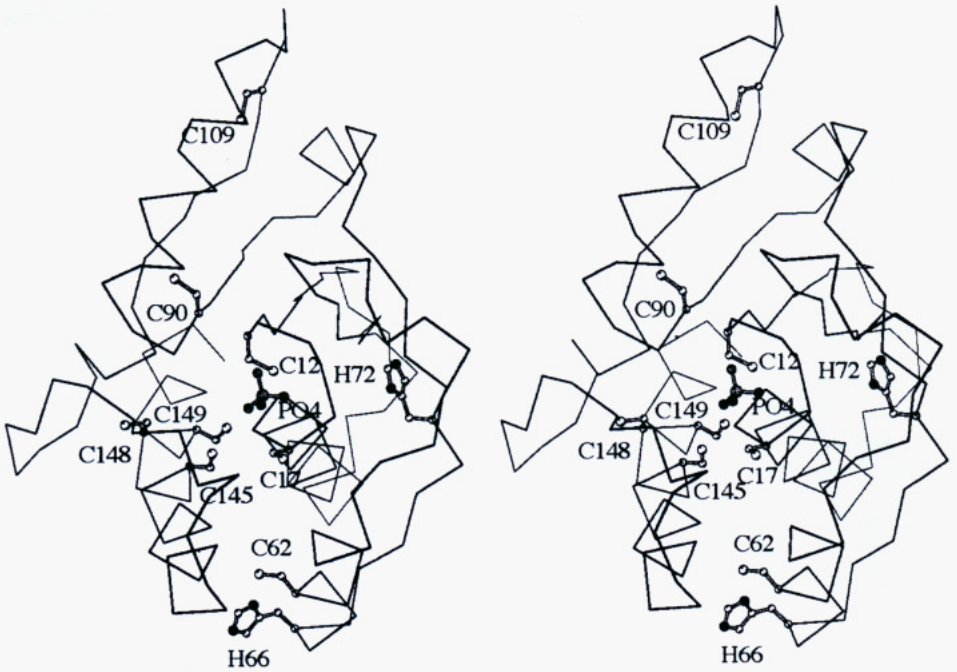

FIGURE 4: Stereoview of the BHPTP molecule showing the locations of the conserved cysteine and histidine residues. The orientation of this figure duplicates that shown in the ribbon drawing of Figure 2a. A phosphate ion is included to mark the active site position between cysteine 12 and cysteine 17. The two conserved histidines in the structure, one of which is implicated in the activity of the enzyme, are located at the right and bottom.

