Protein Tyrosine Phosphatase Substrate Specificity: Size and Phosphotyrosine Positioning Requirements in Peptide Substrates†

Zhong-Yin Zhang,‡ Derek Maclean,§ Dennis J. McNamara,§ Tomi K. Sawyer,§ and Jack E. Dixon*,‡

Department of Biological Chemistry, Medical School, Walther Cancer Institute, The University of Michigan, Ann Arbor, Michigan 48109, and Department of Chemistry, Parke-Davis Research Division, Warner-Lambert Company, Ann Arbor, Michigan 48105

Received October 18, 1993; Revised Manuscript Received November 30, 1993®

ABSTRACT: The structural requirements of substrates for two recombinant protein tyrosine phosphatases (PTPases) are probed using various-sized synthetic phosphotyrosine (pY)-containing peptides corresponding to the autophosphorylation site in EGF receptor (EGFR) at Y992. The peptide $EGFR_{988-998}$ (DADEpYLIPQQG) is chosen as a template due to its favorable kinetic constants. The contribution of individual amino acids on both sides of pY to binding and catalysis was assessed by kinetic analysis using a continuous, spectrophotometric assay. For both *Yersinia* PTPase and a soluble recombinant mammalian PTPase of 323 amino acid residues (rat PTP1), efficient binding and catalysis required six amino acids including the pY residue, i.e., four residues N-terminal to pY and one residue C-terminal to pY. Thus, PTPase substrate specificity is primarily dictated by residues to the N-terminal side of pY. The pY moiety and the rest of the peptide interact with PTPases in a cooperative manner. The presence of pY in the peptide substrate is necessary but not sufficient for high-affinity binding, since phosphotyrosine and other simple aryl phosphates exhibit weak binding, and dephosphorylated peptides do not bind to PTPases. Two variations on the pY moiety are also examined in order to assess their utility in PTPase inhibitor design. It is demonstrated that the thiophosphoryl analog in which one of the phosphate oxygens is replaced by sulfur can be hydrolyzed by PTPases, whereas the phosphonomethylphenylalanine analog in which the tyrosyl oxygen is replaced by a CH_2 group is a competitive and nonhydrolyzable inhibitor, with K_i values of 18.6 and 10.2 μM, respectively, for the *Yersinia* PTPase and the rat PTP1.

The reversible tyrosine phosphorylation of proteins has been recognized as a fundamental mechanism that cells utilize to regulate activities including proliferation, differentiation, and metabolism (Yarden & Ullrich, 1988; Fischer et al., 1991). Extensive research efforts have been directed toward protein tyrosine kinases (PTKs),[1] because PTKs are growth factor receptors, and cell cycle regulators, and are encoded by several oncogenes. Since the tyrosine phosphorylation "status" of a cell is maintained by PTKs and protein tyrosine phosphatases (PTPases), a complete understanding of the physiological role of tyrosine phosphorylation requires the characterization of the PTPases in addition to the PTKs. It is clear that the catalytic activity of receptor kinases as well as their associations with other SH2 domain-containing signaling molecules are tightly regulated by PTPases. Like PTKs, PTPases constitute a large diversified family of catalysts which can be divided into two structurally distinct groups, the receptor-like and the soluble cytoplasmic enzymes (Fischer et al., 1991; Walton & Dixon, 1993).

One of the central questions in PTPase research is how phosphatases distinguish the diversity of substrates that they encounter in the cell. Despite the rapid progress in the identification and characterization of new PTPases, there have been relatively few biochemical analyses of the mechanisms that govern PTPase substrate specificity. We envision that PTPase substrate specificity can be controlled at several levels. Sequence surrounding the site of phosphorylation, localization or "positional" information dictated by sequences outside of the active site, PTPase inhibitors and activators, posttranslational modifications, and competition with SH2 domain-containing proteins are all likely factors to contribute to the regulation and substrate specificity of PTPases. We also believe that, for at least some PTPases, substrate recognition may be mediated primarily by determinants in the vicinity of the phosphorylated tyrosine residue. cdc25 represents an extreme example which works only on its physiological substrate, cdc2 (Gautier et al., 1991). Other PTPases may be less discriminatory in terms of sequence specificity for protein dephosphorylation.

Using synthetic phosphotyrosine (pY)-containing peptides, several laboratories (Cho et al., 1991, 1993; Chatterjee et al., 1992; Zhang et al, 1993a,b) have demonstrated that PTPases displayed a range of k_{cat}/K_m values for the hydrolysis of short peptide fragments of various phosphoproteins, indicating that PTPase substrate specificity could be at least in part controlled at the primary structure level. These studies have identified peptide sequences with good kinetic properties toward a few PTPases, yet systematic searches for optimal sequences for PTPases are needed. Utilizing a sensitive and continuous assay that we developed for PTPases (Zhang et al, 1993a), we have analyzed the substrate specificity of the PTPases in greater detail to address some of the structural requirements for optimum substrate binding and catalysis. A strategy was employed in which each amino acid within the pY-containing peptide substrate $EGFR_{988-998}$ (DADEpYLIPQQG) (Rotin et al., 1992; McNamara et al., 1993) was sequentially

† The contribution of the first two authors should be considered equal. This work was supported by the National Institutes of Health Grant NIDDKD 18849 and the Walther Cancer Institute.

