Screening Combinatorial Libraries by Mass Spectrometry. 2. Identification of Optimal Substrates of Protein Tyrosine Phosphatase SHP-1†

Peng Wang,‡ Hua Fu,‡ Donald F. Snavley, Michael A. Freitas, and Dehua Pei*

Department of Chemistry and Ohio State Biochemistry Program, The Ohio State University, 100 West 18th Avenue, Columbus, Ohio 43210

Received January 25, 2002; Revised Manuscript Received March 23, 2002

ABSTRACT: Protein tyrosine phosphatases (PTPs) are a large family of enzymes that catalyze the hydrolytic removal of the phosphoryl group from phosphotyrosyl (pY) proteins. In this work, we have developed a novel combinatorial library method, termed "enzyme-catalyzed loss of isotope peak signal enhancement (ECLIPSE)", to determine the substrate specificity of PTPs. This method involves partial labeling of pY at a nonbridging phosphate oxygen atom with 50% ^{18}O ($^{16}O/^{18}O = 1:1$). A 361-member solution-phase peptide library with randomization at the −1 and −2 positions (relative to pY), RNNXXpYA-NH_2 (X = 19 α-amino acids except for Cys), was synthesized with the partially ^{18}O-labeled pY by the split-synthesis method. Each member of the resulting pY peptide library appeared as a doublet peak in the mass spectrum (*m*/*z* m and m + 2.0043). Limited treatment of the library with a PTP removed the mass-degenerate phosphoryl group from the most preferred substrates to generate products as singlet peaks, which were readily identified and sequenced by tandem mass spectrometry. Screening of the pY library against the catalytic domain of SHP-1 revealed that SHP-1 prefers an acidic residue at the −2 position, with aspartic acid being slightly better than glutamic acid. At the −1 position, SHP-1 also prefers an acidic residue, although a variety of other amino acids are also tolerated. On the other hand, positively charged residues at these positions render the corresponding peptides very poor substrates of SHP-1. Several selected peptides were individually synthesized and assayed against SHP-1, and the kinetic data confirmed the screening results. These results demonstrate that ECLIPSE is a viable method for studying the substrate specificity of PTPs.

Reversible phosphorylation of proteins on tyrosyl residues is one of the key events that mediate the execution and regulation of many cellular processes. A proper level of phosphorylation is critical for these processes and is controlled by the opposing functions of protein tyrosine kinases (PTKs)[1] and protein tyrosine phosphatases (PTPs). Indeed, the imbalance of these two activities has been associated with a number of human diseases and conditions (*1*). A large number of PTKs and PTPs have been identified, but their precise mechanisms of action in vivo have been largely unknown. A major challenge in this field has been the determination of the in vivo targets/substrates of these PTKs and PTPs.

SHP-1 and SHP-2 are members of a subfamily of intracellular PTPs, which each contain two Src homology 2 (SH2) domains N-terminal to their PTP domain. SHP-1 is expressed primarily in cells of the hematopoietic origin, whereas SHP-2 is ubiquitously expressed. The two enzymes are actively involved in a variety of signaling pathways (reviewed in refs *2* and *3*). Despite their high sequence similarity (∼60% identity), SHP-1 and SHP-2 play very different roles in vivo. For example, SHP-2 generally acts as a positive regulator in many signaling pathways, whereas SHP-1 is primarily a negative regulator. The molecular basis for their contrasting roles is not yet clear. Some studies suggest that SHP-1 and SHP-2 may be recruited to different phosphotyrosyl (pY) receptors via their SH2 domains and subsequently dephosphorylate pY proteins in the vicinity of the receptors. However, domain-swapping studies between SHP-1 and SHP-2 have shown that their catalytic domains have distinct substrate specificity and that the intrinsic substrate specificity of the PTP domain also plays an important role in controlling their in vivo selectivity (*4*, *5*). It is currently believed that the in vivo selectivity of PTPs is likely controlled by a combination of specific targeting strategies and the intrinsic specificity of their catalytic domain (*3*).

One approach to identifying potential substrates of PTKs and PTPs is to study their intrinsic substrate specificity using purified proteins or synthetic peptides, followed by database searches using the consensus peptide sequence(s). Substrate specificity data would also be very useful in the design of selective inhibitors to modulate cellular signaling processes. Several laboratories have used synthetic pY peptides corre-

† This work was partially supported by a grant from the National Institutes of Health (AI40575).

* To whom correspondence should be addressed at Department of Chemistry, The Ohio State University, 100 West 18th Avenue, Columbus, OH 43210. Telephone: (614) 688-4068. Fax: (614) 292-1532. E-mail: pei.3@osu.edu.

‡ These two authors contributed equally to this work.

[1] Abbreviations: PTP, protein tyrosine phosphatase; pY, phosphotyrosine; ESI-MS, electrospray ionization mass spectrometry; FTICR, Fourier transform ion cyclotron resonance; ECLIPSE, enzyme-catalyzed loss of isotope peak signal enhancement.

10.1021/bi025591f CCC: $22.00
Published on Web 04/19/2002

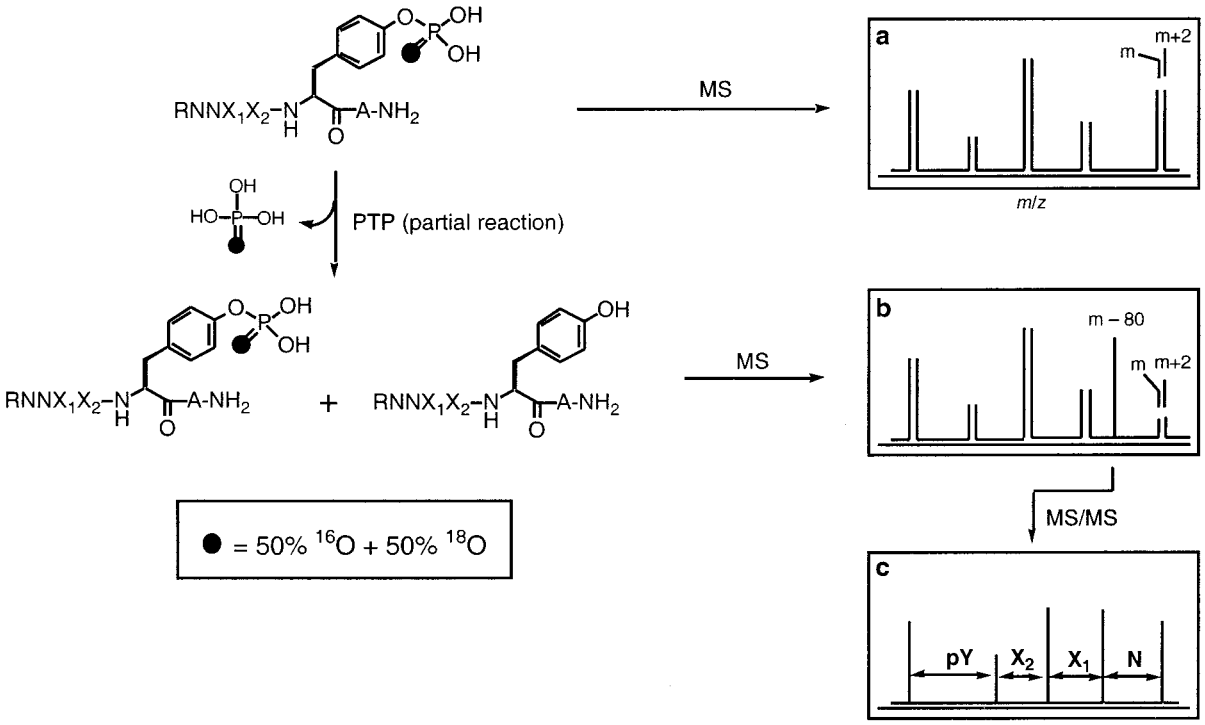

FIGURE 1: Scheme showing library screening by ECLIPSE. Each member of the starting library appears as a pair of doublets at *m*/*z* m and m+2, whereas a product appears as a singlet at *m*/*z* m-80 in the mass spectrum.