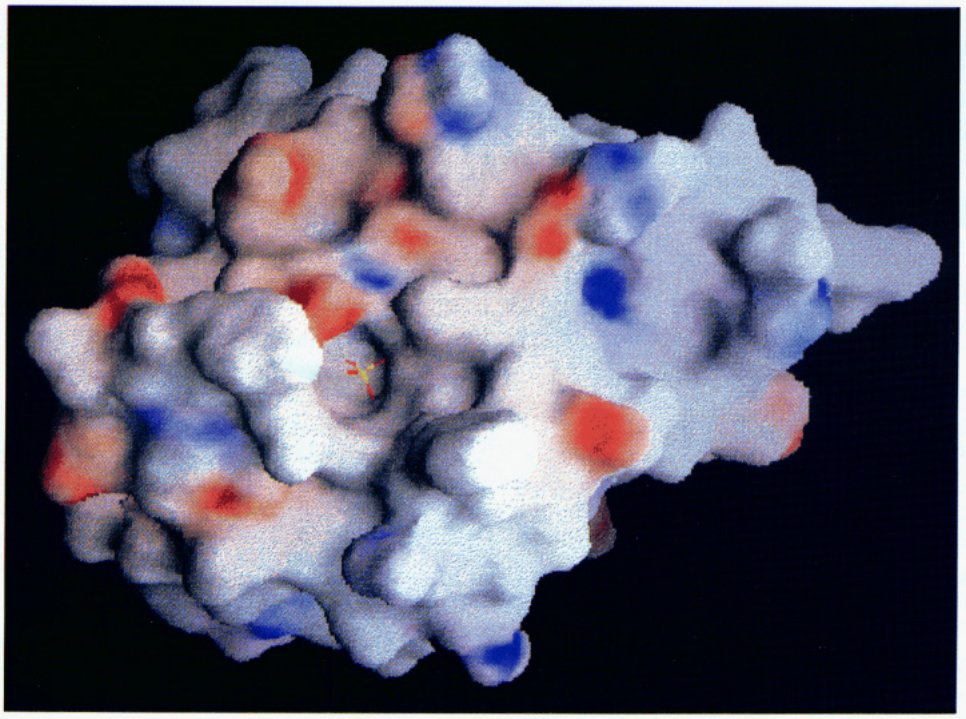

FIGURE 5: A smoothed surface representation of the BHPTP structure. The surface structure shown is rotated by approximately 90° from the orientation shown in Figure 2a. A ball and stick model of the phosphate ion marks the active site loop. The inner rim of the deep hole in which the phosphate sits is formed by the conserved motif loop (residues 12–18) and side chain of the essential arginine 18. The next circular terrace is formed by three stretches of residues: one from the extended loop and short helix (46–58), one from the loop connecting the first β and α structures of the second $\beta\alpha\beta$ motif (90–95), and the third from the loop residues (128–132) that enter the C-terminal α helix. The tryptophan and tyrosine residues forming the aromatic claws can be seen on either side of the active site hole protruding from the surface. Surface structure rendering was done with the program GRASP (Nicholls *et al.*, 1991).

directed mutagenesis studies of the nonfusion protein, identified that Cys 12 is the nucleophilic cysteine and that Cys 17 affects, but is not essential for, enzyme activity (Davis *et al.*, 1994b). In the BHPTP structure, the nonessential cysteine residue of the active site motif, Cys 17, is positioned near the phosphate ion, but not close enough for strong hydrogen bonding, and so may contribute only in a small way to the general stabilization of the phosphoenzyme complex. Chemical modification at this site, however, would clearly block substrate access to the active site loop. On the other hand, the position of Cys 12 in the structure clearly indicates its function in the nucleophilic attack of the enzyme on a phosphotyrosine substrate. As expected, the side chain of the nucleophilic Cys 12 in the structure of the BHPTP–phosphate complex is placed at an ideal position for a typical S_N2 displacement reaction, directly centered between the three oxygens of the tetrahedral

