INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI®

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Specificity of SH2 Domains and Protein Tyrosine Phosphatases

Gregory Huyer

A Thesis in the

Department of Chemistry and Biochemistry

Presented in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy at

Concordia University

Montréal, Québec, Canada

June, 1997

ABSTRACT

Specificity of SH2 Domains and Protein Tyrosine Phosphatases

Gregory Huyer, Ph.D.

Concordia University, 1997

A multicellular organism must be able to coordinate the actions of its cells in order to function effectively. As such, cells have developed mechanisms to sense and respond to changes in their environment and signals from other cells. One common mechanism by which signals are transmitted within cells involves protein phosphorylation. For example, signal transmission from many extracellular signaling molecules is initiated by receptors that are themselves protein tyrosine kinases (PTKs) or have an associated PTK activity. SH2 domains and protein tyrosine phosphatases (PTPs) play essential roles in transmitting these signals. SH2 domains are small domains of ~100 amino acids that bind to phosphotyrosine (pY) in the context of adjacent amino acids, and PTPs counteract the activity of PTKs by removing phosphates from tyrosine residues. Many SH2 domains and PTPs exist, yet in spite of significant overlap in the intracellular machinery of tyrosine phosphorylation signaling pathways, specificity of the different extracellular signals is maintained.

The goal of the work presented in this thesis was to explore the basis of SH2 domain and PTP specificity. In the first part, the SH2 domains of the PTP SHP-2 were

characterized. The specificity of these SH2 domains was shown to involve two residues N-terminal to the target pY in addition to three C-terminal residues. A single residue of the SH2 domains was found to play a critical role in directing the specificity requirement for N-terminal residues. This unique binding specificity defines an additional class of SH2 domain binding with general implications for SH2 domain–protein interactions.

In the second part, PTP specificity was studied. First, the mechanism of inhibition of two non-specific PTP inhibitors, vanadate and pervanadate, was elucidated. While little was revealed regarding PTP specificity, different mechanisms of inhibition were uncovered for these inhibitors having important implications for their use. Second, kinetic studies with pY-peptide substrates were used to attempt to uncover specificity determinants for three transmembrane PTPs. Finally, a general peptide affinity selection technique that was initially explored with the SH2 domains of SHP-2 was refined to define the *in vitro* sequence selectivities of PTPs. This technology should aid in predicting potential *in vivo* targets for PTPs and in the development of specific PTP inhibitors.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Mike Gresser for creating the opportunity for me to complete my Ph.D. His understanding, patience, faith, and support made my studies an enjoyable and positive experience, helping me to rediscover that excitement for research and the self-confidence to succeed at it. None of this would have been possible without the constant support and patient assistance of Chidambaram Ramachandran, to whom I am deeply indebted and from whom I have learned so much. Thanks as well to Ann English for her supervision, enthusiasm, tireless proof-reading, and the sessions at Hurley's. I am also grateful to everyone at Merck Frosst who have contributed to my work and to making the time I spent here enjoyable: to Sylvie, Qingping, Kathryn, Deena, Cary, and Khalid of the phosphatase group; to Huang, for tearing apart my manuscripts and providing so many useful insights; to Brian, Paul, Mark, Yves, and Rino, for patiently helping me with my molecular biology forays; to Chris, for his help with the structural studies and for stimulating discussions; to Jim, Chun, and especially John, for teaching me everything I wanted to know about mass spectrometry but was afraid to ask; and to Bob, for letting me pretend to be an organic chemist again. I would also like to acknowledge the many students at Merck (too numerous to name) who helped to make it a fun place to be, and for the great friendships that were created. The understanding and support of my family and friends outside of Merck helped me to keep my sanity and perspective, reminding me that there was more to life than the lab. To all of you, a very big thank you and my eternal gratitude.

DEDICATION

To my grandparents,

Clara and Isaac Braunstein,

For helping me to remember the important things in life;

and to Marianne Huyer (1963 – 1995):

We shall not cease from exploration

And the end of all our exploring

Will be to arrive where we started

And know the place for the first time.

— T. S. Eliot

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Principles of Signal Transduction

Multicellular organisms have evolved because of their ability to coordinate the activities of hundreds, thousands, and even millions of cells for the benefit of the organism as a whole. This coordination requires an extensive and regulated communication network that allows cells to respond to each other and to external environmental signals. Signaling networks are even required by certain single-cell organisms; for example, yeast cells coordinate mating and sexual reproduction through signaling pathways similar to those in higher eukaryotes. The importance of the fidelity and control of signal transduction is clear from the many disease states that arise from defects in signaling pathways, including cancer, diabetes, and many metabolic disorders. The processes by which signals are sensed and responded to by cells are very complex. However, an enormous research effort, especially in the last 15 years, has greatly advanced the understanding of the signaling machinery used by cells and revealed certain basic principles of signal transduction.

Messages between cells are sent in the form of signaling molecules. Signaling molecules come in many different forms, including proteins, peptides, amino acids, nucleotides, steroids, fatty acid derivatives, and dissolved gases such as nitric oxide and carbon monoxide. The common element shared by all of these signaling molecules is the existence of a specific receptor for that molecule. Thus, many cells can be exposed to a

particular signaling molecule, but only those cells with a receptor for that molecule will respond, ensuring a specific response.

Perhaps the biggest obstacle faced by cells in responding to signals is the cell membrane, which acts as a barrier between extracellular signals and the intracellular machinery that responds to the signals. Cells have evolved two mechanisms for overcoming this obstacle. First, certain extracellular signaling molecules diffuse across the cell membrane and interact with the intracellular machinery directly to effect a response. For example, steroid hormones are small, lipophilic molecules that can cross the cell membrane and bind to their intracellular cytosolic receptors. Upon hormone binding, a conformational change is induced in the receptor that allows it to bind to hormone response elements in promoters of various genes to stimulate their transcription (1, 1a). Similarly, dissolved gases like nitric oxide and carbon monoxide diffuse directly through the cell membrane and interact with their intracellular targets. Nitric oxide, for example, reacts with heme iron in the active site of the enzyme guanylyl cyclase to stimulate its activity and generate a cellular response (2, 3).

The majority of signaling molecules are membrane impermeable and therefore cannot cross the cell membrane to interact with an intracellular receptor target. Consequently, the receptors for these molecules are transmembrane proteins with a ligand binding domain on the surface of the cell. There are three main classes of surface receptors based on the mechanisms by which they transmit a signal across the membrane. One class includes the ion channel receptors. Ligand binding to the extracellular domain causes the channel to open which allows ions to flow into or out of the cell, thereby

generating a signal within the cell. Certain neurotransmitters bind to receptors of this class to excite postsynaptic cells (4).

The second class of receptors includes those linked to trimeric GTP-binding regulatory proteins (G proteins). These receptors are referred to as the G-protein coupled receptors (GPCR) and belong to a superfamily of related seven-pass transmembrane proteins (5, 6). G proteins are peripheral membrane proteins composed of three subunits: G_α, G_β, and G_γ (7, 8). The G_α subunit binds guanine nucleotides (GDP and GTP) and has intrinsic GTPase activity. The G_β and G_γ subunits exist as a tight complex that associates with GDP-bound G_α, forming the inactive basal state. When ligand binds to a GPCR, a conformational change is induced in the receptor that creates a binding site for G_α and stimulates GDP/GTP exchange by G_α. GTP-bound G_α dissociates from the receptor and from $G_{\beta/\gamma}$, freeing it to regulate various effectors.

There are several classes of G_α proteins, the major ones being $G_{\alpha s}$ and $G_{\alpha i}$ for stimulatory and inhibitory G_α proteins, respectively. Free $G_{\alpha s}$ can modulate the activity of two enzymes: phospholipase C_β (PLC_β) and adenylyl cyclase. PLC_β hydrolyzes the membrane lipid phosphatidyl-4,5-bisphosphate (PIP_2) to form inositol triphosphate (IP_3) and diacylglycerol (DAG), which act as second messengers to modulate the activities of other enzymes. Adenylyl cyclase catalyzes the conversion of ATP to cAMP, another important second messenger that activates a Ser/Thr protein kinase (PKA). $G_{\alpha i}$, by contrast, leads to inhibition of adenylyl cyclase activity. $G_{\beta/\gamma}$ subunits can also play a role in signaling by regulating activity of enzymes and ion channels. The soluble second messengers amplify the original signal many times and transmit a signal from the

membrane to locations throughout the cell. The system returns to its basal state when the GTPase activity of G_α hydrolyzes the bound GTP to GDP, allowing $G_{\beta/\gamma}$ to reassociate with G_α.

The third and most diverse class is the enzyme-linked receptors. The cytosolic regions of these transmembrane receptors either have intrinsic catalytic activity or associate directly with an enzyme. Five types of enzyme-linked receptors have been identified: (i) receptor guanylyl cyclases; (ii) receptor tyrosine phosphatases; (iii) receptor Ser/Thr kinases; (iv) receptor tyrosine kinases (RTKs); and (v) tyrosine-kinase associated receptors. By far the majority of enzyme-linked receptors are either RTKs or tyrosine-kinase associated receptors and will be discussed in greater detail below. The receptor guanylyl cyclases convert GTP to cGMP which acts as a second messenger, like cAMP (see above), to activate a Ser/Thr protein kinase (PKG). Atrial natriuretic peptides (ANPs) use receptors of this class, and their binding activates the receptor guanylyl cyclase activity which leads to a cellular response (9). Receptor tyrosine phosphatases remove phosphates from tyrosine-phosphorylated proteins (see below), regulating their activity and thereby propagating the initial signal. While a number of these receptor types have been identified, no ligands are known. Finally, the receptor Ser/Thr kinases phosphorylate proteins on Ser and Thr residues in response to ligand binding. For example, the receptors for members of the transforming growth factor ß (TGF-ß) family of signaling molecules belong to this class, and ligand binding activates their intrinsic Ser/Thr kinase activity, resulting in the phosphorylation of various intracellular proteins to propagate the signal (10).

Tyrosine Phosphorylation in Signal Transduction

The largest class of enzyme-linked receptors are those that use protein tyrosine phosphorylation to transmit the signal. These receptors either have intrinsic catalytic protein tyrosine kinase (PTK) activity (the receptor tyrosine kinases or RTKs) or act through a separate associated PTK. Tyrosine phosphorylation was only discovered in the early '80's, yet in spite of its short history there has been an explosion in understanding of the role played by reversible tyrosine phosphorylation of proteins in signal transduction, and the importance of this modification has been clearly established. The first receptor identified as a RTK was the epidermal growth factor (EGF) receptor (11). Subsequently, a number of hormones and growth factors have been demonstrated to signal through RTKs (12), including insulin, platelet-derived growth factor (PDGF), vascular endothelial growth factor (VEGF), nerve growth factor (NGF), fibroblast growth factor (FGF), and others. Another class of receptors has been identified more recently that do not have intrinsic PTK activity themselves but instead have an associated PTK activity. These are primarily receptors of the cytokine receptor superfamily, and the associated PTKs are members of the JAK (Janus kinase) family of tyrosine kinases (13).

Regardless of whether the tyrosine kinase activity is intrinsic to or associated with the receptor, the net result of receptor stimulation is autophosphorylation of the receptor or the associated kinases. Autophosphorylation results in stimulation of the tyrosine kinase activity and subsequent phosphorylation of protein targets to transmit and amplify the signal. The most common manner in which ligand binding stimulates the intracellular tyrosine kinase activity by inducing receptor dimerization or clustering. Many of the

ligands for RTKs and tyrosine-kinase associated receptors have a dimeric structure that can bind two receptors simultaneously. Consequently, the two receptors are brought in close contact, and as a result of their low basal tyrosine kinase activity the kinases cross-phosphorylate each other, greatly stimulating their activity. Other ligands bind in a monomeric fashion but still induce receptor clustering, perhaps through a conformational change that promotes interactions between receptors.