sponding to known phosphorylation sites in proteins to define the substrate specificity of PTPs (*6*−*11*). While these studies have clearly demonstrated the existence of primary sequence specificity (i.e., PTPs recognize the linear sequence flanking the pY residue), they have yet to produce a consensus sequence for any PTP. This is because the PTP active site makes contacts with 3−5 residues on either the N- or C-terminal side of pY (*12*, *13*). A complete characterization of each PTP would require the synthesis and testing of a prohibitively large number of single peptides (20^6−20^{10}). Therefore, more recent efforts have been focused on various combinatorial approaches, in which peptide libraries are synthesized simultaneously and screened against PTPs of interest. However, a major difficulty of the combinatorial approaches is to find a reliable method to separate/identify a reaction product (a tyrosyl peptide) from a complex mixture of substrates (pY peptides). Cheung et al. have employed α-chymotrypsin to discriminate between tyrosyl and pY peptides in a resin-bound library, based on the observation that α-chymotrypsin cleaves the dephosphorylated products after the tyrosyl residue but not the unreacted pY peptides (*14*). The drawback of this method is that it requires the exclusion of all aromatic and basic amino acids (e.g., Y, F, W, H, K, and R) from the library. A solution-phase library approach has also been reported independently by two laboratories (*15*, *16*). In this method, peptide libraries containing nonhydrolyzable pY analogues were incubated with PTP1B, and the resulting enzyme−peptide complexes were separated from unbound peptides by affinity or size-exclusion chromatography. The bound peptides were then analyzed by electrospray ionization mass spectrometry (ESI-MS). The major drawback of this method is that it only selects for high-affinity peptides to a PTP, which may or may not be a good substrate of the enzyme.

We have recently reported a novel combinatorial method, which we now refer to as "enzyme-catalyzed loss of isotope peak signal enhancement (ECLIPSE)", for the identification of optimal enzyme substrates (*17*). In this method, discrimination between substrates and products is achieved by partially labeling the substrates with a heavier isotope (heavy/normal isotope = 1:1), so that each member of the substrate library appears as a doublet in a mass spectrum. Enzymatic reaction removes the functional group that contains the isotopic label and the products appear as singlets in the spectrum, allowing for their unambiguous identification (Figure 1). We now report the application of this method for the determination of the substrate specificity of SHP-1. In this case, one of the nonbridging phosphate oxygen atoms of pY is labeled with 50% ^{18}O. Thus, each pY peptide should appear as a doublet in a mass spectrum, separated by 2.0043 Da. However, catalytic turnover by a PTP removes the phosphoryl group and therefore the mass degeneracy, resulting in products as singlet peaks. The product peaks are then selected for sequence analysis by tandem mass spectrometry. Using this method, we have determined the subsite specificity of SHP-1 at the −1 and −2 positions (relative to pY, which is defined as position 0).

MATERIALS AND METHODS

Materials and General Methods. Peptide synthesis reagents and resins were purchased from Advanced ChemTech (Louisville, KY). ^{18}O-Enriched water (95% ^{18}O) was purchased from Isotech Inc. (Oxford, OH). All other chemicals were obtained from Aldrich or Sigma. The catalytic domain of SHP-1, SHP-1(ΔSH2), was purified from a recombinant *Escherichia coli* strain as previously described (*18*). Protein concentrations were determined by Bradford assay using bovine serum albumin as standard. Concentrations of pY

peptides were determined by complete hydrolysis with SHP-1(ΔSH2) and measuring the absorbance increase at 282 nm ($\epsilon = 826\ M^{-1}\ cm^{-1}$ at pH 7.0) (*19*).

*Synthesis of tert-Butyl N$^{\alpha}$-[(Fluoren-9-yl)methoxylcarbonyl]-L-tyrosinate (**1**).* Fluorenylmethyl chloroformate (5.45 g, 21.0 mmol) was slowly added to a stirred solution of L-tyrosine *tert*-butyl ester (5.0 g, 21.0 mmol) in 70 mL of dioxane and 60 mL of 10% Na_2CO_3 in an ice/water bath. The mixture was stirred for 4 h in the ice/water bath and then 12 h at room temperature. The mixture was extracted with ethyl acetate (2 × 200 mL), and the organic phase was combined and evaporated under reduced pressure. The crude product was purified by silica gel chromatography (1:1 ether/hexane) to obtain a white solid (8.2 g, 85% yield). ^{1}H NMR (250 MHz, $CDCl_3$): δ 7.27−7.78 (m, 8H, Ar of Fmoc), 7.01 (d, 2H, Ar of Tyr), 6.73 (d, 2H, Ar of Tyr), 5.27 (d, 1H, NH of Tyr), 4.14−4.62 (m, 4H, CH_2 and CH of Fmoc, CH of Tyr), 3.70 (s, 1H, OH of Tyr), 3.00−3.03 (m, 2H, CH_2 of Tyr), 1.43 (s, 9H, *tert*-butyl).

*Synthesis of ^{18}O-Labeled N$^{\alpha}$-[(Fluoren-9-yl)methoxylcarbonyl]-O-[bis(allyloxy) phosphoryl]-*L-tyrosine (***4***).* 1H-Tetrazole (0.44 g, 6.0 mmol) was added to a solution of bis(allyloxy)diisopropylaminophosphine (1.68 g, 6.0 mmol) and *tert*-butyl *N*-Fmoc-L-tyrosinate (**1**) (1.84 g, 4.0 mmol) in anhydrous THF (15 mL), and the resulting solution was stirred under Ar atmosphere for 45 min at room temperature. A solution of iodine (1.52 g, 6.0 mmol) and 0.57 g (30 mmol) of water ($^{16}O/^{18}O$ = 1:1) in THF (5 mL) was added at −58 °C. After 2 h, an aqueous solution of 10% $Na_2S_2O_5$ (15 mL) was added to the mixture. The solution was stirred for 10 min at room temperature, transferred to a separatory funnel, and extracted with ethyl acetate (40 mL). The organic phase was washed with 5% $NaHCO_3$ (2 × 30 mL) and dried over Na_2SO_4. After silica gel chromatography (2:1 hexane/ethyl acetate), 1.78 g of a white solid (compound **3)** was obtained (72% yield). ^{1}H NMR (250 MHz, CD_3COCD_3): δ 7.16−7.87 (m, 12H, Ar of Fmoc and Tyr), 5.94−5.98 (m, 2H, $OCH_2CH{=}CH_2$), 5.20−5.41 (m, 4H, $OCH_2CH{=}CH_2$), 4.54−4.66 (m, 5H, $OCH_2CH{=}CH_2$, NH of Tyr), 4.16−4.43 (m, 4H, CH_2 and CH of Fmoc, CH of Tyr), 2.91−3.12 (m, 2H, CH_2 of Tyr), 1.41 (s, 9H, *tert*-butyl); ^{31}P NMR (CD_3COCD_3): δ −4.98.