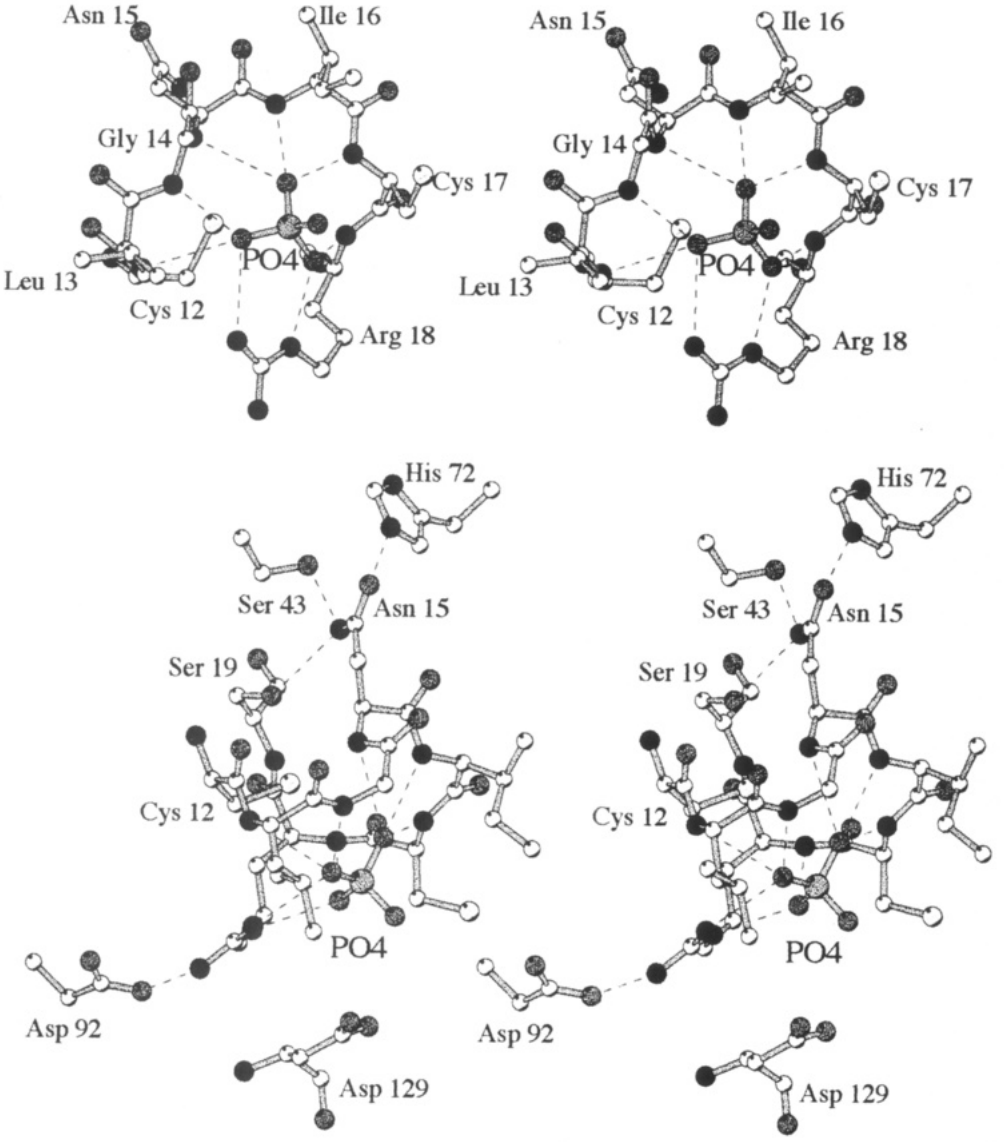

FIGURE 6: Two stereoviews of the active site of the BHPTP molecule. (a, top) Conformation of the active site loop conserved sequence residues corresponding to the low molecular weight phosphatase sequence motif XCXXXXCRS. The view is directly into the active site from approximately the same direction as is shown in Figure 2a. Cys 12 is the last residue of the β strand of the $\beta\alpha\beta$ motif. Residue Arg 18 is in the top of the first turn of the α helix. Backbone NHs from all six amino acids in the loop act to position the phosphate with hydrogen bonds to the oxygens. The side chain of Arg 18 closes the loop and contributes two hydrogen bonds to the phosphate oxygens. The nucleophilic residue Cys 12 is centered directly behind the bound phosphate ion, centered equally between the three phosphate oxygens. (b, bottom) Side view of the active site including residues important for positioning the active site loop and the putative proton donating residue. All hydrogen-bonding contacts to the phosphate shown in (a) are included (residues now unlabeled). The network of hydrogen bonds, including those from the conserved His 72, which position the side chain of Asn 15 is shown at the top of the figure. Asp 92, which stabilizes the position of the essential Arg 18, is shown at the lower left of the figure. Asp 129, which is suggested as a proton donor for the leaving group in the reaction, is shown at the bottom of the figure, on the opposite side of the phosphate from Cys 12.

face of the phosphate ion (Figure 6a). The phosphoenzyme intermediate that is formed in this reaction would then be attached through Cys 12 as oriented by the active site loop, again in agreement with the chemical and spectroscopic studies which first detected the intermediate (Davis *et al.*, 1994a; Wo *et al.*, 1992b).