* To whom correspondence should be addressed.

‡ The University of Michigan, Ann Arbor.

§ Warner-Lambert Co.

® Abstract published in *Advance ACS Abstracts,* February 1, 1994.

[1] Abbreviations: PTPase, protein tyrosine phosphatase; PTK, protein tyrosine kinase; EGFR, epidermal growth factor receptor; SH2, *src* homology 2; pY, phosphotyrosine; pmF, phosphonomethylphenylalanine.

0006-2960/94/0433-2285$04.50/0

substituted by Ala (Ala-scan) (Zhang et al., 1993b). Specific contributions to the binding and catalysis by individual residues were ascertained by kinetic measurements. We demonstrated that peptides with acidic residues N-terminal to pY are important for binding and catalysis, especially at the −1 position. For example, when the Glu residue immediately N-terminal to the target pY residue (i.e., at position −1) is replaced by an Ala, this single substitution decreases the substrate specificity of *Yersinia* PTPase (in terms of k_{cat}/K_m) by 126-fold compared to the wild-type peptide.

In this study, we sought to define the size requirements for pY-containing peptide substrates as well as the optimal position of pY within the peptide sequence, in order to identify the range of molecular features on the peptide substrate that can be favorably accommodated by the PTPase active site. We have also modified the pY moiety in an attempt to achieve nonhydrolyzable PTPase inhibitors. Although many inhibitors of protein phosphatases are known, no potent and selective PTPase inhibitors have been reported. Such inhibitors would be useful in order to assess the physiological function of these enzymes. Thus, we assessed the effect of sulfur substitution on the pY moiety. We also evaluated the possibility of using phosphonopeptide as PTPase inhibitor in which pY was substituted with phosphonomethylphenylalanine (pmF), an unnatural, nonhydrolyzable analog of pY in which the >C-O-PO_3H_2 moiety is replaced by >C-CH_2-PO_3H_2.

MATERIALS AND METHODS

Enzyme Preparation. Homogeneous recombinant *Yersinia* PTPase (Zhang et al., 1992) and the mammalian PTPase, PTPU323 (Guan & Dixon, 1991a), were purified as described. PTPU323 was generated by introducing a stop codon at amino acid residue 323 in the cDNA coding for the rat brain PTP1. The stop codon eliminates a hydrophobic region of the protein located near the C-terminus. The C-terminus truncated molecule, PTPU323, retains all the full-length enzymatic activity (Guan & Dixon, 1991b). In this paper, we shall refer PTPU323 as PTP1. Enzyme and peptide concentrations were determined by amino acid analysis.

Peptide Preparation. Phosphotyrosine-containing peptides were synthesized, purified, and characterized as previously described (Zhang et al., 1993a). Thiophosphoryltyrosyl-containing peptide was prepared by a modification of the described procedure for the preparation of pY-containing peptides. Oxidation of the intermediate resin-bound phosphite triester was carried out with elemental sulfur (100 equiv) in carbon disulfide and dichloromethane. Subsequent preparative steps and characterization were as described for pY-containing peptides. Phosphonomethylphenylalanine-containing peptide was prepared as described (McNamara et al., 1993).

Enzyme Assay and Data Analysis. All enzyme assays were performed at 30 °C in 50 mM 3,3-dimethylglutarate, 1 mM EDTA, pH 6.6 buffer with a constant ionic strength of 0.15 M (adjusted with NaCl). pH 6.6 was chosen due to the optimal kinetic properties of $EGFR_{988-998}$ at this condition (Zhang et al., 1993b). All of the peptide substrates were analyzed by the continuous assay (Zhang et al., 1993a), which is based on the differences in the spectra of the peptide before and after the removal of the phosphate group. The increase in the absorbance at 282 nm of the peptide upon the action of PTPase can be followed continuously, and the resulting progress curve (time course) can be analyzed directly using the integrated form of the Michaelis–Menten equation. This procedure is convenient and efficient, since both k_{cat} and K_m values can be obtained in a single run. For standard spectrophotometric assays, a semimicrocuvette was used. The reaction solution (total volume of 500 μL) containing an appropriate amount of peptide substrate was incubated at 30 °C for at least 15 min before the reaction was started by introducing a catalytic amount of PTPase into the reaction mixture. The entire time course of the PTPase-catalyzed hydrolysis of tyrosine-phosphorylated peptide substrate was recorded by monitoring the increase in absorbance at 282 nm, and the Michaelis–Menten kinetic parameters k_{cat} and K_m were determined by analyzing the experimental data through a nonlinear least-squares fit algorithm (Yamaoka et al., 1981) using the integrated Michaelis–Menten equation:

$$t = p/k_{cat}E_o + (K_m/k_{cat}E_o) \ln [p_\infty/(p_\infty - p)]$$

where k_{cat} is the catalytic turnover number, K_m is the Michaelis constant, E_o is the enzyme concentration, and p and p_∞ are the product concentration at time t and infinity, respectively. This relationship can then be used directly to analyze an array of experimental t–p data pairs by nonlinear least-squares methods where the parameters k_{cat} and K_m are optimized through minimizing $\sum(t_{exp} - t_{calc})^2$. Kinetic parameters for PTPases using pNPP and phosphotyrosine were determined as previously described (Zhang & Van Etten, 1991) by following the production of *p*-nitrophenol and inorganic phosphate, respectively.

