A Combinatorial Approach to Identifying Protein Tyrosine Phosphatase Substrates from a Phosphotyrosine Peptide Library

Yee Wah Cheung, Chris Abell, and
Shankar Balasubramanian*

*University Chemical Laboratory, Lensfield Road
Cambridge CB2 1EW, England*

Received June 4, 1997

Protein phosphorylation plays a pivotal role in the regulation of cellular functions such as growth and differentiation.[1] The study of the substrate recognition of protein kinases and protein phosphatases is critical to our understanding of protein phosphorylation and could yield valuable structural information to assist in the design of selective inhibitors to interfere with cellular signalling processes. Studies on protein tyrosine phosphatases (PTPs) using phosphotyrosyl peptide substrates have demonstrated some level of substrate specificity.[2,3] Such studies have largely been based on peptide sequences derived from phosphoproteins that include the autophosphorylation sites of the insulin receptor[4,5] and the epidermal growth factor receptor.[2,3,6] However, the necessity to synthesize and test single phosphopeptides limits the exploration of PTP specificities. A suitable combinatorial approach would provide a valuable means of rapidly identifying substrate consensus sequences for PTPs.

There are examples of the application of solution-phase combinatorial methods to the discovery of ligands that bind to a target receptor.[7] For a solid-phase library, ligand selection has been mediated by binding assays using a fluorescently labeled target protein[8] or by immunostaining the protein-bound beads.[9−11] The identification of good substrates for enzymatic turnover requires the discrimination of solid-phase products of catalysis from solid-phase substrates, with single-bead sensitivity. This has been demonstrated in the screening of peptide libraries with endopeptidases by a quenched fluorescence assay[12] and also in the detection of protein kinase substrates by the incorporation of radioactive ^{32}P.[13] Immunodetection of solid-phase products has been employed in the selection of catalytic antibodies.[14] We report a novel combinatorial approach for identifying substrates for PTPs from a resin-bound library of phosphotyrosyl peptides.

The assay requires the discrimination between tyrosyl and phosphotyrosyl peptides on-bead. It is well-known that the serine protease, α-chymotrypsin (CT), is selective for the cleavage of amide bonds on the C-terminal side of aromatic residues, in particular tyrosine.[15] The selectivity of CT is primarily due to the presence of a hydrophobic binding pocket adjacent to the scissile bond. We proposed that a phosphotyrosine (pY) residue would bind poorly in this pocket providing an enzymatic discrimination between a tyrosyl- and phosphotyrosylpeptide. This kinetic discrimination was studied by comparing the on-bead rates of CT-mediated cleavage of an N-dabcyl-labeled epidermal growth factor receptor [$EGFR_{988-998}$] peptide, DADEpYLIPQQG, for both phosphorylated and unphosphorylated tyrosine. The discrimination in favor of unphosphorylated peptide was greater than a factor of 100 for CT,[16] sufficient to be exploited as the basis for detecting positives in a combinatorial screen.

The choice of resin for an on-bead enzyme assay is important. Kieselguhr and PEGA resins gave excellent substrate turnover efficiencies for CT, PTP, and proteins as large as 140 kDa,[17] whereas Tentagel was found to be unsuitable owing to very poor enzyme accessibility. Phosphotyrosyl peptides were synthesised on amino-functionalized Kieselguhr (KD) resin, using the unprotected phosphotyrosine monomer.[18,19]

The screening strategy is outlined in Figure 1. A solid-phase phosphopeptide library is first synthesized in which the pY position is doped with ∼30% glycine to generate an encoding strand. The N-termini are left capped with Fmoc protecting groups (**A**, Figure 1, Fmoc = (fluorenylmethoxy)carbonyl). The library is subjected to on-bead dephosphorylation with the chosen PTP resulting in the transformation of substrates to tyrosyl peptides (**B**, Figure 1). Following incubation with PTP, the library is subjected to proteolytic cleavage with CT, which is selective for those sequences that have undergone PTP-mediated dephosphorylation in the previous step. The encoding strand, which lacks tyrosine, remains intact throughout. The free amino termini that result from dephosphorylated sequences are coupled to an amine-reactive fluorophore or chromophore to locate the positive beads. Removal of the terminal Fmoc group from the encoding strands of positive beads permits facile sequence identification by standard peptide sequencing.

To exemplify this methodology, we have studied the substrate specificity of leukocyte antigen receptor (LAR) phosphatase.[20,21] The undecapeptide corresponding to the autophosphorylation site of the epidermal growth factor receptor ($EGFR_{988-998}$, Figure 2) has been shown to be a good substrate for *Yersinia* and mammalian PTPases[3] and was used as a prototype sequence for our combinatorial library. $EGFR_{988-998}$ contains three acidic residues on the N-terminal side of the phosphorylation site. It