Other Residues of Importance Suggested by the BHPTP Structure. Additional important features of the reaction can be postulated on the basis of the relative positions of the residues in this portion of the molecule. The leaving group in the S_N2 reaction could be protonated by a nearby aspartic acid residue (Asp 129) located on the loop which precedes the C-terminal helix, which is 3.66 Å away from the phosphate (Figure 5b). After the formation of a phosphoenzyme intermediate, the dephosphorylation event, which is normally the rate-determining step, would occur by attack of water on the phosphoenzyme intermediate with the subsequent release of inorganic phosphate. The lack of a D_2O solvent isotope effect in this step (Zhang & Van Etten, 1991b) suggests that water attacks the phosphoenzyme intermediate without the assistance of a general base. This suggests that the aspartate anion may not function as a general base in the dephosphorylation step. An ordered water molecule which might serve the function of releasing the phosphate from the phosphoenzyme intermediate is not apparent at the active site.

BHPTP has a 94% and 85% sequence identity to two human low molecular weight PTPase isoenzymes (HCPTP-B and HCPTP-A, respectively). HCPTP-A and HCPTP-B are two isoenzymes produced by alternative mRNA splicing and are expressed in all major human organs, including brain (Dissing *et al.*, 1991; Wo *et al.*, 1992a). Although the region near the active site in these phosphatases is strongly conserved, the major differences between BHPTP and HCPTP-A all lie within a 20 amino acid stretch in the middle of the sequence.

This variable region corresponds to a stretch of residues in the extended loop and helix motif which wraps around the BHPTP molecule. Two of the extended runs of this loop cross over the active site and contribute some basic residues to the region which may be important for substrate binding. The variation between these molecules in amino acid composition in this loop region is highly suggestive of a role for this loop in substrate selectivity.

Roles of Histidines and Serines in Forming the Active Site. Two conserved histidine residues (His 66, His 72) (see Figure 4 for locations) are found in all the members of the low molecular weight PTPase family. Site-directed mutagenesis studies have shown that although mutants of His 66 preserve their activity toward pNPP compared to the wild type (Davis *et al.*, 1994a), His 72 mutants lose about 75% of their activity. These results indicated that His 72 apparently is located at or near the active site, consistent with the vanadate titration experiment noted earlier (Zhou *et al.*, 1993). NMR pH titrations also revealed that His 66 and His 72 have unusually high pK_a values of 8.36 and 9.19, respectively (Davis *et al.*, 1994a). The unchanged pK_a of His 72 from pH titrations of mutants C12A and C17A ruled out the possibility of interaction between histidine and these active site cysteines (Davis *et al.*, 1994b).

The present BHPTP structure resolves the interpretation of these experiments and indicates only an indirect function for His 72 in the formation of the active site. In the BHPTP structure, His 72 is located on the extended loop between $\alpha 2$ and $\alpha 3$ somewhat distant (8.65 Å) from the catalytic cysteine. The only contact this residue could have with the active site is a link through its interaction with Asn 15 on the active site loop (Figure 6b). This conserved histidine is one of three residues (His 72, Ser 43 and Ser 19) that are found to interact with Asn 15 on the active site loop. Ser 43 is positioned for this interaction by its location on the C-terminal portion of the second β strand of the active site $\beta\alpha\beta$ motif. Ser 19, which is the conserved serine of the signature sequence, lies in the first turn of the α helix in this same motif, just after the arginine residue that orients the substrate. The interaction of these three residues with Asn 15 apparently stabilizes the left-handed conformation ($\psi = 35°$, $\phi = 56°$) of this residue and modifies the run of the backbone so that all the NH groups in the active site loop are oriented in the same direction. The tight network of hydrogen bonds between His 72, Ser 43, Ser 19, and Asn 15 appears to serve a structural function, such that the active site can adopt the most favorable geometry for phosphate binding. This positioning not only explains the results of the spectroscopic and mutagenesis studies involving the histidine residues but also identifies a role for the conserved serine in the low molecular weight PTPase signature sequence. The reduced activity of the His 72 mutants now appears to be due to modification of the active site loop structure, which could change the orientation of the phosphorylated substrate and so affect activity. Reasonable hypotheses explaining the shift in pK_a of His 72 can also be proposed on the basis of the present structure. Two acidic residues (Glu 23 and Asp 42) are found near His 72 (<4 Å) and are apparently the buffering residues that raise the pK_a of this histidine.