Compound **3** (1.78 g, 2.87 mmol) was treated with 8 mL of trifluoroacetic acid overnight at room temperature. After the removal of trifluoroacetic acid by rotary evaporation, the residue was dissolved in 20 mL of diethyl ether. The solution was extracted with 5% $NaHCO_3$ solution (3 × 10 mL). The combined aqueous phase was washed with diethyl ether (3 × 10 mL), acidified to pH 2 by the addition of 30% HCl, and extracted with CH_2Cl_2 (3 × 15 mL). The organic phase was dried over Na_2SO_4, and the solvent was evaporated under reduced pressure to afford 1.6 g of compound **4** (99% yield). ^{1}H NMR (250 MHz, $CDCl_3$): δ 10.51 (br. 1H, COOH), 7.08−7,74 (m, 12H, Ar of Fmoc and Tyr), 5.84−5.95 (m, 2H, $OCH_2CH{=}CH_2$), 5.21−5.36 (m, 4H, $OCH_2CH{=}CH_2$), 4.43−4.63 (m, 5H, $OCH_2CH{=}CH_2$ and NH), 4.16−4.43 (m, 4H, CH_2 and CH of Fmoc, CH of Tyr), 2.81−3.18 (m, 2H, CH_2 of Tyr); ^{13}C NMR (62.5 MHz, $CDCl_3$): 199.42, 173.84, 156.29, 149.87, 149.76, 144.26, 144.11, 141.71, 133.71, 132.30, 132.19, 131.27, 128.13, 127.49, 125.49, 120.47, 120.38, 119.33, 69.59, 69.51, 67.39, 54.98, 53.87, 47.55, 37.43. ^{31}P NMR ($CDCl_3$): δ −6.20; ESI-MS: $[M+Na]^+$ *m*/*z* 586:588 (1:1).

Peptide Synthesis. The pY peptide library RNNXXpYA-NH_2 was constructed on 1.0 g of Rink resin (0.7 mmol/g) using standard Fmoc/HBTU/HOBt solid-phase peptide chemistry and ^{18}O-labeled **4**. The two randomized positions were prepared by the split-pool synthesis method (*20*−*22*) with 19 different amino acids (cysteine and methionine are excluded from the library and norleucine is included as methionine replacement). The coupling reactions were carried out with 5 equiv of Fmoc-protected amino acids for 3−4 h and repeated once to ensure complete reaction. After removal of the N-terminal Fmoc group with 20% piperidine in DMF, the resin-bound peptides were treated for 14 h with a cocktail containing diethylamine (10 equiv), HCOOH (10 equiv), Pd-$(PPh_3)_4$ (0.05 equiv), and PPh_3 (0.1 equiv) in THF to remove the O-allyl groups. Deprotection of peptide side chains and cleavage off the resin (∼300 mg) were carried out with 5 mL of TFA plus a small amount of anisole (100 μL), ethanedithiol (150 μL), and thioanisole (250 μL) for 2 h at room temperature. TFA and other volatile substances were removed under a gentle flow of nitrogen, and the residue was triturated for five times with diethyl ether. The resulting peptides were obtained as a white solid.

Individual pY peptides were synthesized in a similar manner using the Rink resin and unprotected *N*-Fmoc-pY (Advanced ChemTech). The identity of all peptides was confirmed by MALDI mass spectrometric analysis. With the exception of peptide RNNTQpYA-NH_2, which was purified by HPLC before enzymatic assays, all other peptides showed good purity (>80%) by HPLC and/or MALDI analysis and were used in PTP assays without further purification.

Library Screening. Typically, 5 mg of the peptide library (∼5 μmol) was added into 1 mL of doubly distilled water, and the sample was sonicated for 2 min to dissolve the peptides. The solution was then diluted in 40 mL of a buffer containing 1.5 mM ammonium bicarbonate (pH 7.0). The diluted solution was divided into four equal aliquots (10 mL each). One of the aliquots was kept as control (no PTP treatment), whereas the other three aliquots were incubated at room temperature with 8.0 μg of SHP-1(ΔSH2) for 3.5, 10, and 40 min, respectively, before being quenched by the addition of 0.1 mL of acetic acid. The resulting solutions were quickly frozen in liquid nitrogen and lyophilized for 20 h to dryness. The powder from each reaction was dissolved in 100 μL of a solvent mixture containing 75:25:0.5 (v/v) methanol/water/acetic acid immediately before ESI-MS analysis.

To identify the poor substrates, the pY library (2 mg) was dissolved in 10 mL of 1.5 mM ammonium bicarbonate (pH 7.0) and treated with SHP-1(ΔSH2) (60 μg) overnight at room temperature. The reaction mixture was frozen in liquid nitrogen and lyophilized into a powder, which was analyzed as described above.

Mass Spectrometry. The mass spectra of the peptide library were acquired on a Bruker Apex Ile 7-tesla Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bellerica, MA) under the positive ion mode. Samples were infused in an external microelectrospray source (*23*) at a rate of 15 μL/h. Typical experimental conditions at ESI source were the following: needle voltage at 2000−2300 V, capillary exit at 200 V, skimmer at 5 V, and heated capillary

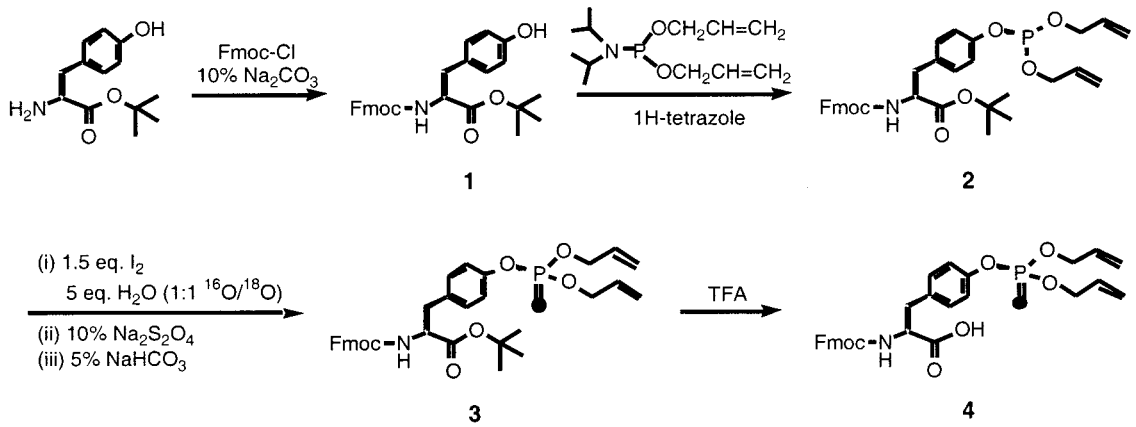

FIGURE 2: Synthesis of partially ^{18}O-labeled pY building block **4**.

temperature at 85 °C. The experiment event sequence and their corresponding parameters were described below. Ions were accumulated in a linear hexapole ion trap for 1.0 s and then transferred to a 3-in. Penning trap with 1.5 V trapping voltage. The broadband ion cyclotron excitation was conducted by chirp mode with an excitation amplitude of 1.0 V. A MIDAS data station (*24*) was used to control the experiment event sequence and to acquire data. Typically, 1 M broadband time-domain data were collected repetitively by ~1000 acquisitions. Hamming apodization was applied to the data before it was processed by fast Fourier transform to yield the FTICR MS spectrum. The product peaks were identified by manual inspection of the expanded spectrum and searching for the presence of singlet peaks.

MS/MS spectra of product ions were acquired on a Micromass Q-TOF2 mass spectrometer (Manchester, UK). The enzyme-treated peptide library was introduced into spectrometer by microspray at a rate of ~20 μL/h. A 1-Da window was used to isolate the product ions, which were fragmented through collision-induced dissociation (CID) with argon. The fragment ions were typically accumulated over a period of 10 min to yield the desired CID spectra.