K_i Determination. The effect of the phosphonomethylphenylalanine (pmF)-containing peptide (DADEpmFLIPQQG) on the PTPase-catalyzed hydrolysis of DADEpYLIPQQG was also evaluated using the integrated Michaelis–Menten equation. Inhibitor concentrations of 0, 15.3, and 30.6 μM were used for the *Yersinia* PTPase, while inhibitor concentrations of 0, 9.55, and 19.1 μM were used for PTP1. Kinetic parameters k_{cat} and K_m were obtained from nonlinear least-squares fits of the integrated Michaelis–Menten equation. Since the presence of pmF-containing peptide had no effect on k_{cat} values within experimental errors, and since the presence of pmF-containing peptide always caused an increase in the apparent K_m values, the mode of inhibition was assumed to be competitive. The K_i value for the phosphonic acid derivative was calculated using $K_m^{app} = K_m^o (1 + [I]/K_i)$, where Ks_m^{app} is the apparent K_m in the presence of a competitive inhibitor, Ks_m^o is the apparent K_m value in the absence of an inhibitor, and [I] is the inhibitor concentration.

RESULTS AND DISCUSSION

Size Requirements and Positioning of pY within the Peptide for the Yersinia PTPase. The pY-containing peptide DADEpYLIPQQG ($EGFR_{988-998}$, peptide **21** in Tables 1 and 2), which corresponds to the autophosphorylation site Tyr992 in the EGF receptor, is by far the best substrate reported for both the *Yersinia* PTPase and the rat PTP1 (Zhang et al., 1993b) with k_{cat}/K_m values of 2.23×10^7 and 2.88×10^7 M^{-1} s^{-1}, respectively, at pH 6.6 and 30 °C. These rate constants are near the diffusion-controlled limit. We observed that k_{cat} values for most peptide substrates were relatively constant (Zhang et al., 1993a,b). If *in vivo* PTPase protein/peptide substrates exhibit similar k_{cat} values, one would not expect to see K_m values much lower than those that were reported for the best peptide substrates. Thus, conceivably PTPase substrate specificity can result from differences in affinity conferred by a linear sequence surrounding pY. Phosphotyrosine itself, however, displays k_{cat}/K_m values of 1.71×10^4

Table 1: Size Requirements for the *Yersinia* PTPase[a]

	substrate	k_{cat} (s^{-1})	K_m (μM)	$10^{-7} \times k_{cat}/K_m$ (M^{-1} s^{-1})
1	pNPP	89.3 ± 1.3	1700 ± 39	$5.25 \times 10^{-3} \pm 1.1 \times 10^{-4}$
2	**pY**	164 ± 4.4	9580 ± 700	$1.71 \times 10^{-3} \pm 3.5 \times 10^{-4}$
3	Ac-**pY**-NH_2	583 ± 128	7080 ± 1450	$8.23 \times 10^{-3} \pm 8.2 \times 10^{-4}$
4	**EpYL**	890 ± 50	4860 ± 280	$1.83 \times 10^{-2} \pm 1.2 \times 10^{-3}$
5	**EpYL**-NH_2	1145 ± 81	3240 ± 328	$3.53 \times 10^{-2} \pm 1.1 \times 10^{-3}$
6	Ac-**EpYL**	1382 ± 34	1621 ± 59	$8.52 \times 10^{-2} \pm 4.4 \times 10^{-3}$
7	Ac-**EpYL**-NH_2	1199 ± 38	798 ± 58	0.150 ± 0.0062
8	**EpYLI**	1088 ± 29	3560 ± 96	$3.06 \times 10^{-2} \pm 3.4 \times 10^{-4}$
9	**DEpYL**	1336 ± 22	561 ± 23	0.238 ± 0.012
10	**DEpYLI**	1379 ± 74	659 ± 94	0.209 ± 0.020
11	Ac-**DEpYLI**-NH_2	1250 ± 145	720 ± 80	0.174 ± 0.014
12	**ADEpYLI**-NH_2	1142 ± 24	214 ± 17	0.534 ± 0.030
13	Ac-**ADEpYLI**-NH_2	1445 ± 87	269 ± 67	0.537 ± 0.035
14	**ADEpYLIP**	1286 ± 58	408 ± 40	0.315 ± 0.0079
15	**pYLIPQQG**	327 ± 41	2200 ± 240	$1.49 \times 10^{-2} \pm 7.7 \times 10^{-4}$
16	**DADEpY**-NH_2	1403 ± 22	291 ± 20	0.482 ± 0.022
17	**DADEpYL**	1381 ± 12	100 ± 6.3	1.38 ± 0.048
18	**DADEpYL**-NH_2	1182 ± 16	90.7 ± 8.1	1.41 ± 0.10
19	Ac-**DADEpYL**	1467 ± 82	1240 ± 129	0.118 ± 0.0039
20	Ac-**DADEpYL**-NH_2	1340 ± 25	101 ± 16	1.33 ± 0.087
21	**DADEpYLIPQQG**	1314 ± 18	59 ± 4.8	2.23 ± 0.084

[a] All of the error analyses were done automatically by computer using the nonlinear least-squares fit algorithm with the damping Gauss–Newton method.