(1) Johnson, L. N.; Barford, D. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 199−232.
(2) Zhang, Z. Y.; Maclean, D.; McNamara, D. J.; Sawyer, T. K.; Dixon, J. E. *Biochemistry* **1994**, *33*, 2285−2290.
(3) Zhang, Z.-Y.; Theme-Sefler, A. M.; Maclean, D.; McNamara, D. J.; Dobrusin, E. M.; Sawyer, T. K.; Dixon, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4446−4450.
(4) Cho, H.; Ramer, S. E.; Itoh, M.; Kitas, E.; Bannwarth, W.; Burn, P.; Saito, H.; Walsh, C. T. *Biochemistry* **1992**, *31*, 133−138.
(5) Ramachandran, C.; Aebersold, R.; Tonks, N. K.; Pot, D. A. *Biochemistry* **1992**, *31*, 4232−4238.
(6) Zhang, Z.-Y.; Dixon, J. E. *Adv. Enzymol.* **1994**, *68*, 1−35.
(7) Houghten, R. A.; Pinilla, C.; Blondelle, S. E.; Appel, J. R.; Dooley, C. T.; Cuervo, J. H. *Nature* **1991**, *354*, 84−86.
(8) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82−84.
(9) Buettner, J. A.; Dadd, C. A.; Baumbach, G. A.; Masecar, B. L.; Hammond, D. J. *Int. J. Pept. Protein Res.* **1996**, *47*, 70−83.
(10) Yu, Z. G.; Chu, Y. H. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 95−98.
(11) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. *J. Am. Chem. Soc.* **1995**, *117*, 3900−3906.
(12) Meldal, M.; Svendsen, I.; Breddam, K.; Auzanneau, F. I. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3314−3318.
(13) Wu, J. Z.; Ma, Q. Y. N.; Lam, K. S. *Biochemistry* **1994**, *33*, 14825−14833.
(14) Tawfik, D. S.; Green, B. S.; Chap, R.; Sela, M.; Eshhar, Z. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 373−377.
(15) Kraut, J. *Annu. Rev. Biochem.* **1977**, *46*, 331−358.
(16) Phosphorylated and unphosphorylated $EGFR_{988-998}$ were synthesised on Kieselguhr resin with an N-terminal chromophore, 4-(dimethylamino)-azobenzene-4'-carboxylic acid (dabcyl) attached. CT cleavage of both substrates, under identical conditions, was monitored by release of the dabcylated fragment into solution phase. Only background cleavage (enzyme independent) was detectable for the phosphorylated substrate, whereas the unphosphorylated substrate cleaved 100-fold faster, based on initial rates. A 10-fold increase in CT resulted in no increase in rate for the phosphorylated substrate.
(17) Near quantitative dephosphorylation of the resin-bound substrate could be achieved on Kieselguhr and PEGA resins with LAR PTP (35 kDa), compared to less than 2% with TentaGel resin. Gel filtration studies showed that the exclusion limit of TentaGel resin was below 14 kDa.
(18) Ottinger, E. A.; Shekels, L. L.; Bernlohr, D. A.; Barany, G. *Biochemistry* **1993**, *32*, 4354−4361.
(19) Peptide synthesis was carried out using Fmoc amino acids (5-fold excess except for pY which was 7.5-fold) and PyBOP/HOBt/DIPEA. Double-couplings were used for all residues after and including pY, and a capping step (20% Ac_2O) was at the end of the synthesis. Purity of individual phosphopeptides on resin was determined to be generally greater than 90% by amino acid analysis.
(20) Pulido, R.; Serrapages, C.; Tang, M.; Streuli, M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11686−11690.
(21) Cho, H. J.; Ramer, S. E.; Itoh, M.; Winkler, D. G.; Kitas, E.; Bannwarth, W.; Burn, P.; Saito, H.; Walsh, C. T. *Biochemistry* **1991**, *30*, 6210−6216.

S0002-7863(97)01825-8 CCC: $14.00

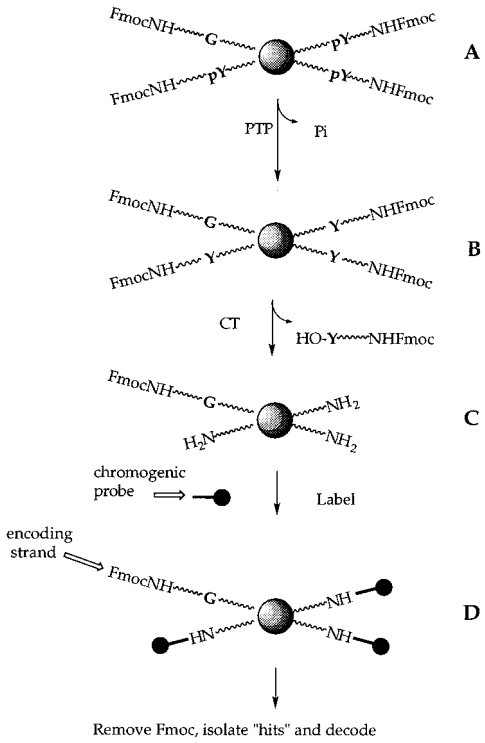

Figure 1. The screening strategy: (i) Typically 40 mg of N-terminal Fmoc-protected phosphopeptide library (**A**) is subjected to dephosphorylation by incubation with LAR PTP (200 units) in 30 mM Hepes (pH 7.4), 150 mM NaCl, 6.25 mM DTT in a total volume of 1 mL at 37 °C for 60 min, releasing typically ∼1% available Pi; (ii) dephosphorylated peptide (**B**) is selectively cleaved by incubation with chymotrypsin (100 units) in 20 mM Tris-HCl (pH 8.0), 160 mM NaCl in a total volume of 1 mL, at 20 °C for 45 min; (iii) cleaved peptide (**C**) is identified by coupling the free N-terminus to carboxyfluorescein using the coupling reagents PyBOP/HOBt/DIPEA; (iv) labeled beads (**D**) are isolated, the terminal Fmoc group is removed by treatment with 20% piperidine/DMF, and the peptide is sequenced.