The majority of tyrosine-kinase associated receptors are cytokine receptors, and their mechanism of signal transmission has been fairly well described (14, 15). As for the RTKs, ligand-induced receptor oligomerization leads to cross-phosphorylation and activation of the associated JAKs. The activated JAKs phosphorylate various targets, including STAT proteins (signal transducers and activators of transcription). Tyrosine phosphorylation of STATs leads to their dimerization through their SH2 domains (discussed below). STATs are active in their dimeric forms, which translocate to the nucleus and stimulate transcription by binding to specific elements in the promoters of genes. Hence, activated receptors initiate a cascade of protein phosphorylation that serves to transmit and amplify signals, resulting in a specific cellular response.

The RTKs transmit their signals in a much less direct route by initiating phosphorylation cascades to amplify the original signal. The best characterized cascade is the MAP (mitogen-activated protein) kinase pathway (16, 17) (Fig. 1.1). In the first step of the pathway, RTK activation leads to stimulation of Ras, a small GTP-binding protein anchored to the cytoplasmic face of the plasma membrane. Ras activation is not direct but instead acts through the recruitment of a GDP/GTP exchange factor (GEF), like Sos, to

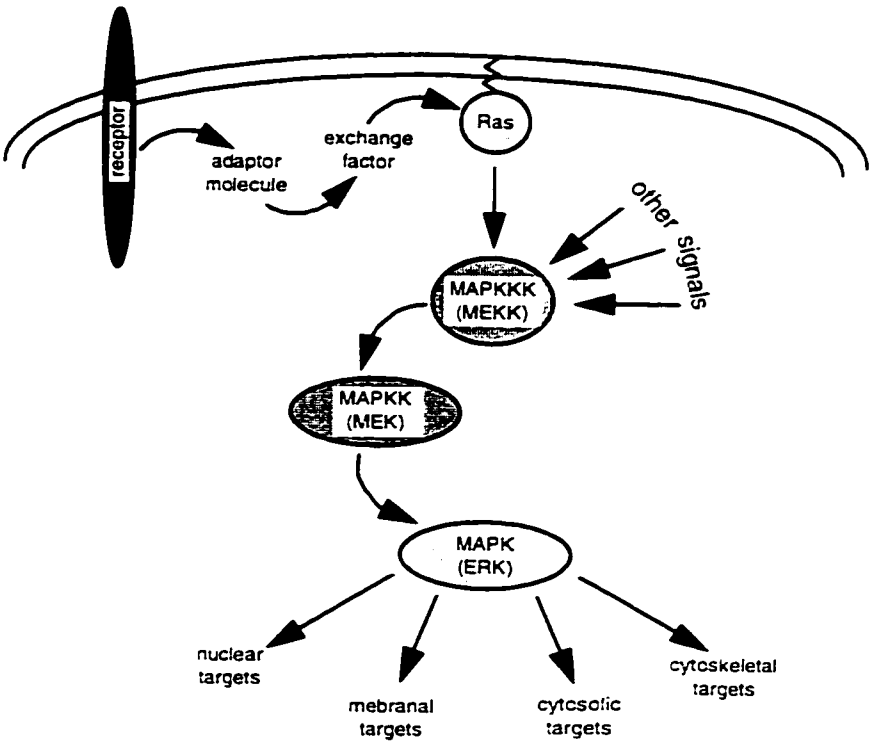


Fig. 1.1. **MAP kinase signaling pathway.** Several MAPKKKs, MAPKKs, and MAPKs are known. Ras can become activated by many extracellular signals which in turn activates the MAPKKK Raf. Other signals can activate Raf and other MAPKKKs. Many cellular targets are regulated by the MAPKs.

the membrane that stimulates Ras to bind GTP, thereby activating it. Activated Ras stimulates the Ser/Thr kinase Raf, one of the MAP-kinase-kinase-kinases (MAPKKK), also called MAP/ERK-kinase-kinases (MEKK). Raf phosphorylates and activates one or more MAP-kinase-kinases (MAPKK), also known as MEKs. These kinases are unusual in their dual-specificity for threonine and tyrosine. Their substrates are the MAP kinases (also called ERKs, for extracellular-signal regulated kinases) which become activated upon threonine and tyrosine phosphorylation. MAP kinases are Ser/Thr kinases that phosphorylate a number of targets leading to changes in the cell. Certain MAP kinases translocate to the nucleus where they phosphorylate transcription factors to activate gene

transcription. Part of the specificity of the MAP kinase pathway comes from the fact that there are multiple kinases at every level of the pathway. Thus, while RTKs can activate the MAP kinase pathway through the MAPKKK Raf, other signals can activate other MAPKKKs, creating parallel signaling cascades that ultimately lead to different cellular responses.

If all RTKs transmitted their signals exclusively through Ras/Raf and the MAP kinase pathway, it would be extremely difficult for signal specificity to be maintained. As such, RTKs also utilize alternative signaling proteins and enzymes to dial in specificity, including PTKs, PTPs, phosphatidylinositol 3-kinase (PI3-K), and structural proteins. There is a great deal of overlap in the use of this signaling machinery by different RTKs. Specificity is maintained by the many permutations and combinations of stimulation and inhibition of the different pathways and machinery, ultimately leading to different net signal outputs.

SH2 Domains

A common theme in signaling mediated by tyrosine phosphorylation is the regulated formation of specific protein complexes. For example, proteins are recruited to RTKs following their activation and tyrosine phosphorylation, and STATs dimerize after tyrosine phosphorylation (as described above). The formation of many protein complexes is directed by SH2 domains. SH2 domains were identified as a sequence of approximately 100 amino acids conserved among the Src family of tyrosine kinases (including Src, Fps, Yes, Abl, and others) and important for their biological function (18). In addition to the

SH2 domain, two other regions of homology exist: the catalytic domain ("SH1") and a non-catalytic region termed "SH3." SH2 domains were subsequently shown to bind to other proteins through a specific interaction with phosphotyrosine (pY) (reviewed in 19). SH2 domains have now been identified in as many as 100 proteins, both enzymes (including PTKs, PTPs, phospholipases, and others) and structural proteins with no intrinsic catalytic activity but often containing other protein binding domains.

The interactions directed by SH2 domains are specific and regulated. The specificity is derived from the amino acid residues immediately surrounding the target pY as these residues are involved in binding. The interaction is regulated by the fact that the tyrosine phosphorylation is reversible, and binding only occurs when the tyrosine is phosphorylated. As such, highly specific protein complexes can form as a result of SH2 domain interactions.

The major consequence of the formation of specific protein–protein complexes mediated by SH2 domains is the recruitment of particular proteins to specific locations in the cell. Proteins can bind via their SH2 domains to RTKs which generally become multiply phosphorylated on Tyr residues when activated, as well as to other proteins that become Tyr-phosphorylated by RTKs. One consequence of such SH2 domain-mediated localization is to direct an SH2 domain-containing enzyme to the vicinity of its substrates. Binding to a pY target also results in the activation of a number of SH2 domain-containing enzymes. For example, the activities of the SH2 domain-containing PTPs, SHP-1 and SHP-2, are increased by as much as 20-fold by occupation of their SH2 domains (20, 21). Src-family PTKs are also regulated by their SH2 domains as a result of intramolecular

association between the SH2 domain and a pY residue near the C-terminal of the protein (22, 23). In this state, the PTK activity is inhibited; dephosphorylation of the pY residue results in a derepression of the PTK as well as freeing the SH2 domain to bind to a pY target in another protein. Activation of Ras (see above) also involves SH2 domain-mediated protein interactions. The Ras activator Sos associates with Grb2, an adapter protein that binds via its SH2 domains directly to a RTK or through an associated pY protein, thereby recruiting Sos from the cytosol to the vicinity of the plasma membrane where it can activate Ras (17).

Besides SH2 domains, other pY-binding domains have been identified. One such domain is referred to as PTB, for phosphotyrosine binding (24). These domains have little or no sequence or structural homology to SH2 domains, defining a distinct binding domain. Recent evidence suggests that some members of this family do not require the tyrosine to be phosphorylated, leading to the suggestion that those domains be called PI, for protein interaction (25). Another pY-binding domain was identified in a protein called STYX (for phospho-serine, threonine, and tyrosine interaction) (26). The binding domain has high homology to the dual-specificity PTPs (see below), except that the catalytic Cys residue is replaced by Gly. As a result, the domain will bind phosphorylated substrates but does not hydrolyze them. Specific and regulated protein–protein interactions are an essential part of signal transduction and as such it is not surprising that several different interaction domains have evolved.

The small size and stability of SH2 domains have facilitated structural determinations both by X-ray crystallography and NMR (for examples, see 27-37). Structures for

at least 12 SH2 domains with or without bound pY-peptides have been determined. All SH2 domains share a highly-conserved structure of an antiparallel ß-sheet sandwiched by two α-helices. The nature of the pY binding pocket is highly conserved amongst the SH2 domains. From the structures with bound peptides, the determinants of peptide sequence specificity are apparent and provide a structural rationale for the sequence specificities observed *in vitro* and *in vivo* (discussed in detail below). The structures of the PTB domains of Shc (38) and IRS-1 (39) have also been determined and are quite distinct from SH2 domains in that they form a ß-sandwich capped by a single α-helix. Interestingly, the pY-binding pockets are not conserved between these two PTB domains.

Protein Tyrosine Phosphatases

As indicated above, tyrosine phosphorylation is a reversible protein modification. The enzymes responsible for removing the phosphate are protein tyrosine phosphatases (PTPs). The first PTP was identified ten years ago as an activity purified from human placenta and termed PTP1B (40, 41). The sequence of this protein showed high homology to the intracellular domain of CD45, a transmembrane protein surface antigen of immune cells (42). Subsequently, CD45 was also shown to have PTP activity (43). All PTPs share a characteristic 11-residue "signature motif" (I/V)HCXAGXXR(S/T)G, where "X" indicates any amino acid.

A large family of PTPs with more than 75 members has been identified to date. The enzymes fall into two main classes: receptor-like and cytosolic (Fig. 1.2). The receptor-like PTPs are all type I transmembrane proteins with a variable extracellular