PTP Assay. Assays were performed with synthetic pY peptides as substrates in a quartz microcuvette. A typical reaction contained 100 mM Hepes (pH 7.4), 100 mM NaCl, 2 mM EDTA, 5 mM tris(carboxyethyl)phosphine, and 0−1 mM pY peptide (final reaction volume of 108 μL). The reaction was initiated by the addition of SHP-1 (final concentration 41 nM), and the reaction progress was monitored continuously at 282 nm on a UV−Vis spectrophotometer. The initial rates were calculated from the early regions of the curves (<60 s) and fitted to the Michaelis−Menten equation to obtain the k_{cat}/K_M values. Determination of k_{cat} and K_M values was not possible because none of the pY peptides tested reached saturation at 1 mM substrate concentration.

RESULTS

Library Design and Synthesis. A 361-member peptide library RNNXXpYA-NH_2 (X = Nle or any of the 18 natural amino acids except for Cys and Met), in which the two residues immediately N-terminal to pY were randomized, was designed to evaluate the contribution of the −1 and −2 residues to PTP substrate recognition and catalysis. Methionine was replaced with norleucine (M) to avoid any oxidation at the side chain during experiments. The N-terminal RNN motif was added to improve the solubility of the library peptides. The arginine would also enhance the ionization efficiency of the peptides during MS analysis. An alanine was placed at the +1 position to avoid any unfavorable interactions between a peptide and the PTP active site. To facilitate library screening by mass spectrometry, the pY residue was partially labeled with ^{18}O at one of the nonbridging phosphate oxygen atoms ($^{16}O/^{18}O$ = 1:1) (Figure 1).

The isotopically labeled Fmoc-pY **4** was synthesized in four steps from commercially available *t*-butyl tyrosinate as detailed under Materials and Methods (Figure 2). The peptide library was constructed from Fmoc-pY **4** and other amino acids by the split-pool synthesis method (*20*−*22*). This method ensures that all library members are present at approximately the same concentration.

MS Analysis of Untreated Library. Theoretically, the peptide library should contain 361 distinct peptides with a molecular mass range between 830.3 and 1090.4 Da. Due to mass degeneracy (e.g., peptides RNNDFpYA-NH_2 and RNNFDpYA-NH_2 share the same exact mass), however, these peptides have only 141 unique masses. Analysis of the untreated library on a 7T FTICR mass spectrometer (average resolving power ~100,000 at *m*/*z* 900) indicated that virtually all of the peptides with distinct masses were well resolved in the spectrum. Because of the isotopic enrichment with ^{18}O, each peptide appears as a doublet of equal intensity (*m*/*z* m and m+2) (Figure 3a). Manual inspection of the spectrum revealed a total of >130 unique doublet peaks, representing >90% of all possible signals. The observed difference in peak intensity was likely due to different ionization efficiencies, the presence of isobaric peptides, and the uneven amplitude of the excitation pulse. A few singlet peaks of relatively low intensities were also visible in the spectrum. These were caused by background hydrolysis of the pY side chain during library construction and screening. Fortunately, the extent of background hydrolysis was small as compared to PTP-mediated hydrolysis, so that the screening was not adversely affected.

Identification of Optimal Substrates of SHP-1. The peptide library was completely dissolved in an ammonium carbonate buffer and treated with SHP-1(ΔSH2) for various lengths of time (3.5−40 min). Since all of the peptides in the library were present at the approximately same concentration, the most preferred substrates should react first. Further, the enzymatic reaction was carried out under very dilute condi-

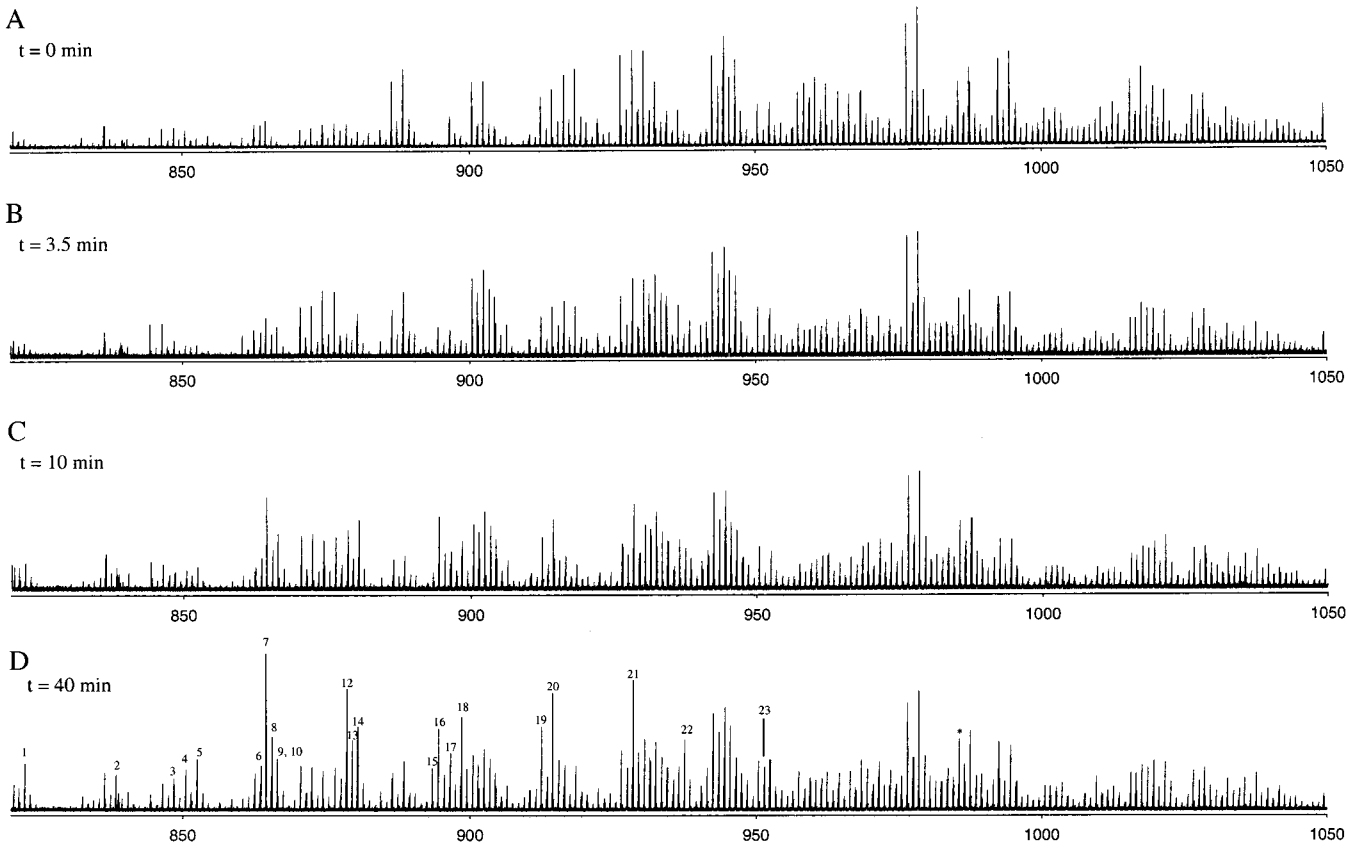

FIGURE 3: ESI−FTICR mass spectra (positive ion mode) of the 361-member pY peptide library before enzymatic treatment (A) and after treatment with SHP-1 (8.0 μg) for 3.5 (B), 10 (C), and 40 min (D). Peaks numbered 1−23 (according to the numbering in Table 1) indicate the most prominent product peaks identified.

tions (the concentration of each peptide was ~0.3 μM); therefore, SHP-1 was likely operating under the k_{cat}/K_M conditions for most of the substrates (vide infra). Dephosphorylation by SHP-1 removed the isotopic label, resulting in products as singlet peaks in the mass spectrum, which can be readily identified and sequenced by tandem mass spectrometry.