Table 2: Size Requirements for the Rat PTP1

	substrate	k_{cat} (s^{-1})	K_m (μM)	$10^{-7} \times k_{cat}/K_m$ (M^{-1} s^{-1})
1	pNPP	48.6 ± 1.0	1474 ± 31	$3.30 \times 10^{-3} \pm 1.3 \times 10^{-4}$
2	**pY**	44.1 ± 0.73	4930 ± 340	$8.94 \pm 10^{-4} \pm 7.8 \times 10^{-5}$
3	Ac-**pY**-NH_2	62.0 ± 3.3	1050 ± 39	$5.90 \times 10^{-3} \pm 4.4 \times 10^{-4}$
4	**EpYL**	70.9 ± 1.8	930 ± 41	$7.62 \times 10^{-3} \pm 2.6 \times 10^{-4}$
5	**EpYL**-NH_2	71.7 ± 1.9	438 ± 34	$1.53 \times 10^{-2} \pm 8.2 \times 10^{-4}$
6	Ac-**EpYL**	75.1 ± 0.96	146 ± 12	$3.41 \times 10^{-2} \pm 2.2 \times 10^{-3}$
7	Ac-**EpYL**-NH_2	68.1 ± 0.62	116 ± 6.1	$5.87 \times 10^{-2} \pm 2.6 \times 10^{-3}$
8	**EpYLI**	68.2 ± 0.76	1010 ± 54	$6.75 \pm 10^{-3} \pm 1.1 \times 10^{-5}$
9	**DEpYL**	71.6 ± 0.45	60.8 ± 2.3	0.118 ± 0.0041
10	**DEpYLI**	66.5 ± 1.1	59.9 ± 1.0	0.111 ± 0.0084
11	Ac-**DEpYLI**-NH_2	71.5 ± 0.68	20.2 ± 3.2	0.354 ± 0.052
12	**ADEpYLI**-NH_2	73.9 ± 0.38	19.1 ± 1.3	0.387 ± 0.024
13	Ac-**ADEpYLI**-NH_2	71.4 ± 0.85	15.0 ± 0.28	0.476 ± 0.023
14	**ADEpYLIP**	88.0 ± 4.0	82.1 ± 8.4	0.107 ± 0.0071
15	**pYLIPQQG**	43.8 ± 0.82	377 ± 19	$1.16 \times 10^{-2} \pm 3.5 \times 10^{-4}$
16	**DADEpY**-NH_2	72.8 ± 0.58	16.5 ± 1.3	0.441 ± 0.030
17	**DADEpYL**	69.0 ± 1.2	17.8 ± 1.2	0.388 ± 0.023
18	**DADEpYL**-NH_2	71.8 ± 0.51	3.20 ± 0.51	2.24 ± 0.18
19	Ac-**DADEpYL**	59.8 ± 0.55	13.2 ± 1.8	0.453 ± 0.022
20	Ac-**DADEpYL**-NH_2	67.6 ± 1.2	3.60 ± 0.47	1.88 ± 0.12
21	**DADEpYLIPQQG**	75.7 ± 1.0	2.63 ± 0.37	2.88 ± 0.24

and 8.94×10^3 M^{-1} s^{-1}, respectively, for the *Yersinia* PTPase and the rat PTP1 (Tables 1 and 2). Thus, the k_{cat}/K_m values for pY are 1300- and 3200-fold lower than those of $EGFR_{988-998}$ with the *Yersinia* PTPase and PTP1, respectively. The most commonly used low molecular weight artificial substrate, *p*-nitrophenyl phosphate (pNPP), displays kinetic constants similar to those of pY. It is observed that the presence of pY within the peptides is crucial for substrate binding because unphosphorylated peptides do not have any effects on the turnover of phosphopeptides at up to millimolar concentrations (Zhang et al., 1993a). Yet pY alone can not contribute to biological specificity. These suggest to us that the binding of a pY-containing peptide must be a cooperative event that involves the recognition of both pY itself and structural features from surrounding residues.

The relative positioning of pY within the peptide sequence also seems important for PTPase action. This is apparent on examining the kinetic constants of a pair of peptide substrates, $EGFR_{1171-1180}$ (AEpYLRVAPQS, k_{cat} = 1155 s^{-1} and K_m = 1.99 mM) and EGFR1167–1176 (TAENAEpYLRV, k_{cat} = 1146 s^{-1} and K_m = 0.367 mM) (Zhang et al., 1993b). Thus, by extending the N-terminal of the target tyrosine from −2 to −6, and at the same time shortening the C-terminal from +7 to +3, an increase of more than 5-fold in the k_{cat}/K_m value was observed. A similar observation has been made by Cho et al. (1991) on the catalytic domain of LAR.