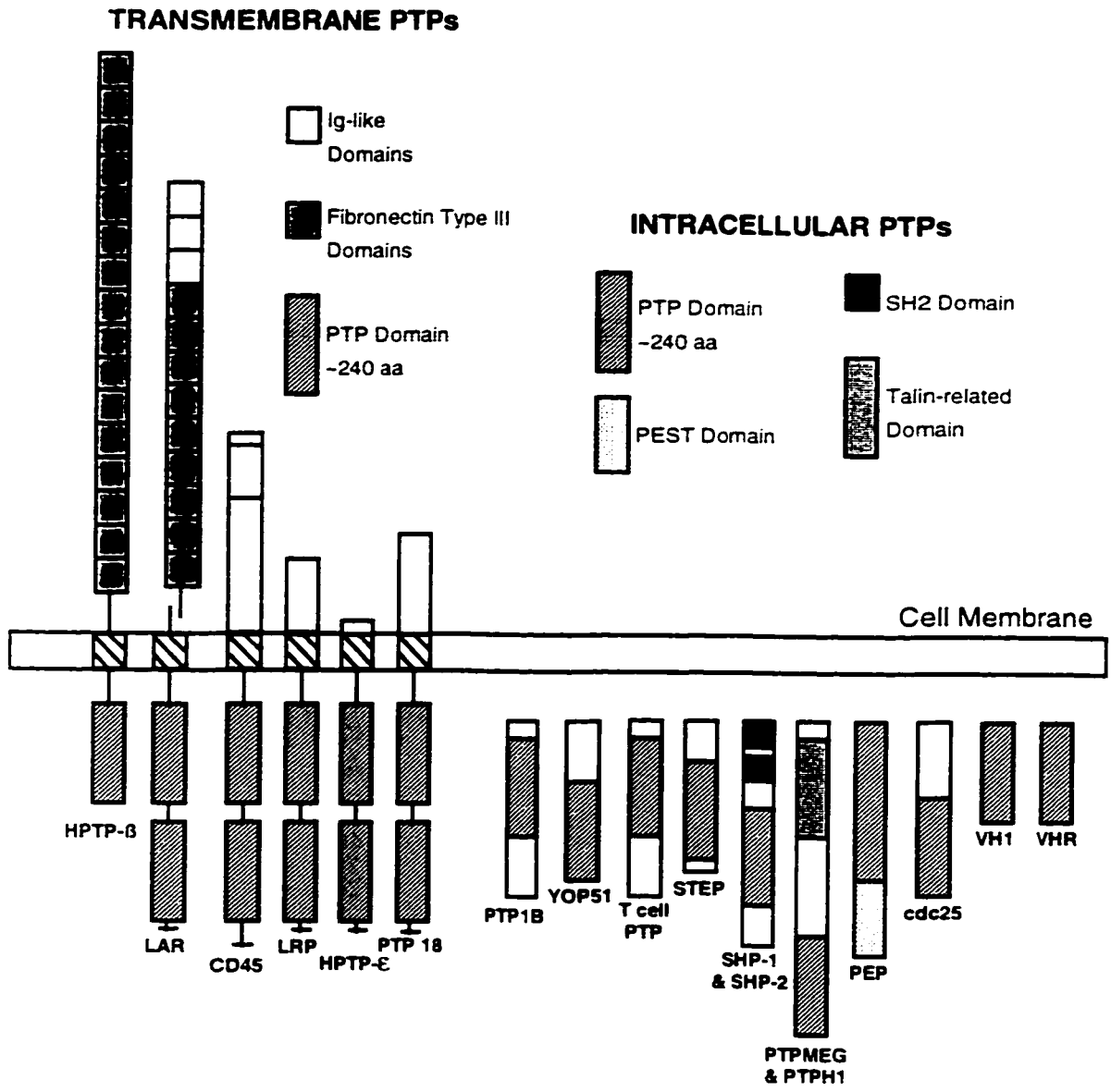


Fig. 1.2. **Protein tyrosine phosphatase families.** The PTPs can be divided into two main classes: transmembrane and intracellular. The transmembrane PTPs can be further subdivided by the number of PTP domains, and the intracellular PTPs by their specificity for phosphotyrosine or dual-specificity for phosphoserine/phosphothreonine in addition to phosphotyrosine (the latter group includes cdc25, VH1, and VHR). Some of the conserved protein domains of the PTPs are indicated.

domain and two intracellular PTP domains (a few exceptions have one PTP domain). Generally, almost all of the PTP activity is associated with the membrane proximal (D1) domain. The notable exception is PTPα: both D1 and D2 are active, although D1 is the more active of the two (44). There is one report of catalytic activity associated with D2 of CD45 (45); however, it was not clearly demonstrated that the activity was due to CD45. The extracellular domains of the transmembrane PTPs are predicted to interact with specific ligands, but none has been identified.

The second class of PTPs are the cytosolic enzymes. These all have a single PTP domain and a variety of N- or C-terminal domain extensions. The cytosolic PTPs can be further subdivided into two classes: those with strict specificity for phosphotyrosine, and those with dual specificity for phosphoserine (pS) and phosphothreonine (pT) in addition to phosphotyrosine. The first dual-specificity PTP, VH1, was identified only six years ago as an open reading frame from *Vaccinia* virus (46). Dual-specificity PTPs hydrolyze pS and pT much more slowly that pY, but the same catalytic mechanism is used for all dephosphorylation reactions (47).

The catalytic mechanism of PTPs (Fig. 1.3) has been elucidated from a number of kinetic, mutagenic, and structural studies (48, 49). All PTPs have an essential catalytic Cys (part of the signature motif) to which the phosphate of the pY substrate becomes covalently attached during catalysis; this thiophosphate intermediate is then hydrolyzed by water (50, 51). The pK_a of the nucleophilic Cys sulfhydryl has been shown to be dramatically lowered in the active site of the *Yersinia* PTP (Yop51) to 4.7 from an average value of ~8.5 for free cysteine (52); thus, the Cys thiol is deprotonated at

physiological pH. Structural studies show that the thiolate anion is stabilized in the *Yersinia* PTP by an extensive network of hydrogen bonds (53). By contrast, the structure of PTP1B shows that an Arg side chain forms a hydrogen bond or salt bridge with the thiolate to stabilize it (54). In the first step of the reaction, the thiolate anion undergoes an in-line nucleophilic attack on the phosphate of the pY, forming a thiol-phosphate ester. The tyrosine leaving group is protonated by an invariant aspartic acid residue contained on a flexible loop that upon substrate binding moves ~7 Å to bring this general acid in proximity to the active site (53). In the second step, a water molecule that becomes activated by a general base (perhaps the same Asp that acts as a general acid in the first step of the reaction) hydrolyzes the thiophosphate linkage, releasing inorganic phosphate and regenerating the Cys for another round of catalysis. Replacement of the active-site Cys with any other residue abolishes catalytic activity. Interestingly, substitution of the Cys with Ser (identical to Cys except for a hydroxyl side chain instead of a sulfhydryl) results in an inactive mutant that can still bind substrate but does not hydrolyze it (50, 60).

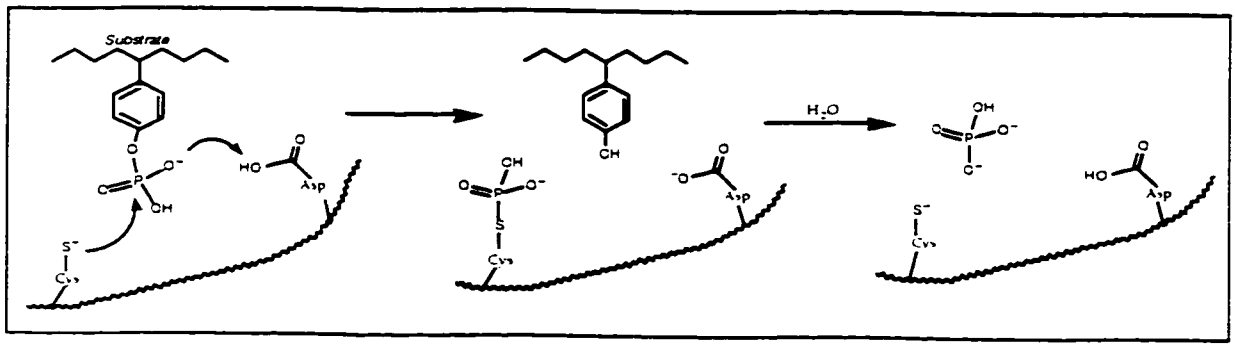


Fig. 1.3. **Catalytic mechanism of PTPs.** See text for details. In the first step of the reaction, the phosphate is cleaved from the phosphotyrosine substrate and becomes covalently attached to the nucleophilic catalytic Cys. In the second step, the thiol–phosphate bond is hydrolyzed, releasing inorganic phosphate.

X-ray crystal structures of four PTPs have been determined: PTP1B (54), the *Yersinia* PTP (53), RPTPα (receptor PTPα, distinct from PTPα) (55), and the dual-specificity PTP VHR (56). The structure of a catalytically-inactive form of PTP1B (in which the catalytic Cys is replaced by Ser) complexed with a high-affinity pY-peptide substrate has also been elucidated (60). Additionally, several structures exist for a class of PTPs known as the low-molecular-weight PTPs (57-59), small cytosolic PTPs of ~18 kDa of unknown function that contain the PTP signature motif but have no other sequence homology to the PTPs. All of the PTP structures have a similar architecture despite low sequence homology between the PTPs. The structures have been informative in describing the catalytic mechanism of PTPs (see above); as well, the PTP1B–peptide structure has provided details regarding determinants of sequence specificity and recognition.

PTPs play important roles in regulating tyrosine-phosphorylation signaling pathways. As PTPs were discovered after PTKs, they are generally described as opposing the actions of PTKs which are often thought of as the positive mediators of signal transduction. This is certainly true for cases in which tyrosine phosphorylation serves to activate an enzyme or positively transduce a signal. For example, protein signaling complexes that form via SH2 domain interactions upon tyrosine phosphorylation are disassembled as a result of PTP activity that dephosphorylates the pY residues to which the SH2 domains are bound. However, there are also situations in which PTPs act as positive regulators by activating enzymes. The clearest case is the role played by CD45 in T-cell receptor signaling. In response to T-cell activation, CD45 dephosphorylates a

number of intracellular targets, in particular the PTKs Lck and Fyn (61). The dephosphorylation activates these PTKs by relieving the intramolecular SH2 domain-mediated inhibition described above, thereby propagating the initial signal.

Specificity of SH2 Domains and PTPs

The goal of the work described in this thesis was to gain a better understanding of the specificity of SH2 domains and PTPs. Specificity is essential in signaling pathways to prevent inappropriate crosstalk between pathways that may result in undesired cellular responses. As noted above, receptors that signal through tyrosine phosphorylation use many of the same components of the signaling machinery and the same phosphorylation cascades are often activated. It is the specific involvement of the appropriate components, including SH2 domains and PTPs, that ensures the fidelity of the signal. The determinants of SH2 specificity are becoming clearer from structural and binding studies, as described below. However, PTP specificity is still somewhat of a black box and the focus of much current research.

SH2 domains and PTPs have in common the fact that they both bind to pY. PTPs take this binding one step further and also hydrolyze the phosphate–tyrosine linkage. Not surprisingly, there is a reasonable amount of structural conservation in the pY binding pocket of SH2 domains and PTPs. However, there is no apparent sequence conservation, suggesting that the similarity represents convergent evolutionary solutions to pY binding. An important reason for the specificity displayed by SH2 domains and PTPs towards pY is the depth of the binding pocket: pS and pT cannot reach far enough into the binding

pocket for the phosphate to make the important interactions with residues at the bottom of the pocket (53, 54, 60). The exception is the dual-specificity PTPs which have a shallower binding groove as shown in the crystal structure, allowing pS and pT to reach the catalytic Cys residue (56).

Specificity of SH2 domains is defined to a large extent by the residues surrounding the target pY. The role of residues flanking the pY was apparent from the fact that SH2 domains of proteins that bind to the PDGF receptor do not recognize identical sites, as mutation or deletion of a particular pY residue only abrogated binding of certain proteins (18). As well, a comparison of sequences surrounding the pY binding sites for PI-3 kinase in multiple proteins revealed that certain residues are conserved, implying that these residues are important (62). The first crystal structures of SH2 domains with bound peptides clearly demonstrated the involvement of residues surrounding the pY in mediating binding. The first SH2 domains crystallized, from Src and Lck, have a similar binding mechanism in which a hydrophobic binding pocket in the SH2 domain selects for hydrophobic aliphatic residues at the pY+3 position (*i.e.*, three positions C-terminal to the pY) (28, 29). The crystal structure of the SH2 domain of PLCγ suggested a different mode of binding in which the pY peptide binds in a shallow groove and makes extensive interactions as far as the pY+5 residue (30).

Techniques for studying the binding of SH2 domains to pY peptides *in vitro* greatly facilitated the study of their sequence specificities. Studies with individual peptides were carried out by a number of techniques, including surface plasmon resonance to measure binding directly in real time (63-65), and competitive radioactive peptide binding

assays (66, 67). The use of individual peptides as required in these studies is labourious and does not permit an examination of all possible sequence permutations and combinations. Therefore, an affinity selection technique from peptide libraries was developed to screen SH2 domains rapidly to determine their consensus sequences for high affinity binding (68). In general, all of these studies suggested that the primary determinants of SH2 binding specificity were the residues C-terminal to the target pY, and in particular the pY+3 residue was critical.

The sequence specificity of PTPs is much more poorly understood. In fact, since many PTPs are quite promiscuous *in vitro*, it was suggested that PTPs have little intrinsic catalytic specificity and instead that specificity is achieved by strictly controlling their subcellular localization and access to substrates. However, it is becoming apparent that at least some PTPs have intrinsic catalytic specificity. For example, when studied with pY-peptide substrates, a particular PTP can display widely varying catalytic efficiencies towards the different peptides (69-75), indicating that the residues surrounding the pY clearly play a role in recognition by the PTP. As well, chimeric proteins have been used *in vivo* to show that, in at least two cases, the PTP domain of one protein cannot substitute for another (76, 77), consistent with intrinsic specificity differences. In general, though, only a few careful studies of PTP specificity have been performed.