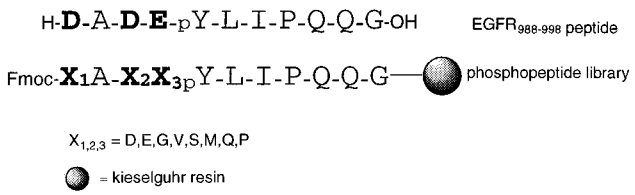

Figure 2. The bead-bound phosphopeptide library was based around the epidermal growth factor receptor, $EGFR_{988-998}$, sequence. The acidic side chains on the N-terminal side of pY were each substituted by one of eight amino acids (D, E, G, V, S, M, Q, P) to generate a 512-member library.

has been noted that several PTPs exhibit a general requirement for >4 residues on the N-terminal side of the phosphotyrosine, of which at least one should bear an acidic side-chain.[2,22] The phosphopeptide library was synthesised on Kieselguhr resin using a "split and mix approach,"[7] inserting one of eight possible amino acids in each of the positions designated X_1, X_2, and X_3 (Figure 2), to generate a 512-member library. The pool of variable amino acids was chosen to include acidic, hydrophobic, hydrophilic, large, and small side chains, in order to explore the requirement for carboxyl side chains in these positions for

(22) Hippen, K. L.; Jakes, S.; Richards, J.; Jena, B. P.; Beck, B. L.; Tabatabai, L. B.; Ingebritsen, T. S. *Biochemistry* **1993**, *32*, 12405−12412.

Table 1. Sequences of Seven Selected Positive Beads Isolated from Screening the Library with Leukocyte Antigen Receptor Protein Tyrosine Phosphatase

	residue at position		
sequence	X_1	X_2	X_3
1[a]	E	E	Q
2	E	D	Q
3	D	E	P
4	E	Q	P
5	E	D	E
6	E	Q	D

[a] Sequence appeared twice.

LAR PTP.[23] The resin-bound library was incubated with LAR PTP, to mediate the release of ∼1% of the total available phosphate (Pi), as detected by a malachite green assay.[24] It was then subjected to CT as depicted in Figure 1. Positive beads were stained by coupling any free amino termini to carboxyfluorescein. Analysis of the beads showed that ∼4% of the library was lightly labeled, with <1% of the beads heavily labeled.[25] Seven of the heavily stained beads were sequenced to give the results shown in Table 1.

Sequences 1−3 (Table 1) were resynthesized and assayed individually (resin-bound) with LAR PTP for release of Pi[24] against time. The release profile for all three sequences were comparable with that of $EGFR_{988-998}$, and furthermore, the initial rates of dephosphorylation were ∼30-fold greater than those for the sequence Fmoc-GAPGpYLIPQQG-KD, which has no acidic residues in positions X_1, X_2, or X_3 and did not label strongly under the screening conditions.

The combinatorial screening experiment generated six distinct substrate sequences for LAR PTP, each of which contained at least two acidic residues in the three variable positions. There also appears to be a preference for an acidic side chain derived from Glu rather than Asp in position X_1. These results are broadly in agreement with the substrate specificity studies by Zhang *et al*. on *Yersinia* PTPase and rat PTP1, which concluded that acidic residues on the N-terminal side of pY are important for binding and catalysis.[2]

In summary, we have demonstrated a novel combinatorial screening approach to identify substrate sequence motifs for PTPs. The main advantages over existing methods for studying PTPs include screening efficiency and the ability to identify individual sequences rather than simply a positional preference for particular residues. The methodology will be expanded to study and compare the specificities of a range of PTPs.

Acknowledgment. We thank Zeneca for a studentship to Y.W.C. and supplying purified LAR PTP. S.B. is a Royal Society University Research Fellow.

Supporting Information Available: Experimental protocols for phosphopeptide library synthesis and screening, kinetic studies of LAR PTP and CT against resin-bound substrates, and typical sequencing results for a "hit" and Pi release data (14 pages). See any current masthead page for ordering and Internet access instructions.

JA971825K

(23) A current limitation of this methodology is that the aromatic residues must be excluded from the library since CT will cleave adjacent to these residues.

(24) Lanzetta, P. A.; Alvarez, L. J.; Reinach, P. S.; Candia, O. A. *Anal. Biochem.* **1979**, *100*, 95−97.

(25) The intensity of staining was detected by eye. Although the chromogenic intensity increases with loading, the fluorescence actually drops above a critcial loading due to quenching from neighboring fluorophores, see: Scott, R. H.; Balasubramnian, S. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 1567−1572.