To determine the structural determinants for PTPase recognition, we chose the phosphopeptide $EGFR_{988-998}$ (DADEpYLIPQQG) as a template. Table 1 summarizes the Michaelis–Menten kinetic parameters of pNPP, pY, and various pY-containing peptides with the homogeneous recombinant *Yersinia* PTPase at pH 6.6 and 30 °C. When all of the N-terminal residues of $EGFR_{988-998}$ are removed while keeping all of the C-terminal residues, the resulting peptide pYLIPQQG (peptide **15**) is only about 8.7-fold better than pY. On the other hand, when all of the C-terminal residues

are eliminated while keeping all of the N-terminal residues of $EGFR_{988-998}$, the peptide DADEpY-NH_2 (peptide **16**) is 282-fold better than pY alone and only 4.6-fold worse than the parent peptide. It seems that residues C-terminal to the pY residue contribute to substrate binding but the predominant determinants reside in the residues N-terminal to pY.

In order to more carefully characterize the importance of individual residues, we carried out a systematic analysis of structure–activity relationships within a series of phosphopeptides of increasing size. Starting with pY, residues were systematically added from both sides of pY until the kinetic parameters are comparable to those found for the parent peptide $EGFR_{988-998}$. By acetylating the free amino group and amidating the free carboxyl group of pY (giving **3**), k_{cat}/K_m is increased by 5-fold compared with that of pY. Addition of one residue to either terminus of pY yields the tripeptide EpYL (peptide **4**) which is turned over 10 times more rapidly than pY alone. Elimination of the negative charge on the C-terminal carboxylate of EpYL by amidation (peptide **5**) increases the k_{cat}/K_m value by 2-fold, while the elimination of the positive charge on the N-terminal amino group of EpYL by acetylation (peptide **6**) increases the k_{cat}/K_m value by 4-fold. This suggests that the electrostatic charges on the N- and C-termini of peptide **4** are poorly tolerated. Simultaneous elimination of the charge on both termini of EpYL (peptide **4**) (giving peptide **7**) results in an 8-fold increase in k_{cat}/K_m, indicating that these modifications are additive in their effects. The tetrapeptide EpYLI (peptide **8**) which has one more residue added to the C-terminal of EpYL shows kinetic parameters similar to those of EpYL, whereas the tetrapeptide DEpYL (peptide **9**) which has one more residue added to the N-terminal of EpYL shows a k_{cat}/K_m value that is 13-fold higher than that of EpYL. The addition of an Ile residue on the C-terminal side of the peptide DEpYL generating the pentapeptide DEpYLI (peptide **10**) does not show any further improvement on substrate binding. Therefore, it appears that the extension of pY-containing peptide beyond the +1 position by an Ile residue does not contribute significantly to the PTPase substrate specificity. Furthermore, the acetylation of the N-terminal residue and the amidation of the C-terminal residue of the pentapeptide DEpYLI generates peptide **11** which behaves with indistinguishable kinetics from DEpYLI. When the acetyl group in peptide **11** is replaced with an Ala residue to give the hexapeptide ADEpYLI-NH_2 (peptide **12**), a 3-fold increase in k_{cat}/K_m value is observed. This indicates that the acetyl functionality is insufficient to substitute for Ala at the −3 position. The parent peptide DADEpYLIPQQG has a k_{cat}/K_m value that is only 4-fold better than that of peptide **12** (ADEpYLI-NH_2). Acetylation of the N-terminal amino group in peptide **12** (giving peptide **13**) does not produce a noticeable effect on the kinetic parameters, suggesting that the acetyl group is not sufficient to replace an Asp residue at the −4 position. Interestingly, the attachment of a Pro residue to the C-terminal of ADEpYLI yielding peptide **14** actually decreases the affinity of the *Yersinia* PTPase toward the peptide, although we have shown earlier that substitution of the Pro residue by an Ala in $EGFR_{988-998}$ increased the K_m by 4-fold (Zhang et al., 1993b).

The hexapeptide DADEpYL (peptide **17**), which is obtained by incorporating just one more Leu residue to the C-terminal side of pY in peptide DADEpY, displays a k_{cat}/K_m value (3-fold better than DADEpY) of 1.38×10^7 M^{-1} s^{-1} that is almost as high as that of the parent peptide $EGFR_{988-998}$. C-terminal amidation of peptide **17** (giving peptide **18**) does not show any noticeable improvement. Surprisingly, N-terminal acetylation of DADEpYL (peptide **19**) decreases the substrate specificity by nearly 12-fold. Finally, amidation of the C-terminal of Ac-DADEpYL (peptide **20**) generates Ac-DADEpYL-NH_2 which is found to be kinetically similar to DADEpYL, displaying a k_{cat}/K_m value approaching that of the parent peptide. It is not clear why acetylation of the N-terminal of DADEpYL makes the *Yersinia* PTPase more sensitive to the negative charge on the C-terminal carboxylate. In any event, it is probably a good idea to always amidate the C-terminal carboxylate in order to remove any potential charge–charge repulsion. Thus, an optimal small peptide substrate for the *Yersinia* PTPase is DADEpYL-NH_2.