Methods of Studying Specificity

As indicated above, structural studies are very informative in providing molecular details of sequence specificity, in particular for SH2 domains. However, for PTPs, only

one structure with bound peptide exists (60). It is not a trivial matter to generate an X-ray or NMR structure of a protein, with or without bound ligand. Although a structural determination provides a snapshot of the protein–ligand interactions that identifies sites of contact, kinetic and thermodynamic studies are generally required to assess the relative importance of the contacts in the binding interactions.

A less direct approach is to examine the interactions between SH2 domains or PTPs with individual pY-protein or -peptide targets. By exposing the protein to a variety of pY targets and assessing differences in affinities, important residues in the sequence can be identified and a consensus sequence for high affinity binding determined. This approach in particular has been successful with SH2 domains (78-80). However, the approach is labourious since many different peptides or proteins must be examined, and it is not practical to examine all possible sequence combinations.

More recently, combinatorial libraries have been used to define substrate sequence specificities. The library approach has the advantage of exposing the protein of interest to many different sequence combinations at once, and the highest affinity peptides are isolated by affinity selection and identified. This technique has been applied successfully to a number of proteins, including SH2 domains (68, 81), protein kinases (82, 83), and PDZ domains (84). A more biological approach has also been used with phage display libraries (85). In this method, random peptide sequences are inserted into a surface phage protein, and the target protein is used to select phage displaying high affinity peptide sequences. Because the peptide is part of a larger protein it has a more biologically relevant secondary structure, and the phage technique permits amplification of selected

sequences and direct determination by DNA sequencing. However, in order to use phage display for SH2 domains and PTPs, methods need to be developed to incorporate phosphotyrosine (or an analogue) into the displayed sequence.

Organization of Thesis

Initially, studies were undertaken to examine the sequence specificity requirements of SH2 domains, using the SH2 domains of the phosphatase SHP-2 as a model. When these studies were started, detailed information about SH2 domain specificity was just beginning to be reported. For example, only three crystal structures were known (27-30), and the structure of the N-terminal SH2 domain of SHP-2 was solved while these studies were in progress (31). The first series of experiments was to examine the sequence specificity requirements surrounding Tyr_{1009} of the PDGF receptor, the *in vivo* target of the N-terminal SH2 domain of SHP-2. The results of this study have been published (86) and are reported in Chapter 2. This work demonstrated, among other things, that residues N-terminal to the pY of the target peptide are necessary for high-affinity binding of this SH2 domain, in contrast to most SH2 domains whose specificity is controlled exclusively by residues C-terminal to pY. This unexpected requirement for N-terminal residues was followed up in Chapters 3 and 4.

Chapter 3 describes an affinity selection approach with peptide libraries to explore further the SH2 domain specificity of SHP-2. A published protocol was followed (68) with certain modifications. Specifically, instead of using peptide sequencing for the analysis as previously employed, we chose to use continuous flow–liquid secondary ion

mass spectrometry (CF-LSIMS) to characterize the selected peptides. Unfortunately, a number of technical limitations were encountered with this MS approach that could not be solved due to limited instrument access. In spite of these problems, the study yielded promising preliminary results and was key to developing the technology used in Chapter 7.

In Chapter 4, the structural basis was explored for the specificity of the SH2 domains of SHP-2 for residues N-terminal to the target pY. An examination of the sequence of the N-terminal SH2 domain of SHP-2 showed a Gly residue at the αA2 position instead of the normally highly-conserved Arg residue. Because of the αA2 Gly, a gap is apparent in the crystal structure that is filled by a residue N-terminal to the target pY, suggesting a structural reason for the N-terminal residue specificity. The αA2 Gly was mutated to Arg and indeed found to alter the requirement for residues N-terminal to pY, demonstrating the involvement of the αA2 residue in directing SH2 domain specificity.

The second part of the thesis work was directed at understanding the specificity of PTPs. One approach to determine enzyme specificity in signaling pathways is to specifically inhibit a particular enzyme and observe the effects in the cell. Unfortunately, no specific inhibitors of PTPs, and very few non-specific inhibitors, are known. The most commonly used PTP inhibitors are vanadate and pervanadate, but the details of their mechanisms of inhibition had not been examined. These inhibitors have also been used in cell and animal models as insulin mimetics (87, 88), even without an understanding of their modes of action. To address this fundamental gap in the literature, a detailed study of vanadate and pervanadate inhibition was undertaken with the enzyme PTP1B. The results

of this study have been published (89) and are reported in Chapter 5. The different modes of vanadate and pervanadate inhibition reflect differences in how they interact at the active site of PTP1B and the results provide key insights into their *in vivo* activity.

The most frequently used approach to probe PTP specificity is to evaluate the efficiency of a PTP towards individual pY-peptide substrates (for example, 69-75). In Chapter 6, the peptide substrate specificities of three PTPs (CD45, LAR, and PTPβ) were examined by this classical approach. These PTPs did not appear to discriminate between the various peptide substrates used, suggesting that they have little intrinsic specificity. However, the study was limited because of the time and expense involved in testing individual peptides and therefore a complete range of substrate sequences could not be examined.

To address the limitations of studying PTP sequence specificity with individual pY-peptide substrates, a more general approach was developed in Chapter 7. Specifically, the affinity selection approach with peptide libraries described in Chapter 3 was adapted to define the sequence specificity of PTP1B. The technology developed in Chapter 7 incorporated a number of improvements that overcame the limitations encountered in Chapter 3. The success of this study provides a methodology that can be extended to any PTP and should assist in the identification of *in vivo* substrates of PTPs by defining their optimal target sequences. Furthermore, an understanding of PTP sequence specificity differences will provide a basis for the development of specific PTP inhibitors, yielding important tools for studying signaling pathways and even therapeutic agents.

Contributions of Colleagues

The collaborative nature of research, especially at Merck Frosst, means that I have benefited greatly from the expertise of many colleagues. I am indebted to them for their contributions to the work described in this thesis. For unpublished studies, the work performed by others is indicated in the Acknowledgments section of each chapter. However, Chapters 2, 5, and 7 are multi-author papers or submitted manuscripts which I wrote in their entirety. The contributions of the co-authors are acknowledged below.

Chapter 2 was published in *Biochemistry* [Huyer, G., Li, Z. M., Adam, M., Huckle, W. R., and Ramachandran, C. (1995) *Biochemistry* **34,** 1040-1049] and is reproduced with the permission of the journal. Z. M. Li and M. Adam provided the subclones of the different GST fusion proteins used for the binding studies. W. R. Huckle provided the tyrosine kinase that was used to phosphorylate the PDGF receptor peptide. C. Ramachandran performed the BIAcore experiments and most of the phosphatase kinetic assays. I developed the radioactive competitive binding assay and performed all of the experiments associated with the assay, as well as expressing and purifying all of the fusion proteins used.

Chapter 5 is reproduced from the *Journal of Biological Chemistry* [Huyer, G., Liu, S., Kelly, J., Moffat, J., Payette, P., Kennedy, B., Tsaprailis, G., Gresser, M. J., and Ramachandran, C. (1997) *J. Biol. Chem.* **272,** 843-851] with the permission of the journal. The contributions of the co-authors are as follows. S. Liu performed the initial studies on vanadate and pervanadate, in particular describing the EDTA effect on vanadate. While ultimately none of her work was incorporated into this manuscript, her

initial efforts were essential to get this study off the ground and she is therefore acknowledged with an authorship. J. Kelly and G. Tsaprailis performed the mass spectrometric analysis on PTP1B samples that I prepared with and without pervanadate treatment; the analysis of the tryptic peptides and the MS/MS were done by J. Kelly with my assistance. The subcloning of PTP1B to make the FLAG fusion protein and the expression and purification from *E. coli* were performed in the laboratory of B. Kennedy with the assistance of J. Moffat and P. Payette. Finally, M. Gresser and C. Ramachandran provided intellectual support, supervision, and many critical insights.

Chapter 7 has been submitted for publication with the following author list: G. Huyer, J. Kelly, J. Moffat, R. Zamboni, M. J. Gresser, and C. Ramachandran. The technical expertise of J. Kelly was essential to develop and validate the mass spectrometric analytical method. As well, he greatly assisted in running the samples I prepared and in extracting the relevant data for me to analyze. J. Moffat performed the large-scale preparations of GST-PTP1B following protocols I developed. Furthermore, as an undergraduate co-op student under the joint supervision of myself and C. Ramachandran, he was instrumental in performing the initial development experiments to validate the affinity selection methodology in a controlled system. While none of his results were incorporated into the manuscript, his work laid the foundation for the methodology I ultimately developed. R. Zamboni synthesized the peptide libraries and the F_2Pmp monomer, as well as optimizing the peptide synthesis protocol and teaching me how to synthesize the individual F_2Pmp-containing peptides. Finally, as throughout my thesis work, M. Gresser and C. Ramachandran provided invaluable intellectual contributions and

supervision. In addition, the initial idea to develop an affinity selection approach came from M. Gresser.

REFERENCES

1. Parker, M. G. (1993) *Curr. Opin. Cell Biol.* **5,** 499-504

1a. Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., and Evans, R. M. (1995) *Cell* **83,** 835-839

2. Lowenstein, C. J., and Snyder, S. H. (1992) *Cell* **70,** 705-707

3. Knowles, R. G., and Moncada, S. (1992) *Trends Biochem. Sci.* **17,** 399-402

4. Barnard, E. A. (1992) *Trends Biochem. Sci.* **17,** 368-374

5. Strader, C. D., Fong, T. M., Tota, M. R., Underwood, D., and Dixon, R. A. F. (1994) *Annu. Rev. Biochem.* **63,** 101-132

6. Wess, J. (1997) *FASEB J* **11,** 346-354

7. Hepler, J. R., and Gilman, A. G. (1992) *Trends Biochem. Sci.* **17,** 383-387

8. Rens-Domiano, S., and Hamm, H. E. (1996) *FASEB J* **9,** 1059-1066

9. Maack, T. (1992) *Annu. Rev. Physiol.* **54,** 11-27

10. Massagué, J. (1992) *Cell* **69,** 1067-1070

11. Buhrow, S. A., Cohen, S., and Staros, J. V. (1982) *J. Biol. Chem.* **257,** 4019-4022

12. Fantl, W. J., Johnson, D. E., and Williams, L. T. (1993) *Annu. Rev. Biochem.* **62,** 453-481

13. Ihle, J. (1995) *Adv. Immun.* **60,** 1-35

14. Darnell, J. E., Jr., Kerr, I. M. and Stark, G. R. (1994) *Science* **264,** 1415-1421

15. Ihle, J. (1995) *Nature* **377,** 591-594

16. Nishida, E., and Gotoh, Y. (1993) *Trends Biochem. Sci.* **18,** 128-131

17. Seger, R., and Krebs, E. G. (1995) *FASEB J* **9,** 726-735

18. Sadowski, I., Stone, J. C., and Pawson, T. (1986) *Mol. Cell. Biol.* **12,** 4396-4408

19. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991) *Science* **252,** 668-674

20. Sugimoto, S. *et al.* (1994) *J. Biol. Chem.* **269,** 13614-13622

21. Pluskey, S., Wandless, T. J., Walsh, C. T., and Shoelson, S. E. (1995) *J. Biol. Chem.* **270,** 2897-2900

22. Cooper, J. A., and Howell, B. (1993) *Cell* **73,** 1051-1054

23. Xu, W., Harrison, S. C., and Eck, M. J. (1997) *Nature* **385,** 595-602

24. Kavanaugh, W. M., and Williams, L. T. (1994) *Science* **266,** 1862-1865

25. Margolis, B. (1996) *J. Lab. Clin. Med.* **128,** 235-241

26. Wishart, M. J., Denu, J. M., Williams, J. A., and Dixon, J. E. (1995) *J. Biol. Chem.* **270,** 26782-26785

27. Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. D., Silverman, L., and Kuriyan, J. (1992) *Nature* **358,** 646-653