MS analysis of the SHP-1-treated library revealed a total of 23 singlets of significant intensities at various *m*/*z* values (Table 1 and Figure 3B−D). Most of these peaks were already clearly visible after 3.5 min of PTP treatment. In contrast, a few weak singlets appeared only after 40 min of reaction; these peaks were not further studied. The 23 singlets were verified as products by the observed loss/reduction of the corresponding substrate peaks at *m*/*z* m+80 and m+82. Their possible amino acid compositions were inferred from the observed accurate molecular masses. Their actual sequences were then determined by MS/MS analysis on an ESI Q-TOF instrument, which has higher sensitivity than the FTICR spectrometer. For example, a strong singlet peak of *m*/*z* 878.458 emerged in the enzyme treated library (Figure 4A and Table 1, entry 12). The molecular mass of this product is consistent with peptides containing sequence EZ (Z = leucine, isoleucine, or norleucine) or ZE at the randomized positions. MS/MS experiments indicated that the major product had the sequence RNNEZYA-NH_2 (Figure 4B). The intensities of the fragment ions corresponding to sequence RNNZEYA-NH_2 were much weaker and not significantly above the noise level. Unfortunately, due to mass degeneracy, it was not possible to determine whether all three sequences (EL, EI, and EM) were present or in what proportions they were present in the product. Note that due to the presence of an arginine at the N-termini of the peptides, the b fragment ions dominated the CID spectrum, significantly simplifying the spectral interpretation. All together, good MS/MS spectra were obtained for 16 product peaks. These spectra suggested that the major products all had either

Table 1: Most Abundant Products from the SHP-1-Treated Library

entry no.	sequence[a]	observed *m*/*z*	MS/MS[b]	rank[c]
1	DA	822.390	+	2
2	EA	836.408	+	3
3	DP	848.408	+	3
4	DV	850.425	+	2
5	DT	852.403	+	3
6	EP	862.424	−	3
7	DZ/EV	864.442	+	2
8	DN	865.397	+	3
9	DD	866.381	−	1
10	ET	866.419	−	3
11	SF/FS	870.428	−	3
12	EZ	878.458	+	2
13	EN/DQ	879.416	+	3
14	DE	880.402	+	1
15	EQ	893.428	−	3
16	EE	894.412	+	1
17	FZ/ZF	896.481	−	3
18	DF	898.424	+	2
19	EF	912.440	+	3
20	DY	914.420	−	3
21	EY	928.439	+	3
22	DW	937.435	+	2
23	EW	951.453	+	3

[a] Z, leucine, isoleucine, or norleucine. [b] +, MS/MS performed; −, no MS/MS performed. [c] 1−3, classes 1−3 peptides.

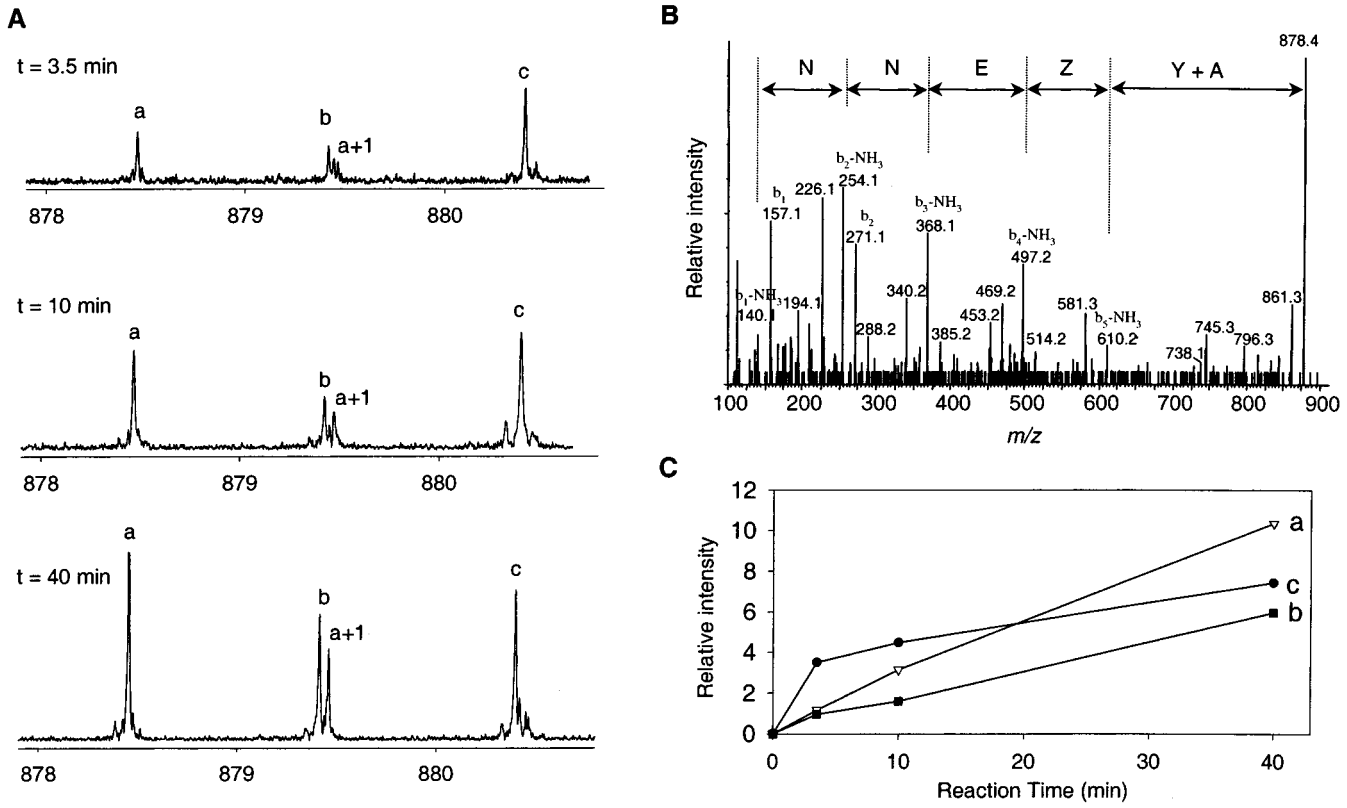

FIGURE 4: (A) Expanded spectra for the *m*/*z* 878−881 region in Figure 3, showing the appearance of product peaks with increasing SHP-1 reaction time. a, peptide RNNDZYA-NH_2; b, peptide RNNDQ(EN)YA-NH_2; c, peptide RNNDEYA-NH_2. Peak a+1 is due to the existence of natural isotopes in a (e.g., ^{13}C). (B) A Q-TOF MS/MS spectrum of the product peak a (*m*/*z* 878.4). (C) Plot of peak intensities for products a−c in part A as a function of time (reaction progress curves). Z, Leu, Ile, or Nle.

aspartic or glutamic acid at the −2 position (Table 1). For the DD peptide, no MS/MS experiment was necessary. For the rest of peptides (DY/YD, EP/PE, ET/TE, EQ/QE, SF/FS, and FZ or ZF), weak product signals and/or the presence of other intense peaks of similar *m*/*z* values prevented unambiguous sequence assignment. On the basis of the consensus of the sequenced products, we assume that the former products were DY, EP, ET, and EQ, respectively. Therefore, the consensus sequence for the catalytic domain of SHP-1 is (D/E)XpY (X = any amino acid except for Arg, Lys, His, or Gly). This is consistent with the previous observation that PTPs generally prefer acidic over basic residues at positions N-terminal to pY (*6*−*11*).