It is well-established that PTPase-catalyzed hydrolysis reaction involves a covalent phosphoenzyme intermediate, and a thiophosphate enzyme intermediate has been trapped using either pNPP or pY-containing peptide substrates (Guan & Dixon, 1991b; Cho et al., 1992). Thus, the PTPase-catalyzed reaction involves at least two chemical steps, i.e., the formation and the breakdown of the phosphoenzyme intermediate. During the course of this study, we noticed that k_{cat} values for the *Yersinia* PTPase-catalyzed hydrolysis of phosphate monoesters at pH 6.6 increase significantly from simple aryl phosphates to pY-containing peptides (Table 1). Also it seems that the k_{cat} values of peptide substrates are sensitive to the amino acid side chain at −1 position, and, in addition, the *Yersinia* PTPase requires at least two residues N-terminal to pY in order to achieve the maximal k_{cat} values. We do not understand this fully, but it appears that, at pH 6.6, factors in addition to the breakdown of the phosphoenzyme intermediate may also contribute to the rate-determining step in the *Yersinia* PTPase-catalyzed reaction. On the other hand, k_{cat} values for all of the substrates examined are amazingly similar for PTP1. This observation is consistent with PTP1 forming a covalent phosphoenzyme intermediate which is on the catalytic pathway and the breakdown of the intermediate is rate-limiting under the described conditions.

Size Requirements and Positioning of pY within the Peptide for the Rat PTP1. The structural requirements for the mammalian cytoplasmic PTP1 from rat are also examined with the same set of compounds and pY-containing peptides. The Michaelis–Menten kinetic parameters for the homogeneous recombinant rat PTP1 at pH 6.6 and 30 °C are summarized in Table 2. It is apparent that, in general, the trend of effects of the incremental increase in peptide length, and absence or presence of terminal charges on the peptide for the PTP1 binding and catalysis is similar to that observed for the *Yersinia* PTPase as discussed above. In only a small number of instances have differences between the two enzymes been noticed. For example, as shown in Table 2, Ac-DEpYLI-NH_2 (peptide **11**) displays a k_{cat}/K_m value that is 3-fold higher than that of DEpYLI (peptide **10**) and very similar to that of ADEpYLI-NH_2 (peptide **12**). So in the case of PTP1 and in contrast to the *Yersinia* PTPase, an acetyl group can effectively substitute for the Ala residue at position −3. On the other hand, like the *Yersinia* PTPase, the acetyl group is not an effective substitute for an Asp residue at position −4 for PTP1. Another striking difference between the two enzymes is that, unlike the *Yersinia* PTPase, PTP1 will not tolerate a negative charge on the C-terminal carboxyl group at the +1 position under any circumstances. Thus, an optimal small peptide substrate for PTP1 is DADEpYL-NH_2 (peptide **18**), which displays k_{cat}/K_m values very similar to that of the parent peptide $EGFR_{988-998}$ (DADEpYLIPQQG).

In summary, data from the *Yersinia* PTPase and PTP1 establish that the minimal phosphopeptide for optimal binding

and catalysis by both enzymes requires six amino acids including the pY residue. A good concensus substrate is DADEpYL-NH_2, which includes four amino acid residues N-terminal to pY and one amino acid C-terminal to pY. The fact that DADEpYL-NH_2 exhibits k_{cat}/K_m values against both PTPases that are approaching the diffusion-controlled limit suggests that short phosphopeptide sequences corresponding to a natural phosphorylation motif in a physiological substrate may confer *in vivo* substrate specificity by the PTPases.

PTPase Specificity Is Mainly Controlled by Residues N-Terminal to pY. The most important finding in this study is that efficient PTPase binding and catalysis requires six amino acid residues including pY. It appears that PTPase substrate specificity is primarily dictated by the properties of amino acid residues N-terminal to pY; thus four amino acid residues N-terminal to pY are desirable, and only one amino acid residue C-terminal to pY is necessary for the optimal binding. This is consistent with our previous results utilizing an alanine scan (Zhang et al., 1993b) which demonstrated that substitution of acidic residues N-terminal to pY by Ala resulted in substantial loss in substrate specificity while substitutions of residues on the C-terminal side of pY by Ala has a very modest effect on binding and catalysis. Another important feature for both enzymes is that in general the negative charge on the carboxylate at the +1 position is detrimental for PTPase recognition.