From the structure, His 66 is immediately eliminated from consideration as part of the active site in contrast to the conclusions of Chiarugi *et al.* (1994). This histidine is located >15 Å away from the active site on the opposite side of the molecule, near the C-terminus of the short α helix at the center of the extended loop (Figure 4). In a similar manner to the perturbation of the pK_a of His 72, the pK_a of His 66 may be shifted by an acidic residue (Glu 139) that is near (<4 Å) its position in the structure.

Comparison to the High Molecular Weight PTPases. The BHPTP structure is distinct not only from serine or threonine, acid, and alkaline phosphatases but also from the known structure of the high molecular weight PTPase, PTP1B (Barford *et al.*, 1994). Due to the lack of sequence homology, it might be expected that the high molecular weight PTPases might share similar structural features while the low molecular weight PTPases would adopt a different type of three-dimensional fold. Nonetheless, these different tyrosyl phosphatases still preserve the most essential cysteine and arginine features of the active site, at either end of a phosphate binding loop.

The recently published structure of the tyrosyl phosphatase PTP1B (Barford *et al.*, 1994) allows a comparison of the overall features of the two molecules and a comparison of how these distinct, but catalytically related molecules create an active site designed for tyrosine dephosphorylation. Both structures reveal mixed α/β enzymes with central twisted β sheets surrounded on both sides by α helices. The PTP1B structure has a more extended and mixed central β sheet consisting of 10 β strands, as compared to the four-strand central sheet of the smaller BHPTP enzyme. The middle four parallel strands of PTP1B, which are widely separated in the sequence, appear to form a structure homologous to BHPTP, although in the larger enzyme the two $\beta\alpha\beta$ motifs of this portion of the sheet interdigitate instead of lying side by side, as in BHPTP. In contrast to the BHPTP structure, the signature sequence motif of the high molecular weight PTPase is not found in the β/α loop of the first $\beta\alpha\beta$ motif of the structure but rather at the C-terminus of the final β strand in the four-strand parallel portion of the sheet in a loop that connects the central sheet to a cluster of C-terminal helices. The α helix which ends the active site loop in PTP1B appears to lie nearly perpendicular to the two $\beta\alpha\beta$ motifs in the sheet, in contrast to the parallel position of the helix in BHPTP, which is included in this tightly wound $\beta\alpha\beta$ structural unit.

The details of the active sites in these two molecules are more closely related, although differences specific to each class of enzyme can be found. Both structures have their active site at the C-terminus of a parallel sheet with the conserved signature sequence forming a loop between a β strand and an α helix. For both enzymes the bound inhibitor (phosphate or tungstate) sits on the axis of the α helix that completes the loop. In each case the backbone NHs in the active site loop contribute to the binding of the ion. In both structures, the conserved arginine is intimately involved in binding the inhibitor and completes a loop of hydrogen bonds in the active site. In BHPTP, all three oxygens of one tetrahedral face of the phosphate are positioned by these interactions, where in PTP1B only two oxygens appear to be in contact with main-chain NHs. Where in PTP1B the backbone configuration in the active site that allows these bonds could be due to the presence of a glycine-rich loop, in BHPTP this configuration is facilitated by multiple hydrogen bonds of the active site loop Asn 15 with positioning histidine and serine residues. The histidine at the beginning of the signature sequence in PTP1B appears to position the nucleophilic cysteine in relation to the tungstate; in BHPTP, the location of the active site loop within a $\beta\alpha\beta$ structural motif defines the position of Cys 12 relative to the loop. A residue corresponding to the glutamine which helps position the tungstate in PTP1B is not seen in BHPTP. Surprisingly, no proton donor for the leaving group in the phosphorylation