Identification of Poor Substrates of SHP-1. The disfavored substrates were identified by treating the pY library with excess enzyme for extended periods of time and analyzing the remaining doublet peaks in the mass spectrum (data not shown). Fifty-two sequences (27 doublet peaks) showed no detectable product formation after overnight incubation and therefore represent the least reactive substrates of SHP-1 (class I in Table 2). Each of these sequences contains at least one basic residue (arginine or lysine) at the randomized positions. Another 35 sequences (20 doublet peaks) showed only very small product peaks after overnight incubation with SHP-1. These sequences were also considered as poor substrates of SHP-1 (class II in Table 2). Again, these sequences have basic (e.g., Lys, Arg, and His) or polar, neutral residues (e.g., Gln, Asn, Thr) at one of the randomized positions. Therefore, SHP-1 strongly disfavors basic residues at the −1 and −2 positions.

Table 2: Poor Substrates of SHP-1

class I[a]		class II[a]	
RG/GR	KG/GK	KA/AK	PP
RA/AR	KS/SK	KP/PK	PN/NP
RS/SR	KV/VK	KZ/ZK	PW/WP
RV/VR	KT/TK	KQ/QK	NN
RT/TR	KN/NK	RP/PR	TT
RZ/ZR	KD/DK	HA/AH	QQ
RN/NR	KK	HP/PH	YV/VY
RD/DR	KE/EK	HQ/QH	YT/TY
RQ/QR	KH/HK	HH	YN/NY
RK/KR	KF/FK	HF/FH	YQ/QY
RE/ER	KY/YK		
RH/HR	KW/WK		
RF/FR	RW/WR		
RR			

[a] Class I, pY peptides that showed no detectable product peak after overnight incubation; Class II, pY peptides that showed small product peaks after the incubation. Z, Leu, Ile, or Nle.

Rank Order of Selected SHP-1 Substrates. The analysis described above revealed that SHP-1 generally prefers peptides of the consensus sequence (D/E)XpY. However, it does not provide further information about the relative reactivity of the 23 selected sequences. This is because the intensity of a peak in an MS spectrum depends on many factors including sample concentration, ionization efficiency of the molecule, and the electrospray conditions. Therefore, a product peak of the highest intensity does not necessarily indicate that the corresponding substrate is the best substrate. To rank order the 23 peptides, the time course for their formation was monitored. Thus, the pY peptide library was treated with SHP-1 for varying lengths of time (3.5, 10, and

40 min) before it was analyzed by FTICR-MS. Comparison of the spectra from different reaction times showed that for most of the singlet peaks, their intensity increased with time (Figure 3A−D). Next, the intensity of each singlet peak was normalized relative to an internal standard and plotted against time to generate its reaction progress curve. In this work, peptide RNNZR(RZ)pYA-NH_2 (*m*/*z* 985.473) was used as the internal standard since it had no detectable reaction during the experimental period. Since all of the peptides were present in the initial library at approximately the same concentration, the shape of the reaction progress curve provides an indicator of the potency of that substrate. For example, the intensity of peptide RNNDEYA-NH_2 (*m*/*z* 880.402) approached the plateau value at 10 min, whereas the intensity of peptides RNNEZYA-NH_2 (*m*/*z* 878.458) and RNNEN(DQ)YA-NH_2 (*m*/*z* 879.416) further increased substantially after 10 min (Figure 4A,C). This indicates that at 10 min, peptide RNNDEpYA-NH_2 was almost completely reacted, whereas RNNEZpYA-NH_2 and RNNEN(DQ)pYA-NH_2 were not, suggesting the former as the better substrate. Note that even after 40 min, no significant product signal was observed for sequences QN/NQ (*m*/*z* 878.423), KN/NK (*m*/*z* 878.460), GW/WG (*m*/*z* 879.422), DK/KD (*m*/*z* 879.444), SR/RS (*m*/*z* 879.455), or FP/PF (*m*/*z* 880.443) in the expanded region (Figure 4A), indicating that these peptides are very poor substrates of SHP-1.

The reaction progress curves of all 23 substrates are illustrated in Figure 5. It is apparent from the strong curvature of their reaction progress curves that peptides containing acidic residues at both −1 and −2 positions (e.g., DD, DE, and EE) are the most preferred substrates of SHP-1 (class 1). Furthermore, DD and DE appear to be slightly more reactive than EE. The next group of peptides (class 2) contain mostly aspartate at the −2 position and hydrophobic residues at the −1 position (Z, F, W, V, A) (Table 1). Their reaction progress curves exhibited some curvature, indicating that a significant fraction of these substrates was reacted during screening. The majority of class 3 peptides contained an E at the −2 position and various amino acids at the −1 position. Their reaction progress curves are essentially linear over the 40-min period. Overall, it appears that SHP-1 has a slight preference for an aspartate over glutamate at the −2 position.

Kinetic Analysis of Selected pY Peptides. Five of the selected peptides (with sequences DE, DD, EE, DF, and EF) were individually synthesized and tested against SHP-1(ΔSH2). These peptides showed k_{cat}/K_M values in the range of 1.6−4.3 × 10^4 M^{-1} s^{-1} (Table 3). These are among the most efficient peptide substrates known for SHP-1 (*11*, *25*). The kinetic data are consistent with the ranking order predicted by the time course experiments. For example, the data confirmed our prediction that SHP-1 slightly prefers an aspartate to glutamate at the −2 position (i.e., DE > EE, DF > EF). The MS/MS spectrum for peptide DF/FD showed that the predominant species was DF (fragment ion ratio was 6:1 for DF/FD) (data not shown). We thus synthesized the FD sequence (RNNFDpYA-NH_2) and found that the DF sequence was indeed 2-fold more reactive than the FD sequence (Table 3). Peptide RNNTQpYA-NH_2, which showed a very weak product peak at *m*/*z* 865.422 after 40 min, had a k_{cat}/K_M value of 8670 M^{-1} s^{-1}. Two of the predicted poor substrates, RNNGKpYA-NH_2 and RNNKKpYA-NH_2, had k_{cat}/K_M values of 640 and 460 M^{-1} s^{-1}, respectively. Therefore, sequence variation at the −1 and −2 positions results in at least 100-fold difference in activity. To gain further insight into the cause of this difference, we

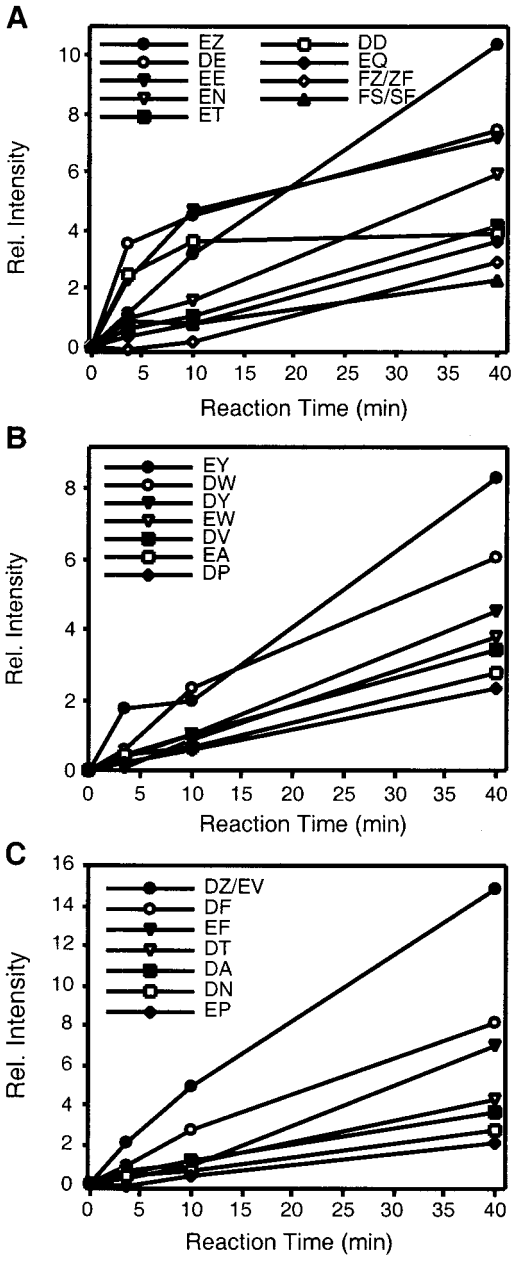

FIGURE 5: Reaction progress curves for the 23 most abundant SHP-1 products.