It is noteworthy to point out that *src* homology 2 (SH2) domains, which are protein modules containing about 100 amino acids that bind to specific pY-containing proteins/peptides, also recognize linear sequences surrounding phosphorylated tyrosine. Interestingly, unlike PTPases, the binding specificity of SH2 domain appears to be largely dictated by amino acids immediately C-terminal of pY. The crystal structures of the *lck* SH2 domain (Eck et al., 1993) and the *src* SH2 domain (Waksman et al., 1993) complexed with a pY-containing peptide reveal that only 5–7 residues make direct contact with the protein, with most extensive interactions centered around pY, the +1, and the +3 positions. Analyses of SH2 domain interaction with a pY-containing peptide library indicate that, in addition to the pY residue, the three residues immediately C-terminal to pY play a major role in defining the SH2 domain specificities (Shoelson et al., 1992; Songyang et al., 1993). Furthermore, functional studies have shown that a phosphopeptide motif as short as five amino acid residues (pYVPML) can effectively block the association of phosphotidylinositol (PI) 3′-kinase to the PDGF receptor (Fantl et al., 1992). It is interesting to note that Y992 of the EGF receptor is a high-affinity site for the SH2 domains of PLCγ (Rotin et al., 1992) and that both the *Yersinia* PTPase and PTP1 dephosphorylate peptides designed around this site most efficiently.

PTPase Inhibitor Design: The Effect of Thiophosphorylation and the Use of a Phosphonomethylphenylalanine-Containing Peptide as a Nonhydrolyzable, Competitive PTPase Inhibitor. While the amino acid sequence around pY appears to be an important structural feature in determining PTPase substrate specificity, it is clear that the pY moiety is absolutely required for PTPase recognition. For example, it is shown that if the phosphate group is removed from the pY moiety, the resulting peptide would not bind to the phosphatase (Zhang et al., 1993a; Ruzzene et al., 1993). It is also shown that the replacement of either Ser-P or O-methylated phosphotyrosine for pY within a suitable peptide substrate gives rise to a totally inert derivative (Ruzzene et al., 1993), suggesting that both the tyrosine residue and the charge on the phosphate group are important for PTPase recognition. Two structural variations which modify the phosphotyrosyl ester functionality and do not destroy the aromatic system and the charges on phosphate are employed to assess pY recognition. These are the phosphonomethylphenylalanine (pmF) in which the tyrosyl ester oxygen is replaced by a CH_2 group and thiophosphoryl analog [pY(s)] in which one of the $-PO_3$ oxygens is replaced by sulfur. Their structures are shown in Figure 1. Our initial idea of using a pY(S)-containing peptide as a PTPase inhibitor came from the report that the thiophosphorylated RCM lysozyme could be used as an affinity agent to purify human PTP1B (Tonks et al., 1988). To our surprise, the peptide DEpY(S)LI is found to be hydrolyzed by both the *Yersinia* PTPase and PTP1. At pH 6.6 and 30 °C, DEpY(S)LI displays k_{cat} values of 988 and 74.4 s^{-1} and K_m values of 1120 and 218 μM, respectively, for the *Yersinia* PTPase and PTP1. In comparison with DEpYLI (Tables 1 and 2), the major effect of sulfur substitution seems to be a significant increase in the apparent K_m values. This is in agreement with results of Cho et al. (1993) on the receptor-like PTPases. The fact that thiophosphorylated RCM lysozyme has been used to purify PTP1B may be due to the inaccessibility of thiophosphate in RCM lysozyme and the low temperature at which the affinity column was operated.

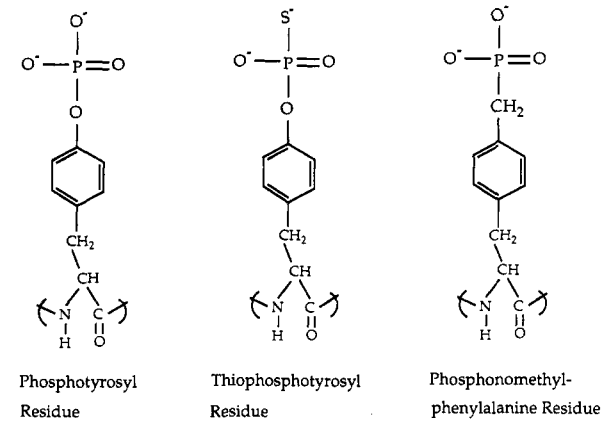

FIGURE 1: Structural comparison of the phosphotyrosine, the thiophosphorylated tyrosine, and the nonhydrolyzable phosphotyrosine analog, the phosphonomethylphenylalanine.

Phosphonomethylphenylalanine (pmF) is a nonhydrolyzable analog of pY since the ester oxygen is replaced with a CH_2 group (Figure 1). The use of pmF-containing peptides as PTPase inhibitors was initially demonstrated by Chatterjee et al. (1992). We made the pmF analog of $EGFR_{988-998}$ (DADEpmFLIPQQG) and evaluated its ability as a PTPase inhibitor for both *Yersinia* PTPase and PTP1. We found that DADEpmFLIPQQG exhibited K_i values of 18.6 ± 3.0 and 10.2 ± 0.9 μM, respectively, against the *Yersinia* PTPase and PTP1, and the pattern of inhibition was competitive. The fact that phosphonic acid analogues of pY-containing peptides are competitive, nonhydrolyzable inhibitors of PTPases and that they display K_i values in the range of K_m values of pY-containing peptides suggests that they can act as direct substrate mimics for PTPases.