step was found in PTP1B, and this may be expected to be a subject of some mechanistic interest. In contrast, the structure of BHPTP strongly suggests that an aspartatic acid residue located near the phosphate but opposite the nucleophilic cysteine is well positioned to perform such a function. The latter hypothesis is currently being tested experimentally.

The structure of BHPTP has revealed a unique way of defining a phophotyrosine phosphatase active site in this, the smallest of these enzymes. The high-resolution structure defines the details of the phosphate binding and catalytic mechanism of this enzyme. In this structure, the roles of all the residues in the conserved sequence motif as well as those residues (Cys 12, Cys 17, Arg 18, and His 72) that are directly implicated in the enzymatic reaction and active site can be clearly demonstrated. A surface groove formed by aromatic residues on the surface of the molecule, as well as the positions of a set of variable residues in the structure, suggests the basis of substrate selectivity. Other residues of potential importance to the reaction mechanism have also been identified and are now being studied by site-directed mutagenesis. Due to the high sequence homology within the low molecular weight PTPase family, BHPTP should also serve as a good model for the rest of the members of this phosphatase family, including the human red cell type isoenzymes. Structural studies of mutant enzymes and of the native enzyme in the presence of known inhibitors should now further define the molecular basis of the activity of this enzyme and open new ways to investigate the nature of its cellular substrates. Investigations of the parameters that affect substrate binding should allow the design of specific inhibitors for these low molecular weight PTPases as well as the definition of possible general principles of tyrosine phosphatase inhibition that might be applicable to the larger family of these enzymes.

ADDED IN PROOF

Site-directed mutagenesis and NMR experiments demonstrate that the D129A mutant, although correctly folded, has drastically reduced catalytic activity (Zhongtao Zhang and Robert L. Van Etten, unpublished results). In contrast, the V_{max} values of the D56A and D92A mutants are similar to that of the wild-type enzyme. This supports the hypothesis advanced here that D129 acts as a proton donor in leaving group protonation.

ACKNOWLEDGMENT

The authors acknowledge Martin Lawrence for his help in data collection and interpretation during the early stages of the structure solution and both Joyce Bell and Mark O'Neil for their valuable assistance in the production and illustration of the manuscript.