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

29. Eck, M. J., Shoelson, S. E., and Harrison, S. C. (1993) *Nature* **362,** 87-91

30. Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., and Forman-Kay, J. D. (1994) *Cell* **77,** 461-472

31. Lee, C.-H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E., and Kuriyan, J. (1994) *Structure* **2,** 423-438

32. Mikol, V., Baumann, G., Keller, T. H., Manning, U., and Zurini, M. G. M. (1995) *J. Mol. Biol.* **246,** 344-355

33. Nolte, R. T., Eck, M. J., Schlessinger, J., Shoelson, S. E., and Harrison, S. C. (1996) *Nature Struct. Biol.* **3,** 364-374

34. Tong, L., Warren, T. C., King, J., Betageri, R., Rose, J., and Jakes, S. (1996) *J. Mol. Biol.* **256,** 601-610

35. Metzler, W. J., Leiting, B., Pryor, K., Mueller, L., and Farmer, B. T. (1996) *Biochemistry* **35,** 6201-6211

36. Breeze, A. L., Kara, B. V., Barratt, D. G., Anderson, M., Smith, J. C., Luke, R. W., Best, J. R., and Cartlidge, S. A. (1996) *EMBO J.* **15,** 3579-3589

37. Thorton, K. H., Mueller, W. T., McConnell, P., Zhu, G., Saltiel, A. R.,and Thanabal, V. (1996) *Biochemistry* **35,** 11852-11864

38. Zhou, M.-M., Ravichandran, K. S., Olejniczak, E. T., Petros, A. M., Meadows, R. P., Sattler, M., Harlan, J. E., Wade, W. S., Burakoff, S. J., and Fesik, S. W. (1995) *Nature* **378,** 584-592

39. Eck, M. J., Dhe-Paganon, S., Trub, T., Nolte, R. T., and Shoelson, S. E. (1996) *Cell* **85,** 695-705

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

41. Tonks, N. K., Diltz, C. D., and Fischer, E. H. (1988) *J. Biol. Chem.* **263,** 6731-6737

42. Charbonneau, H., Tonks, N. K., Walsh, K. A., and Fischer, E. H. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 7182-7186

43. Tonks, N. K., Diltz, C. D., and Fischer, E. H. (1990) *J. Biol. Chem.* **265,** 10674-10680

44. Wang, Y., and Pallen, C. J. (1991) *EMBO J* **10,** 3231-3237

45. Tan, X., Stover, D. R., and Walsh, K. A. (1993) *J. Biol. Chem.* **268,** 6835-6838

46. Guan, K. L., Broyles, S. S., and Dixon, J. E. (1991) *Nature* **350,** 359-362

47. Denu, J. M., and Dixon, J. E. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 5910-5914

48. Zhang, Z.-Y., and Dixon, J. E. (1994) *Adv. Enzymol. Relat. Areas Mol. Biol.* **68,** 1-36

49. Barford, D., Jia, Z., and Tonks, N. K. (1995) *Nature Struct. Biol.* **2,** 1043-1053

50. Guan, K. L., and Dixon, J. E. (1991) *J. Biol. Chem.* **266,** 17026-17030

51. Cho, H., Krisharaj, R., Kitas, E., Bannwarth, W., Walsh, C. T., and Anderson, K. S. (1992) *J. Am. Chem. Soc.* **114,** 7296-7298

52. Zhang, Z.-Y., and Dixon, J. E. (1993) *Biochemistry* **32,** 9340-9345

53. Stuckey, J. A., Schubert, H. L., Fauman, E. B., Zhang, Z.-Y., Dixon, J. E., and Saper, M. A. (1994) *Nature* **370,** 571-575

54. Barford, D., Flint, A. J., and Tonks, N. K. (1994) *Science* **263,** 1397-1404

55. Bilwes, A. M., den Hertog, J., Hunter, T., and Neel, J. P. (1996) *Nature* **382,** 555-559

56. Yuvaniyama, J., Denu, J. M., Dixon, J. E., and Saper, M. A. (1996) *Science* **272,** 1328-1331

57. Su, X.-D., Taddei, N., Stefani, M., Ramponi, G., and Nordlund, P. (1994) *Nature* **370,** 575-578

58. Logan, T. M., Zhou, M.-M., Nettesheim, D. G., Meadows, R. P., Van Etten, R. L., and Fesik, S. W. (1994) *Biochemistry* **33,** 11087-11096

59. Zhang, M., Van Etten, R. L., and Stauffacher, C. V. (1994) *Biochemistry* **33,** 11097-11105

60. Jia, Z., Barford, D., Flint, A. J., and Tonks, N. K. (1995) *Science* **268,** 1754-1758

61. Trowbridge, I. S., and Thomas, M. L. (1994) *Ann. Rev. Immunol.* **12,** 85-116

62. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Garziani, A., Kapeller, R., and Soltoff, S. (1991) *Cell* **64,** 281-302

63. Felder, S., Zhou, M., Hu, P., Urena, J., Ullrich, A., Chaudhuri, M., White, M., Shoelson, S. E., and Schlessinger, J. (1993) *Mol. Cell. Biol.* **13,** 1449-1455

64. Panayotou, G., Gish, G., End, P., Truong, O., Gout, I., Dhand, R., Fry, M. J., Hiles, I., Pawson, T., and Waterfield, M. D. (1993) *Mol. Cell. Biol.* **13,** 3567-3576

65. Payne, G., Shoelson, S. E., Gish, G. D., Pawson, T., and Walsh, C. T. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 4902-4906

66. Piccione, E., Case, R. D., Domchek, S. M., Hu, P., Chaudhuri, M., Backer, J. M., Schlessinger, J., and Shoelson, S. E. (1993) *Biochemistry* **32,** 3197-3202

67. Case, R. D., Piccione, E., Wolf, G., Benett, A. M., Lechleider, R. J., Neel, B. G., and Shoelson, S. E. (1994) *J. Biol. Chem.* **269,** 10467-10474

68. 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., and Cantley, L. C. (1993) *Cell* **72,** 767-778

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

70. Zhang, Z.-Y., Maclean, D., McNamara, D. J., Sawyer, T. K., and Dixon, J. E. (1994) *Biochemistry* **33,** 2285-2290

71. Cho, H., Krishnaraj, R., Itoh, M., Kitas, E., Bannwarth, W., Saito, H., and Walsh, C. T. (1993) *Protein Science* **2,** 977-984

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

73. Dechert, U., Adam, M., Harder, K. W., Clark-Lewis, I., and Jirik, F. (1994) *J. Biol. Chem.* **269,** 5602-5611

74. Dechert, U., Affolter, M., Harder, K. W., Matthews, J., Owen, P., Clark-Lewis, I., Thomas, M. L., Aebersold, R., and Jirik, F. R. (1995) *Eur. J. Biochem.* **231,** 673-681

75. Kwon, M., Oh, M., Han, J., and Cho, H. (1996) *Biochem. Mol. Biol.* **29,** 386-392

76. Musci, M. A., Beaves, S. L., Ross, S. E., Yi, T., and Koretzky, G. A. (1997) *J. Immun.* **158,** 1565-1571

77. Tenev, T., Keilhack, H., Tomic, S., Stoyanov, B., Stein-Gerlach, M., Lammers, R., Krivtsov, A. V., Ullrich, A., and Böhmer, F.-D. (1997) *J. Biol. Chem.* **272,** 5966-5973

78. Escobedo, J. A., Kaplan, D. R., Kavanaugh, W. M., Turck, C. W., and Williams, L. T. (1991) *Mol. Cell. Biol.* **11,** 1125-1132

79. Fantl, W. J., Escobedo, J. A., Martin, G. A., Turck, C. W., del Rosario, M., McCormick, F., and Williams, L. T. (1992) *Cell* **69,** 413-423

80. Nishimura, R., Li, W., Kashishian, A., Mondino, A., Zhou, M., Cooper, J., and Schlessinger, J. (1993) *Mol. Cell. Biol.* **13,** 6889-6896

81. Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R. A., and Cantley, L. C. (1994) *Mol. Cell. Biol.* **14,** 2777-2785

82. Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F., Piwnica-Worms, H., and Cantley, L. C. (1994) *Curr. Biol.* **4,** 973-982

83. Songyang, Z., Carraway, K. L., Eck, M. J., Harrison, S. C., Feldman, R. A., Mohammad, M., Schlessinger, J., Hubbard, S. R., Smith, D. P., Eng, C., Lorenzo, M. J., Ponder, B. A. J., Mayer, B. J., and Cantley, L. C. (1995) *Nature* **373,** 536-539

84. Songyang, Z., Fanning, A. S., Fu, C., Xu, J., Marfatia, S. M., Chishti, A. H., Crompton, A., Chan, A. C., Anderson, J. M., and Cantley, L. C. (1997) *Science* **275,** 73-77

85. Clackson, T., and Wells, J. A. (1994) *Trends Biotechnol.* **12,** 173-184

86. Huyer, G., Li, Z. M., Adam, M., Huckle, W. R., and Ramachandran, C. (1995) *Biochemistry* **34,** 1040-1049

87. Sekar, N., Li, J., and Shechter, Y. (1996) *Crit. Rev. Biochem. Mol. Biol.* **31,** 339-359

88. Bevan, A. P., Drake, P. G., Yale, J.-F., Shaver, A., and Posner, B. I. (1995) *Mol. Cell. Biochem.* **153,** 49-58

89. Huyer, G., Liu, S., Kelly, J., Moffat, J., Payette, P., Kennedy, B., Tsaprailis, G., Gresser, M. J., and Ramachandran, C. (1997) *J. Biol. Chem.* **272,** 843-851

CHAPTER 2

Direct Determination of the Sequence Recognition Requirements of the SH2 Domains of SHP-2*

SUMMARY

SHP-2 is a widely-expressed protein tyrosine phosphatase with two tandem SH2 (Src homology 2) domains and a C-terminal catalytic domain. Glutathione *S*-transferase fusions of the SH2 domains alone and of a catalytically-inactive full-length mutant were made, and binding assays were developed using the purified fusion proteins to directly determine what residues are involved in the recognition of binding targets by the SH2 domains. The binding kinetics of the SH2 domains to a phosphotyrosyl-containing peptide of the sequence surrounding Tyr_{1009} of the platelet-derived growth factor receptor (PDGFR) ß subunit [DTSSVL(pY)TAVQPN] were determined by surface plasmon resonance, confirming that this is a high-affinity binding ligand. Using various N- and C-terminal truncations of this peptide as competitors in the binding assays, the minimum peptide that served as a high-affinity binding ligand was found to be VL(pY)TAV. Systematic Ala substitutions of this peptide indicated that in addition to the phosphotyrosine (pY), the critical residues for recognition and binding are at pY+1 and pY+3 as previously reported, and notably at pY-2 as well. Binding competition results with these and other PDGFR, IRP, and IRS-1 peptides suggested some general rules for sequence

* Reprinted with permission from *Biochemistry* 1995, 34, 1040-1049.

recognition by the SH2 domains of SHP-2. Peptides that bind to the SH2 domains in the binding assays were also found to stimulate the phosphatase activity of SHP-2.

INTRODUCTION

Protein tyrosine phosphorylation plays an essential role in regulating cell function. The tyrosine phosphorylation status of a protein in the cell reflects the balance between the combined activities of protein tyrosine kinases (PTKs)[1] and protein tyrosine phosphatases (PTPs). One mode of signal transmission by tyrosine phosphorylation involves SH2 domains. These domains consist of about 100 amino acids and bind to pY residues in proteins, with the binding specificity apparently determined by the amino acids around the phosphotyrosine (1-5). SH2 domains are present in a variety of signaling proteins, and the proteins can be broadly divided into two classes. The first class of proteins has no inherent enzymatic activity in the polypeptide chain that contains the SH2 domain (*e.g.*, Grb2, p85 subunit of PI-3 kinase). The second class of proteins are enzymes that have SH2 domains in their structure (*e.g.*, Src, PLCγ, SHP-1, SHP-2). The PTPs in this latter class are unique in that by definition they have two different phosphotyrosine-binding motifs, the catalytic domain and the SH2 domain.