In conclusion, we have demonstrated that, for both *Yersinia* PTPase and PTP1, good substrates require six amino acids including the pY residue, and that substrate specificity is mainly dictated by residues N-terminal to pY. Such a motif sequence may be used to search for potential physiological substrate(s) or for rational PTPase inhibitor design. Although

many inhibitors of protein phosphatases are known, there have been no potent and selective PTPase inhibitors yet reported. We believe the utilization of pmF as a nonhydrolyzable analog of pY and the systematic development of peptide mimetics by structure–activity strategy that is based on a specific, optimal phosphopeptide template will result in potent and selective inhibitors for PTPases. Such specific PTPase inhibitors will open new experimental approaches to define the role of these enzymes in cell function. Activation or deactivation of a particular pathway could be achieved by designing a small molecule that specifically disrupts one of the signaling pathways.

REFERENCES

Chatterjee, S., Goldstein, B. J., Csermely, P., & Shoelson, S. E. (1992) in *Peptides (Proceedings of the Twelfth American Peptide Symposium)*, pp 553–555, Leiden, The Netherlands.

Cho, H., Ramer, S. E., Itoh, M., Winkler, D. G., Kitas, E., Bannwarth, W., Burn, P., Saito, H., & Walsh, C. T. (1991) *Biochemistry 30*, 6210–6216.

Cho, H., Krishnaraj, R., Kitas, E., Bannwarth, W., Walsh, C. T., & Anderson, K. S. (1992) *J. Am. Chem. Soc. 114*, 7296–7298.

Cho, H., Krishnaraj, R., Itoh, M., Kitas, E., Bannwarth, W., Saito, H., & Walsh, C. T. (1993) *Protein Sci. 2*, 977–984.

Eck, M. J., Shoelson, S. E., & Harrison, S. C. (1993) *Nature 362*, 87–91.

Fantl, W. J., Escobedo, J. A., Martin, G. A., Turck, C. W., Rosario, M. D., McCormick, F., & Williams, L. T. (1992) *Cell 69*, 413–423.

Fischer, E. H., Charbonneau, H., & Tonks, N. K. (1991) *Science 253*, 401–406.

Gautier, J., Solomon, M. J., Booher, R. N., Bazan, J. F., & Kirschner, M. W. (1991) *Cell 67*, 197–211.

Guan, K. L., & Dixon, J. E. (1991a) *Anal. Biochem. 192*, 262–267.

Guan, K. L., & Dixon, J. E. (1991b) *J. Biol. Chem. 266*, 17026–17030.

McNamara, D. J., Dobrosin, E. M., Zhu, G., Decker, S. J., & Saltiel, A. R. (1993) *Int. J. Pept. Protein Res. 42*, 240–248.

Rotin, D., Margolis, B., Mohammadi, M., Daly, R. J., Daum, G., Li, N., Fischer, E. H., Burgess, W. E., Ullrich, A., & Schlessinger, J. (1992) *EMBO J. 11*, 559–567.

Ruzzene, M., Donella-Deana, A., Marin, O., Perich, J. W., Ruzza, P., Borin, G., Calderan, A., & Pinna, L. A. (1993) *Eur. J. Biochem. 211*, 289–295.

Shoelson, S. E., Chatterjee, S., Chaudhuri, M., & White, M. F. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2027–2031.

Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., & Cantley, L. C. (1993) *Cell 72*, 767–778.

Tonks, N. K., Diltz, C. D., & Fischer, E. H. (1988) *J. Biol. Chem. 263*, 6722–6730.

Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D., & Kuriyan, J. (1993) *Cell 72*, 779-790.

Walton, K. M., & Dixon, J. E. (1993) *Annu. Rev. Biochem. 62*, 101–120.

Yamaoka, K., Tanigawara, Y., Nakagawa, T., & Uno, T. (1981) *J. Pharmacobio-Dyn. 4*, 879–885.

Yarden, Y., & Ullrich, A. (1988) *Annu. Rev. Biochem. 57*, 443–478.

Zhang, Z.-Y., & Van Etten, R. L. (1991) *J. Biol. Chem. 266*, 1516–1525.

Zhang, Z.-Y., Clemens, J. C., Schubert, H. L., Stuckey, J. A., Fischer, M. W. F., Hume, D. M., Saper, M. A., & Dixon, J. E. (1992) *J. Biol. Chem. 267*, 23759–23766.

Zhang, Z.-Y., Maclean, D., Thieme-Sefler, A. M., Roeske, R., & Dixon, J. E. (1993a) *Anal. Biochem. 211*, 7–15.

Zhang, Z.-Y., Thieme-Sefler, A. M., Maclean, D., McNamara, D. J., Dobrusin, E. M., Sawyer, T. K., & Dixon, J. E. (1993b) *Proc. Natl. Acad. Sci. U.S.A. 90*, 4446–4450.