Table 3: Kinetic Properties of pY Peptides

peptide	class	k_{cat}/K_M ($\times 10^3$ $M^{-1}s^{-1}$)[a]
RNNDEpYA-NH_2	1	42.7 ± 3.4
RNNDDpYA-NH_2	1	38.1 ± 3.3
RNNEEpYA-NH_2	1	22.4 ± 1.1
RNNDFpYA-NH_2	2	21.9 ± 1.2
RNNFDpYA-NH_2		11.3 ± 1.6
RNNEFpYA-NH_2	3	16.5 ± 0.9
RNNTQpYA-NH_2		8.67 ± 0.48
RNNAApYA-NH_2		2.02 ± 0.08
RNNKKpYA-NH_2	I (poor)	0.46 ± 0.09
RNNGKpYA-NH_2	I (poor)	0.64 ± 0.07

[a] Data reported are the mean ± SD from at least three independent sets of experiments.

synthesized an all-alanine peptide, RNNAApYA-NH_2, which had a k_{cat}/K_M value of 2020 M^{-1} s^{-1}. These data suggest that a good sequence combination promotes the formation of productive E·S complex and catalysis, whereas a poor sequence interferes with binding to the enzyme active site and slows down the reaction.

DISCUSSION

PTPs were initially thought as promiscuous "housekeeping" enzymes that simply oppose the action of PTKs. Recent work indicates that PTPs comprise a large family whose members play active roles in a wide variety of cellular processes and that PTPs exhibit exquisite substrate specificity in vivo. There is now growing evidence that PTPs achieve their impressive in vivo selectivity through combinations of specific targeting strategies (e.g., SH2 domains) and intrinsic specificity of their catalytic domain (*3*). Thus, information on the substrate sequence specificity (intrinsic specificity) of PTPs will help identify their physiological substrates and understand the cellular functions of these enzymes. In addition, the specificity data will facilitate the design of specific inhibitors against the PTPs. Such inhibitors would provide potential therapeutic agents as well as invaluable tools for deciphering the cellular processes in which PTPs are involved. Optimal substrates would also provide more effective kinetic assays for these enzymes. Therefore, in recent years, considerable efforts have been made by many laboratories to define the substrate specificity of PTPs using both combinatorial approaches and by assaying individual pY peptides (6−11, 14−16). However, as discussed above, the existing methods each suffer from certain drawbacks and consequently, there is only limited knowledge about the specificity of a few PTPs.

In this work, we have applied the ECLIPSE method, which was recently developed by our laboratories, to determine the substrate specificity of SHP-1 at the −1 and −2 positions. Screening of a 361-member pY peptide library revealed that the catalytic domain of SHP-1 has a consensus sequence of (D/E)XpY (X = any amino acid other than Arg, Lys, His, or Gly), with the highest activity toward peptides containing aspartate or glutamate at both −1 and −2 positions. The selected peptides are among the most efficient substrates reported so far for SHP-1 (*11*, *25*). Since SHP-1 can make contacts with 3−5 residues on either side of the pY residue (*13*), we expect that still more efficient substrates will be found by optimizing the sequences at other positions. The ECLIPSE method also identifies the least reactive substrates. For SHP-1, the poorest substrates generally have basic residues at one or both of these positions. Many PTPs have basic residues near their active sites; the positively charged peptides are expected to have unfavorable electrostatic interactions with the enzyme active/binding site. Among the 361 peptides analyzed, there is at least a 100-fold difference in reactivity (compare peptides DE vs KK in Table 3). This difference is likely to be much larger if the other important positions are optimized. Thus, SHP-1 has relatively strong sequence selectivity toward pY peptides. Interestingly, the substrate specificity of SHP-1 at the −1 and −2 positions is similar to that of a number of other PTPs, such as PTP1B (*15*, *16*), LAR (*14*, *26*), *Yersinia* PTP (*27*, *28*), and PTP1 (*28*). Like PTP1B, SHP-1 also demonstrates a remarkable degree of plasticity in accommodating both acidic and large hydrophobic residues at the −1 position (*29*). The similar specificity at the −1 and −2 positions suggests that different PTPs perhaps recognize their in vivo substrates by engaging in specific interactions at other positions.

The ECLIPSE method offers several advantages over other literature methods. First, ECLIPSE does not require the exclusion of any amino acid from the library; therefore, it provides a truly systematic and unbiased assessment of PTP specificity. Cysteine was excluded from the library used in this work to simplify sample handling. However, libraries containing all 20 natural amino acids (as well as unnatural amino acids) can be easily screened by this method in the presence reducing agents. Second, screening by ECLIPSE is carried out directly with the enzyme of interest in the solution phase, eliminating any potential impact of the solid support, encoding tags, or coupling enzymes on the enzyme specificity. Therefore, whichever product(s) is formed first reveals the identity of the most preferred substrate(s), which can be further rank ordered according to their catalytic constants (k_{cat}/K_M) through the time-course experiments. Another unique feature of this method is that it can also identify the most disfavored substrates of an enzyme. Such information will be very useful in ruling out certain proteins as potential physiological substrates of a PTP, as a poor substrate in vitro (e.g., due to structural incompatibility with the enzyme active site) is most likely also a poor substrate in the cell. Finally, ECLIPSE is a general method, applicable to any enzyme or synthetic catalyst that catalyzes the removal, addition, or substitution of a functional group from/onto a substrate.

Perhaps one might ask whether isotopic labeling is necessary in this method (or why not simply compare the starting and the enzyme-treated libraries and look for new peaks formed?). Benner and co-workers have indeed used FTICR-MS to screen a very small library (16 members) for the optimal substrate of glutathione-*S*-transferase, by directly comparing the mass spectra of enzyme-treated and untreated libraries and identifying reaction products by the appearance of new peaks in the enzyme-treated spectrum (*30*). While simple to perform, their method is limited to small libraries in which none of the substrates are isobaric with the product(s). For larger libraries, there is a much greater probability of product peaks overlapping with substrate peaks; separation into smaller sublibraries would be necessary prior to MS analysis. However, we are not aware of any current separation technique that is capable of fractionating a complex combinatorial library into exactly the same fractions in two separate runs. This would make it difficult to directly compare the spectra of starting and enzyme-treated libraries. Our method has no such limitation. A product peak completely overlapped with a substrate peak can still be identified by the uneven intensities of the doublet. For example, in our previous work of screening a peptide deformylase library, products MDR and MER shared the same molecular formulas with substrates formyl-MSR and formyl-MTR, respectively (*17*). The two products were still unambiguously identified by ECLIPSE. Furthermore, since ECLIPSE identifies a product based on its peak shape (singlet vs doublet), comparison with the starting library is unnecessary. Consequently, ECLIPSE should be capable of screening much larger libraries by interfacing the mass spectrometer with various separation devices (e.g., HPLC). Its unique feature

for product identification should also greatly facilitate the automation of this screening process.