---

[1] Abbreviations: PTK, protein tyrosine kinase; PTP, protein tyrosine phosphatase; SH2, Src homology 2; pY, phosphotyrosyl; PLC, phospholipase C; PDGFR, platelet-derived growth factor receptor ß; GT, glutathione; GST, glutathione *S*-transferase; GST–SH2, fusion of GST with the SH2 domains of SHP-2; GST–SHP-2, fusion of GST with SHP-2; GST–SHP-2•C→S, fusion of GST with the catalytically-inactive SHP-2 (Cys_{459}→Ser); GST–[ΔSH2]SHP-2•C→S, fusion of GST with the mutated catalytic domain of SHP-2 (without the SH2 domains); SHP-2•C→S, thrombin-cleaved form of the GST fusion; PBS, phosphate-buffered saline; PMSF, phenylmethylsulfonyl fluoride; BAEE, N^{α}-benzoyl-L-arginine ethyl ester; IRS-1, insulin receptor substrate-1; IRP, insulin receptor protein; RCM lysozyme, reduced, carboxyamidomethylated and maleyated lysozyme.

The crystal structures of Src and Lck SH2 domains with binding peptides revealed that the major sites of interaction with their pY peptide targets are at the pY residue and a hydrophobic residue three positions C-terminal to pY (pY+3), as well as with the residue at pY+1 (6-9). The peptide fits into the SH2 domain like a two-pronged plug fitting into a socket. Recently the structure of the PLCγ SH2 domain has been solved (10). This SH2 domain has a hydrophobic peptide binding groove, and contacts are made with the pY peptide target at the same residues described for the Src SH2 domain (pY, pY+1, and pY+3) as well as residues outside this region, especially pY+4 to pY+6. These data suggest that the PLCγ SH2 domain may define another class of SH2 domain. Binding specificity data using peptide library screens are also consistent with different subtypes of SH2 domains (11).

Recently we and others have cloned a widely expressed SH2 domain-containing tyrosine phosphatase known as SH-PTP2 (12, 13) and also referred to as PTP2C (14), SH-PTP3 (15), Syp (16), and PTP1D (17). To standardize the nomenclature, this enzyme is now referred to as SHP-2. It has been shown that this protein associates through its SH2 domains with PDGFR ß subunit at pY residue 1009 (18). In this paper we have studied the binding of the SH2 domains alone, as well as a catalytically inactive full-length SHP-2, to an immobilized 13mer peptide of the PDGFR containing the pY residue 1009. A competitive binding assay was also developed for measuring the relative affinities of pY peptides to the SH2 domains. The results from this study show that two amino acids N-terminal, and three amino acids C-terminal, to the phosphotyrosine are essential for high affinity binding, with the major interactions at pY+1, pY+3, and pY-2, in addition to the

phosphotyrosine. Furthermore, the minimal peptide that binds with high affinity to the SH2 domains also stimulates the phosphatase activity of the enzyme.

EXPERIMENTAL PROCEDURES

Materials – All chemicals used were of reagent grade from Sigma. Triton X-100, biotin, and streptavidin were from Pierce. Anti-phosphotyrosine antibody was from UBI. GT-Sepharose was purchased from Pharmacia as a suspension in 20% ethanol, and GT-agarose was from Sigma. [γ-^{32}P]ATP was from NEN, and Aquasol scintillation fluid was from DuPont. Protein determination was by the method of Bradford (19), using a kit from Bio-Rad.

Preparation of pY Peptides – The pY peptides were synthesized and purified by California Peptide Research Inc. (Napa, CA). The phosphopeptides with free carboxyl terminals were synthesized using Boc-amino acids while the carboxy amide peptides were synthesized with Fmoc-amino acids. Fmoc dimethyl phosphotyrosine derivative was used for both syntheses. Mass spectral analysis and amino acid composition analysis gave the expected results.

Biotinylation of the pY Peptide – About 0.7 mg of the PDGFR 1009 1-13 pY peptide [DTSSVL(pY)TAVQPN] was dissolved in 0.6 ml of 100 mM sodium phosphate (pH 8.0) and incubated with a 15-fold molar excess of NHS-LC-biotin [sulfosuccinimidyl-6-(biotinamido)hexanoate] for 2 h at room temperature, at which point Tris was added to a final concentration of 0.1 M. The mixture was incubated for 30 min and then acidified

with TFA, and the biotinylated peptide was purified by reverse phase HPLC. Mass spectral analysis confirmed that there was one mole of biotin per mole peptide.

Preparation of GST Fusion Proteins – GST fusion proteins of SHP-2 (GST–SH2, GST–SHP-2) and a catalytically-inactive form in which Cys_{459} of the catalytic domain has been changed to Ser (GST–SHP-2•C→S, GST–[ΔSH2]SHP-2•C→S) have been reported previously (12, 20). The SHP-2 sequences fused to GST were as follows: full-length fusions, all residues except the first three (*i.e.*, residues 4-593); GST–SH2 (SH2 domains), residues 4-208; and GST–[ΔSH2]SHP-2•C→S (mutated catalytic domain alone), residues 210-593. *Escherichia coli* cells expressing the various GST fusions were grown in Terrific Broth + ampicillin (100 μg/ml) at 37°C to A_{595} ~0.6. The temperature of the culture was dropped to 27°C and expression was induced with 50 μM IPTG overnight at 27°C. Cells were harvested and resuspended in ice-cold lysis buffer (PBS containing 1 mM PMSF, 1 mM BAEE, 10 μg/ml aprotinin, 10 μg/ml leupeptin, 5 mM DTT, and 1% Triton X-100). The cells were either frozen in liquid nitrogen and stored at -80°C or lysed immediately. Lysis was performed using a Bead Beater (Biospec Products) with 0.1 mm diameter glass beads. Fusion protein was purified by adding GT-Sepharose to the cleared lysate and incubating for 1 h at 4°C with end-over-end mixing. The beads were pelleted and washed extensively with ice-cold PBS. Purity of the preparation was assessed by SDS-PAGE by the method of Laemmli (21) followed by Coomassie staining. The concentration of fusion protein on the beads was estimated by visual comparison of the Coomassie-stained SDS-PAGE to ovalbumin standards. Thrombin digestion to remove the GST from the GST–SHP-2 and GST–SHP-2•C→S fusions was carried out on the GT-

Sepharose beads essentially as described (20). Fusion proteins were eluted from the GT-Sepharose beads with 10 mM GT.

Real Time Binding Measurements With BIAcore – All experiments were carried out in 20 mM HEPES (pH 7.3), 150 mM NaCl, 5 mM EDTA, 5 mM DTT, and 0.005% Tween 20. All of the proteins were desalted before use in the above buffer. Streptavidin (50 µg/ml) in 20 mM sodium acetate (pH 4.5) was immobilized onto the sensor chip according to the manufacturer's instructions using *N*-hydroxysulfosuccinimide (NHS) and *N*-ethyl-*N'*-(dimethylpropyl)carbodiimide (EDC). The excess reactive groups were blocked using 1 M ethanolamine. A flow rate of 5 µl/min was used for all experiments. Routinely 15 µl of biotinylated peptide (5 µg/ml) was injected onto the chip. Binding of antiphosphotyrosine monoclonal antibody (injected at 10 µg/ml) resulted in a 1100-1500 change in resonance units in various experiments. The sensor surface was regenerated at the end of each experiment with 100 mM HCl which did not dissociate the high affinity biotin-streptavidin interaction. This was confirmed by the reproducibility of the change in resonance units upon binding of antiphosphotyrosine monoclonal antibody to the pY peptide on the sensor chip after several cycles of regeneration.

Phosphorylation of Peptide Substrate – A GST-tyrosine kinase fusion protein immobilized on GT-agarose beads[2] was used to phosphorylate the PDGFR 1009 1-13 peptide. 20 µl of the beads were washed in reaction buffer (50 mM imidazole pH 7.2, 10 mM DTT, 30 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM sodium orthovanadate, and 0.05% Triton X-100) and mixed with 40 nmol peptide substrate, 370 µCi [γ-^{32}P]ATP (3000 Ci/mmol), and 60 to 80

[2] W.R. Huckle and H.S. Earp, manuscript in preparation.

nmol of unlabeled ATP, in a final volume of 210 µl. The reaction was allowed to proceed overnight at room temperature with end-over-end mixing and then diluted in 50 mM imidazole pH 7.2 plus 10 mM DTT, with 100% TFA added to a final concentration of 10%. The sample was loaded on a C_{18} column (Extract-Clean), and the column was washed with 0.1% TFA. Peptide was eluted with 0.1% TFA in 50% acetonitrile, dried, and resuspended in dH_2O. The stoichiometry of phosphorylation was about 22%. Quality of phosphopeptide product was assessed by HPLC (data not shown) and phosphoamino acid analysis that identified phosphotyrosine as the only phosphoamino acid species present (data not shown).

Competitive Assay for SH2 Domain Binding – Fusion protein immobilized on GT-Sepharose beads was mixed with GT-agarose or GT-Sepharose beads to a concentration of about 22 pmol fusion protein per µl beads. Binding assays were carried out using the Millipore MultiScreen Assay System, which uses 96-well plates with a 0.45 µm Durapore membrane on the bottom of the wells. A typical well contained 12.5 µl of beads with the immobilized fusion protein (22 pmol/µl beads), 0.2 µM ^{32}P-labeled pY peptide, and 75 µg/ml BSA, to a final volume of 100 µl in PBS. Binding was allowed to proceed for 15 min at room temperature with shaking to keep the beads from settling. Unbound peptide was washed away on the Millipore vacuum manifold with 2, 100 µl washes with ice-cold PBS. The beads and filters were allowed to air dry and then punched out into scintillation vials with the Millipore multiple punch apparatus. Aquasol scintillation fluid was added and samples were counted in a Beckman LS6000IC scintillation counter.

Phosphatase Kinetic Assays – RCM lysozyme was prepared and phosphorylated with [γ-^{32}P]ATP as described (22) using wheat germ lectin-purified human insulin receptor prepared from an over-expressing cell line (23). ^{32}P-labeled RCM lysozyme was desalted on a PD-10 Sephadex G-25M column from Pharmacia. The tyrosine phosphatase activity was measured by following the release of ^{32}P from ^{32}P-labeled RCM lysozyme. The assay mixture contained 50 mM imidazole pH 7.2, 5 mM DTT, 1 mM EDTA, 100 mM NaCl, 100 µg/ml BSA, and 1 µM ^{32}P-labeled RCM lysozyme, in a 50 µl volume. Enzyme was added to initiate the reaction (80 ng of full-length or 40 ng of [ΔSH2]SHP-2), and the mixture was incubated at 25°C for 10 min. The reaction mixture was stopped by adding 25 µl of cold TCA followed by 25 µl of 10 mg/ml BSA, then incubated on ice for 10 min and centrifuged at 16,000 x *g* for 5 min. Released ^{32}P was determined by adding 75 µl of the supernatant to 10 ml Aquasol scintillant and counting in a Beckman LS6000IC scintillation counter. The reaction rates were linear under these conditions. Typical rates measured were 0.82 nmol/mg/min for full-length SHP-2 and 19.6 nmol/mg/min for [ΔSH2]SHP-2.