REFERENCES

Baldijao, C., Guija, E., Bittencourt, H., & Chaimovich, H. (1975) *Biochim. Biophys. Acta 391,* 316–325.
Barford, D., Flint, A. J., & Tonks, N. K. (1994) *Science 263,* 1397–1403.
Boivin, P., & Galand, C. (1986) *Biochem. Biophys. Res. Commun. 134,* 557–564.
Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235,* 458–460.
Camici, G., Manao, G., Cappugi, G., Modesti, A., Stefani, M., & Ramponi, G. (1989) *J. Biol. Chem. 264,* 2560–2567.
CCP4 (1979) Daresbury Laboratory, Warrington, U.K.
Charbonneau, H., & Tonks, N. K. (1992) *Annu. Rev. Cell Biol. 8,* 463–493.
Chernoff, J., & Li, H. C. (1985) *Arch. Biochem. Biophys. 240,* 135–145.
Chiarugi, P., Cirri, P., Camici, G., Manao, G., Fiaschi, T., Raugei, G., Cappugi, G., & Ramponi, G. (1994) *Biochem. J. 298,* 427–433.
Cohen, P. (1989) *Annu. Rev. Biochem. 58,* 453–508.
Davis, J. P., Zhou, M.-M., & Van Etten, R. L. (1994a) *J. Biol. Chem. 269,* 8734–8740.
Davis, J. P., Zhou, M.-M., & Van Etten, R. L. (1994b) *Biochemistry 33,* 1278–1286.
Dissing, J., Johnsen, A. H., & Sensabaugh, G. F. (1991) *J. Biol. Chem. 266,* 20619–20625.
Edelman, A. M., Blumenthal, D. K., & Krebs, E. G. (1987) *Annu. Rev. Biochem. 56,* 567–613.
Fischer, E. H., Charbonneau, H., & Tonks, N. K. (1991) *Science 253,* 401–406.
Heinrikson, R. L. (1969) *J. Biol. Chem. 244,* 299–307.
Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M. (1991) *Acta Crystallogr. A47,* 110–119.
Kabsch, W. (1988) *J. Appl. Crystallogr. 21,* 916–924.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24,* 946–950.
Laidler, P. M., Taga, E. M., & Van Etten, R. L. (1982) *Arch. Biochem. Biophys. 216,* 512–521.
Lawrence, G. L., & Van Etten, R. L. (1981) *Arch. Biochem. Biophys. 206,* 122–131.
Nicholls, A. N., Sharp, K. A., & Honig, B. (1991) *Proteins: Struct., Funct., Genet. 11,* 281–296.
Oakada, M., Owada, K., & Nakagawa, H. (1986) *Biochem. J. 239,* 155–162.
Ramponi, G., Manas, G., Camici, G., Cappugi, G., Ruggiero, M., & Bottano, D. P. (1989) *FEBS Lett. 250,* 469–473.
Rehkop, D. M., & Van Etten, R. L. (1975) *Hoppe Seyler's Z. Physiol. Chem. 356,* 1775–1782.
Saini, M. S., Buchwald, S. C., Van Etten, R. L., & Knowles, J. R. (1981) *J. Biol. Chem. 256,* 10453–10455.
Schultz, G. E. (1992) *Curr. Opin. Struct. Biol. 2,* 61.
Sokal, R. R., Oden, N. L., & Wilson, C. (1991) *Nature 351,* 143–145.
Taga, E. M., & Van Etten, R. L. (1982) *Arch. Biochem. Biophys. 214,* 505–515.
Tronrud, D. G., Ten Eyck, L. F., & Matthews, B. W. (1987) *Acta Crystallogr. A43,* 489–501.
Van Etten, R. L., Waymack, P. P., & Rehkop, D. M. (1974) *J. Am. Chem. Soc. 96,* 6782–6785.
Waheed, A., Laidler, P. M., Wo, Y.-Y. P., & Van Etten, R. L. (1988) *Biochemistry 27,* 4265–4273.
Wo, Y.-Y. P., MacCormack, A., Shabanowitz, J., Hunt, D. S., Davis, J. P., Mitchell, G., & Van Etten, R. L. (1992a) *J. Biol. Chem. 267,* 10856–10865.
Wo, Y.-Y. P., Zhou, M.-M., Stevis, P., Davis, J. P., Zhang, Z.-Y., & Van Etten, R. L. (1992b) *Biochemistry 31,* 1712–1721.
Zhang, M., Van Etten, R. L., Lawrence, C. M., & Stauffacher, C. V. (1994) *J. Mol. Biol. 238,* 281–283.
Zhang, Z.-Y., & Van Etten, R. L. (1990) *Arch. Biochem. Biophys. 282,* 39–49.
Zhang, Z.-Y., & Van Etten, R. L. (1991a) *J. Biol. Chem. 266,* 1516–1525.
Zhang, Z.-Y., & Van Etten, R. L. (1991b) *Biochemistry 30,* 8954–8959.
Zhang, Z.-Y., & Dixon, J. (1994) *Adv. Enzymol. 68,* 1–36.
Zhang, Z.-Y., Davis, J. P., & Van Etten, R. L. (1992) *Biochemistry 31,* 1701–1711.
Zhou, M.-M., Davis, J. P., & Van Etten, R. L. (1993) *Biochemistry 32,* 8479–8485.