Some precautions should be taken to obtain optimal results by the ECLIPSE method and to properly interpret the results. First of all, all of the library members must be soluble during the course of experiments. Fortunately, mass spectrometry is one of the most sensitive analytical methods and ECLIPSE screening is usually performed under highly dilute conditions to ensure that the enzyme is operating under the k_{cat}/K_M conditions (e.g., each peptide was present at ~0.3 μM in this work). Following the enzymatic reaction, the sample can be concentrated and dissolved in solutions containing organic solvents. To ensure that all of the compounds have the minimal solubility necessary for screening, each library member can be derivatized with a positively charged, hydrophilic moiety (e.g., RNN). The positive charge also provides higher as well as more uniform ionization efficiency for the library members during mass spectrometric analysis in the positive ion mode. Negatively charged group may also be used for MS analysis in the negative ion mode. In interpreting the MS data, one must keep in mind that an intense product peak does not necessarily indicate the corresponding substrate as a good substrate. Instead, time-course experiments should be performed to rank order the substrates in a semiquantitative fashion. Like any other combinatorial library methods, ECLIPSE may miss certain sequences if they are difficult to synthesize (therefore not present in the library) or have extremely poor ionization efficiency. However, even if all of the preferred substrates were not identified, the overall consensus sequence would likely stay the same. Finally, it should be noted that ECLIPSE only determines the primary sequence specificity of an enzyme. Other factors such as secondary structures and localization that may influence the in vivo specificity of a PTP cannot be addressed by this method.

In conclusion, we have demonstrated ECLIPSE as an effective method to systematically evaluate the substrate specificity of PTPs. The results show that the catalytic domain of SHP-1 is rather selective for peptides of the consensus sequence (D/E)XpY. This method should be widely applicable to a variety of other enzymes. Its application to other enzymes and more diverse libraries to assess the specificity of SHP-1 at other subsites is already under way in our laboratories.

REFERENCES

1. Hunter, T. (2000) *Cell 100*, 113−127.
2. Neel, B. G., and Tonks, N. K. (1997) *Curr. Opin. Cell Biol. 9*, 193−204.
3. Tonks, N. K., and Neel, B. G. (2001) *Curr. Opin. Cell Biol. 13*, 182−195.
4. Tenev, T., Keilhack, H., Tomic, S., Stoyanov, B., Stein-Gerlach, M., Lammer, R., Krivtsov, A. V., Ullrich, A., and Bohmer, F. D. (1997) *J. Biol. Chem. 272*, 5866−5973.
5. O'Reilly, A. M., and Neel, B. G. (1998) *Mol. Cell Biol. 18*, 161−177.
6. Cho, H., Ramer, S. E., Itoh, M., Winkler, D. G., Kitas, E., Bannwarth, W., Burn, P., Saito, H., and Walsh, C. T. (1991) *Biochemistry 30*, 6210−6216.
7. Zhang, Z.-Y., Maclean, D., McNamara, D. J., Sawyer, T. K., and Dixon, J. E. (1994) *Biochemistry 33*, 2285−2290.
8. Harder, K. W., Owen, P., Wong, L. K. H., Aebersold, R., Clark-Lewis, I., and Jirik, F. R. (1993) *Biochem. J. 298*, 395−401.
9. Bobko, M., Wolfe, H. R., Saha, A., Dolle, R. E., Fisher, D. K., and Higgins, T. J. (1995) *Bioorg. Med. Chem. Lett. 5*, 353−356.
10. Zhang, Z.-Y., Thieme-Sefler, A., M., Maclean, D., McNamura, D. J., Dobrusin, E. M., Sawyer, T. K., and Dixon, J. E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 4446−4450.
11. Dechert, U., Affolter, M., Harder, K. W., Matthews, J., Owen, P., Clarke-Lewis, I., Thomas, M. L., Aebersold, R., and Jirik, F. R. (1995) *Eur. J. Biochem. 231*, 673−681.
12. Jia, Z., Barford, D., Flint, A. J., and Tonks, N. K. (1995) *Science 268*, 1754−1758.
13. Yang, J., Cheng, Z., Niu, T., Liang, X., Zhao, Z. J., and Zhou, G. W. (2000) *J. Biol. Chem. 275*, 4066−4071.
14. Cheung, Y. W., Abell, C., and Balasubramanian, S. (1997) *J. Am. Chem. Soc. 119*, 9568−9569.
15. Huyer, G., Kelly, J., Moffat, J., Zamboni, R., Jia, Z., Gresser, M. J., and Ramachandran, C. (1998) *Anal. Biochem. 258*, 19−30.
16. Pellegrini, M. C., Liang, H., Mandiyan, S., Wang, K., Yuyev, A., Vlattas, I., Sytwu, T., Li, Y.-C., and Wennogle, L. P. (1998) *Biochemistry 37*, 15598−15606.
17. Wang, P., Snavley, D. F., Freitas, M. A., and Pei, D. (2001) *Rapid Commun. Mass Spectrom. 15*, 1166−1171.
18. Pei, D., Neel, B. G., and Walsh, C. T., (1993) *Proc. Natl. Acad. Sci. U.S.A. 91*, 1092−1096.
19. Zhang, Z.-Y., Maclean, D., Thieme-Sefler, A. M., Roeske, R. W., and Dixon, J. E. (1993) *Anal. Biochem. 211*, 7−15.
20. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) *Nature 354*, 82−84.
21. Furka, A., Sebestyen, F., Asgedom, M., and Dibo, G. (1991) *Int. J. Peptide Protein Res. 37*, 487−493.
22. Houghten, R. A., Pinilla, C., Blondelle, S. E., Appel, J. R., Dooley, C. T., and Cuervo, J. H. (1991) *Nature 354*, 84−86.
23. Emmett, M. R., and Caprioli, R. M., (1994) *J. Am. Soc. Mass Spectrom. 5*, 605−613.
24. Senko, M. W., Canterbury, J. D., Guan, S., and Marshall, A. G., (1996) *Rapid Commun. Mass Spectrom. 10*, 1839−1844.
25. Pei, D., Lorenz, U., Klingmuller, U., Neel, B. G., and Walsh, C. T. (1994) *Biochemistry 33*, 15483−15493.
26. Cho, H., Krishnaraj, R., Itoh, M., Kitas, E., Bannwarth, W., Saito, H., and Walsh, C. T. (1993) *Protein Sci. 2*, 977−984.
27. Zhang, Z.-Y., Maclean, D., McNamara, D. J., Sawyer, T. K., and Dixon, J. E. (1994) *Biochemistry 33*, 2285−2290.
28. Zhang, Z.-Y., Thieme-Sefler, A. M., Maclean, D., McNamara, Dobrusin, E. M., D. J., Sawyer, T. K., and Dixon, J. E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 4446−4450.
29. Sarmiento, M., Puius, Y. A., Vetter, S. W., Keng, Y.-F., Wu, L., Zhao, Y., Lawrence, D. S., Almo, S. C., and Zhang, Z.-Y. (2000) *Biochemistry 39*, 8171−8179.
30. Wigger, M., Nawrocki, J. P., Watson, C, H., Eyler, J. R., and Benner, S. A. (1997) *Rapid Commun. Mass Spectrom. 11*, 1749−1754.

BI025591F