RESULTS

Interaction of PDGFR 1009 pY Peptide With SHP-2 – The SH2 domains of SHP-2 have been shown to interact *in vivo* with the pY residue at position 1009 of the PDGF receptor ß chain (18). To measure the kinetics of this binding interaction, a 13 amino acid pY peptide centered around phosphotyrosine 1009 was chemically synthesized, biotinylated, and immobilized through streptavidin on the dextran coated BIAcore sensor chip. The

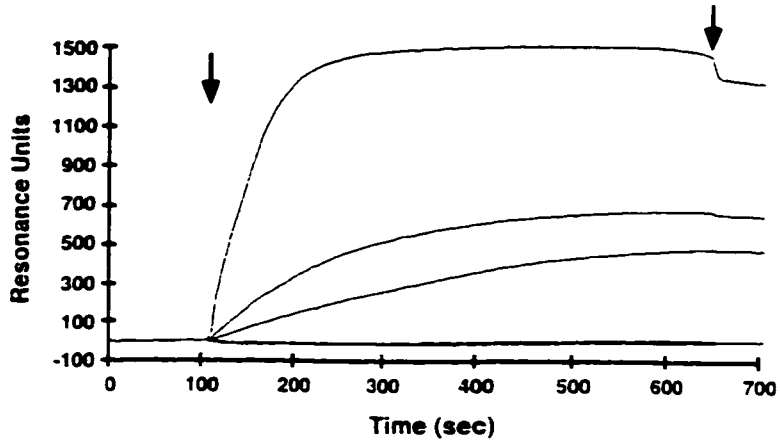


Fig. 2.1. **Overlay sensograms for the binding of various GST–SHP-2 fusions to immobilized PDGFR 1009 1-13 pY peptide.** Biotinylated PDGFR 1009 1-13 pY peptide was bound to the sensor surface through immobilized streptavidin as described under "Experimental Procedures." All proteins were injected at 10 μg/ml in running buffer. The plots from top to bottom are for anti-phosphotyrosine monoclonal antibody, GST–SHP-2•C→S, GST–SH2, GST–[ΔSH2]SHP-2•C→S, and GST. First arrow, injection of protein; second arrow, wash with running buffer.

interaction of various proteins with the pY peptide on the sensor chip can be followed continuously by surface plasmon resonance. Fig. 2.1 shows an overlay sensogram when various proteins were injected over the sensor chip. GST–SH2, GST–SHP-2•C→S, and anti-phosphotyrosine antibody all bound to the sensor surface whereas GST alone and GST–[ΔSH2]SHP-2•C→S did not show any binding. Dephosphorylation of the phosphopeptide by injecting active tyrosine phosphatase eliminated the binding, and part of the binding could be restored by phosphorylation of the immobilized dephosphorylated peptide with tyrosine kinase (data not shown). These findings confirm that the SH2 domains of SHP-2 specifically interact with the pY residue of the PDGFR 1009 pY peptide. It was necessary to use the catalytically-inactive Cys_{459}→Ser mutant fusions to prevent the dephosphorylation of the pY peptide by SHP-2. In this mutant the active site

Cys, which forms a thiol-phosphate intermediate during dephosphorylation, is changed to Ser, rendering it catalytically inactive.

To determine the association and dissociation rate constants of the SH2 domains, varying concentrations of the SH2 domain-containing protein constructs were injected over the same surface, regenerating the surface at the end of each injection. The overlay plot for the GST–SH2 interaction with the peptide at different concentrations of protein is shown in Fig. 2.2a. In order to determine the association rate constants, the dR/dt *vs.* resonance units values were plotted (Fig. 2.2b). The slope of the plot of dR/dt *vs.* concentration of protein injected (Fig. 2.2c) accurately estimates the association rate constant. After the completion of protein injection (second arrow in Fig. 2.2a), when there is running buffer flowing over the sensor surface, the protein appears to remain on the surface; *i.e.*, there is very little apparent dissociation of the protein. Increasing the flow rate did not modify the rate of dissociation. The possibility that the protein actually dissociated from the peptide but bound again to the peptide surface before it was removed by the running buffer was examined by injecting non-biotinylated peptide over the sensor surface (Fig. 2.2d). Increasing concentrations of the non-biotinylated peptide caused a dramatic increase in the rate of dissociation, reaching a maximum at about 100 μg/ml of non-biotinylated peptide. The dissociation followed first order kinetics. The summary of the association and dissociation rate constants and the K_d for GST–SH2, GST–SHP-2•C→S, and SHP-2•C→S are shown in Table 2.1. The rate constants for dissociation for all three proteins were similar, ranging from 0.011 to 0.015 s^{-1}. However, the rate of association for GST–SH2 was 0.6 x 10^5 $M^{-1}s^{-1}$, approximately 9- to 12-fold lower than

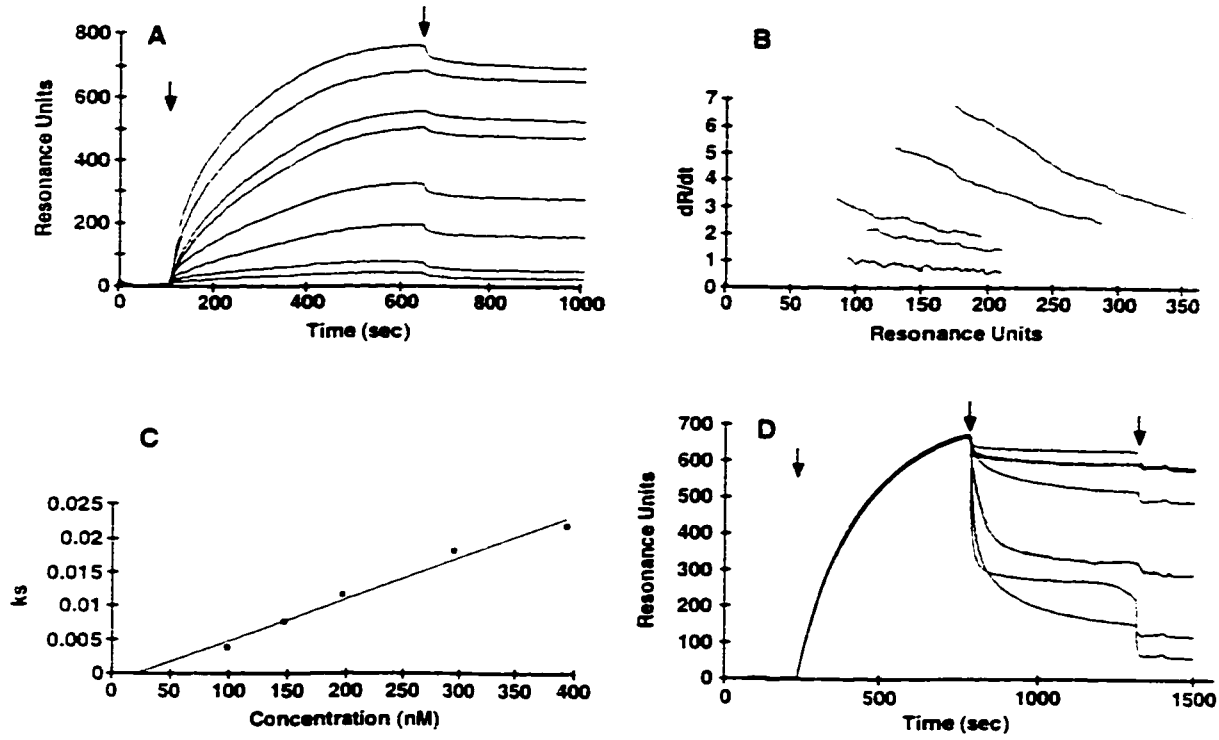


Fig. 2.2. **Kinetics of binding of GST–SH2 to the PDGFR 1009 1-13 pY peptide.** (A) Overlay plots showing change in resonance units with time as varying concentrations of GST–SH2 are injected over the sensor chip. Concentrations of GST–SH2 from top to bottom are (in nM) 392, 294, 196, 147, 98, 49, 25, and 12.5. First arrow, injection of protein; second arrow, wash with running buffer. (B) Plot of the rate of change of resonance units (dR/dt) *vs.* resonance units, as calculated by the BIAcore software. The initial fast change in the refractive index is not included as this is due to the bulk flow of the buffer. (C) Replot of the slopes of the curves in "B" *vs.* concentration of GST–SH2 injected. (D) Dissociation of bound GST–SH2 from the biotinylated peptide on the sensor chip by injection of varying concentrations of non-biotinylated peptide. GST–SH2 was injected at 196 nM at the point shown by the first arrow. Non-biotinylated PDGFR 1009 1-13 pY peptide was injected in running buffer at the second arrow, using the second loop on the BIAcore, at the following concentrations (from top to bottom, in μg/ml): 0, 0.1, 1, 10, 20, 50, and 100. Third arrow, wash with running buffer.

that for GST–SHP-2•C→S or SHP-2•C→S. The dissociation constant for GST–SH2 was 242 nM while the values for GST–SHP-2•C→S and SHP-2•C→S were 19 nM and 18 nM, respectively. The dissociation constants reported here are similar to the values reported earlier using the BIAcore for other SH2 domain-containing proteins (24-26). The binding was inhibited by preincubation of the proteins with non-biotinylated immobilized peptide, or with phosphotyrosine or phenyl phosphate, while non-pY PDGFR 1009 peptide had no effect on binding (data not shown). The binding of GST–SH2 as measured by the change in resonance units was not affected by mixing GST with GST–SH2 before injection (data not shown); as well, the binding constants for GST–SHP-2•C→S and SHP-2•C→S are very similar (Table 2.1). These observations are consistent with the GST fusions binding only as monomers.

Table 2.1. **Association and dissociation rate constants for the interaction of the SH2 domains of SHP-2 with PDGFR 1009 1-13 pY peptide.**

Protein	$k_a \pm$ S.E. (10^{-5} $M^{-1}s^{-1}$)	$k_d \pm$ S.E. (s^{-1})	K_d (nM)
GST–SH2	0.6 ± 0.05	0.015 ± 0.001	242
GST–SHP-2•C→S	5.7 ± 0.9	0.011 ± 0.0002	19
SHP-2•C→S	7.4 ± 0.2	0.014 ± 0.001	18

SH2 Binding Assay – PDGFR 1009 1-13 peptide (DTSSVLYTAVQPN) was phosphorylated on the tyrosine by tyrosine kinase with [γ-^{32}P]ATP, and the radiolabeled peptide was used as a ligand for binding experiments. Preliminary binding experiments were done by mixing the SH2 domains immobilized on GT-Sepharose beads with the peptide ligand in microfuge tubes. Beads were pelleted by centrifugation, the supernatant

was removed, and the beads were washed to remove unbound peptide. However, a large amount of the bound peptide dissociated during the washes; as a result, the signal for peptide retained by the immobilized proteins was only about five-fold above the background. As an alternative, the Millipore MultiScreen Assay system, using 96-well plates with a 0.45 μm membrane at the bottom of the wells, was adopted. GST fusion protein immobilized on GT-Sepharose beads is incubated with the peptide in the wells, and then unbound peptide is washed away by filtration through the membrane on a vacuum manifold. This method keeps the wash times to a minimum as the washes are immediately removed by vacuum filtration through the membrane. While there is still some dissociation of peptide, the result was a much-improved signal to noise ratio for binding as well as greater reproducibility between replicates. This system also facilitates screening many samples and conditions at the same time. Binding experiments with the various GST fusion proteins immobilized on GT-Sepharose beads show that the binding is specific for the SH2 domain, since binding of the radiolabeled peptide to GST–SH2 and GST–SHP-2•C→S is approximately 40- and 150-fold, respectively, greater than to GST or GST-[ΔSH2]SHP-2•C→S (Fig. 2.3a). The binding with the full-length mutant enzyme is three- to four-fold above the binding with the SH2 domains alone. As well, binding to GST–SH2 reaches maximal levels within 5 minutes after addition of peptide, while for GST–SHP-2•C→S, it takes up to 30 minutes to achieve maximum binding. The time course data also show how reproducible the results are between replicates: with four replicates at each time point, the standard error was generally less than 5% for GST–SH2 and GST–SHP-2•C→S.

Unlabeled PDGFR 1009 1-13 pY peptide competed with the ^{32}P-labeled 1-13 peptide for binding to the SH2 domains, as indicated by a decrease in the cpm retained by the immobilized GST–SH2 fusion as the concentration of unlabeled peptide was increased. Using this method, the inhibitory concentration that reduces the binding of the radio-labeled peptide to the SH2 domains by 50% (IC_{50}) can be determined (Fig. 2.3b). The low value (7.5 μM) is consistent with the tight binding of the SH2 domains to this peptide.

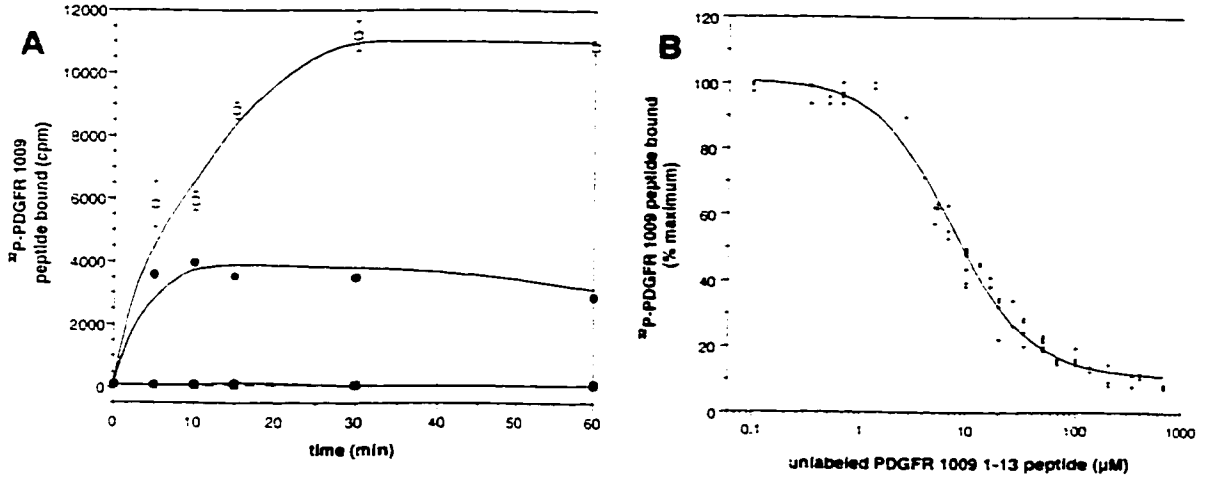


Fig. 2.3. **SH2 domain binding assay using radiolabeled PDGFR peptide.** (A) Time course of binding. ^{32}P-labeled PDGFR 1009 1-13 pY peptide [DTSSVL(pY)TAVQPN] was added to GST (—○—), GST–SH2 (—●—), GST–SHP-2•C→S (—❑—), or GST–[ΔSH2]SHP-2•C→S (– – ■ – –) immobilized on GT-Sepharose beads as described under "Experimental Procedures." "Peptide bound" refers to the cpm retained by the beads after binding and washing, as determined by scintillation counting. The results are the average of four replicates at each time point. (B) Determination of the IC_{50} for the PDGFR 1009 1-13 pY peptide. Binding assays were performed as described under "Experimental Procedures." The data plotted are from four separate experiments, with the amount of peptide bound at each concentration expressed as a % of the maximum binding in that experiment. The IC_{50} was determined by fitting the curve to the equation $\mathbf{y = (m1-m2)\div[1+(x\div IC_{50})^{m4}]+m2}$ where m1 and m2 are the minimum and maximum y values, respectively, and m4 is the slope at the point of inflection (~-1).

Determination of Sequence Recognition Elements for SH2 Domains of SHP-2 – The minimum peptide sequence necessary and sufficient for SH2 domain binding was determined by systematically testing N- and C-terminal truncated versions of the PDGFR 1009 1-14 pY peptide [DTSSVL(pY)TAVQPNE] for their ability to compete with the 1-13 peptide for SH2 domain binding. Both the BIAcore assay and the competitive binding assay were used. In general, the competing peptides were added to the assays at approximately the IC_{50} concentration of the 1-13 peptide. If the truncated peptide is still able to bind to the SH2 domains with a K_d roughly equal to that of the 1-13 peptide, then binding of the 1-13 peptide should be reduced by ~50% in the presence of the truncated peptide; otherwise, no effect should be observed. In the BIAcore assay, the competing peptides were incubated with GST–SH2 for at least 1 h before injecting over the immobilized PDGFR 1009 1-13 pY peptide on the sensor chip. In the competitive assay, the competing peptides were mixed with the radiolabeled PDGFR 1009 1-13 pY peptide and added to the immobilized fusion proteins together. Both approaches gave very similar results. From the N-terminal truncations (Fig. 2.4a), only the two amino acids immediately N-terminal to the pY residue (Val and Leu at pY-2 and pY-1 respectively) are essential for binding: the 5-14 peptide is the smallest that is still able to compete for binding. Similarly, from the C-terminal truncations (Fig. 2.4b), the three amino acids immediately C-terminal to pY (Thr, Ala, and Val at pY+1, pY+2, and pY+3 respectively) are essential for binding: the 1-10 peptide is the smallest that is still able to compete for binding. Similar results were obtained when GST–SHP-2•C→S was used in the two assays (data not shown).

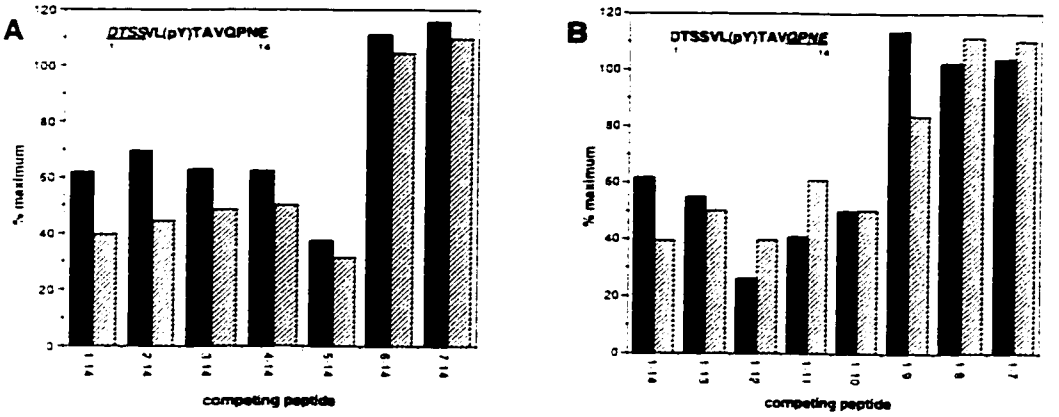


Fig. 2.4. **Competition for SH2 domain binding by radiolabeled PDGFR peptide with various PDGFR 1009 peptides.** Binding competition with various N-terminal (panel A) and C-terminal (panel B) truncated PDGFR 1009 pY peptides. Solid bars, competitive binding assay results; striped bars, BIAcore assay results. Peptides were added to the competitive assay at a concentration of 8.9 μM, and the assays were performed as described under "Experimental Procedures." For the BIAcore assay, GST–SH2 (196 nM) was incubated with the peptides at 17 μM for at least 1 h before injecting over the sensor chip, as described under "Experimental Procedures." Results are expressed as % maximum, which for the competitive assay is the amount of radiolabeled peptide bound as a % of the control (no competing peptide), and for the BIAcore assay is the maximal increase in resonance units as a % of the control (no competing peptide).

Together, these results suggest that the minimum PDGFR 1009 peptide still able to bind to the SH2 domains would be the 5-10 peptide; *i.e.*, VL(pY)TAV. This peptide is in fact a good ligand (Fig. 2.5a), with an IC_{50} value about two-fold lower than that of the larger 1-13 peptide (see Table 2.2). In order to identify essential residues or positions for SH2 domain recognition and binding, an Ala scan was performed on this peptide. Four peptides were synthesized, each with one of the residues replaced by Ala (except for pY). The substitutions with the greatest effect on SH2 domain binding and recognition were at positions 5 and 10 (pY-2 and pY+3), as shown by the 12- to 13-fold increase in the IC_{50}

values (Fig. 2.5b, Table 2.2). The substitution at position 8 (pY+1) had a lesser effect on binding, while the substitution at position 6 (pY-1) had essentially no effect. In addition, double substitution of the Val residues at pY-2 and pY+3 with Ala ("5-10 Ala 5,10") gave rise to a very poor competitor, with an IC_{50} >500 μM (Table 2.2).

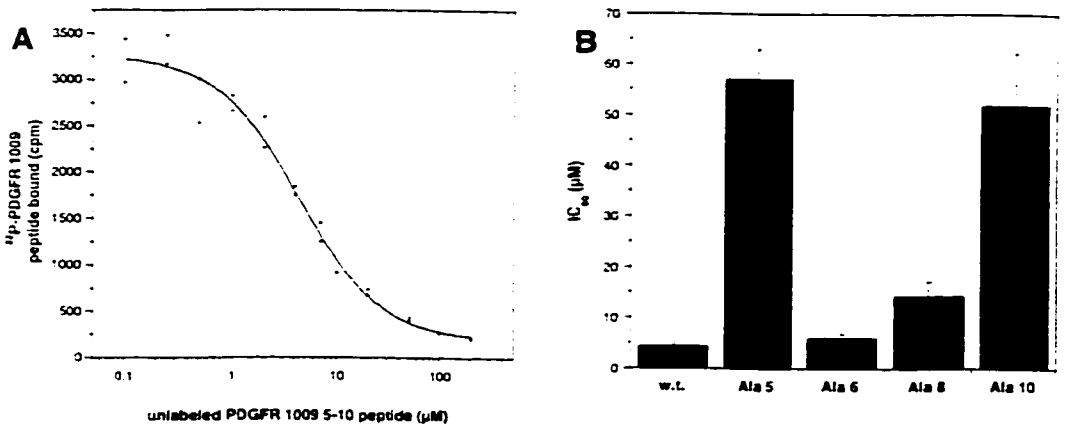


Fig. 2.5. **Competition for SH2 domain binding by radiolabeled PDGFR peptide with the minimum PDGFR peptide (5-10).** (A) Determination of the IC_{50} for the PDGFR 1009 5-10 pY peptide [VL(pY)TAV]. Binding assays and calculations were performed as described under "Experimental Procedures" and in the legend for Fig. 2.3b. (B) Summary of the IC_{50} values for the 5-10 peptide ("w.t.") and the Ala-substituted derivatives.

The second known *in vivo* binding target for SHP-2 is IRS-1, at the pY residue 1172 (27, 28). An 11mer peptide centered around phosphotyrosine 1172 was synthesized and tested in the competitive binding assay (Table 2.2). This peptide was a better competitor than the PDGFR 1009 peptide with an IC_{50} of 0.6 μM. Several other 11mer peptides corresponding to the known SH2 domain binding sites on IRS-1 for Grb2, p85, and Nck were also tested (28, 29) (Table 2.2). Among these, the IRS-1 895 pY peptide (the docking site for Grb2) was an effective competitor. The phosphonomethyl

phenylalanine (PMP) derivative of the best binding peptide (IRS-1 1172) was a poor competitor with an IC_{50} about 80-fold higher than that of the pY peptide (Table 2.2).

Table 2.2. **IC_{50} values for various competitors with the SH2 domains of SHP-2.**

competitor[a]	sequence	IC_{50} ± S.E. (μM)[b]
PDGFR 1009:		
1-13	DTSSVL(pY)TAVQPN	7.5 ± 0.4
1-14	DTSSVL(pY)TAVQPNE	12 ± 2
5-14	VL(pY)TAVQPNE	6.0 ± 1.0
6-14	L(pY)TAVQPNE	97 ± 41
7-14	(pY)TAVQPNE	>500
1-7	DTSSVL(pY)	>500
1-8	DTSSVL(pY)T	>500
1-9	DTSSVL(pY)TA	>500
1-10	DTSSVL(pY)TAV	8.5 ± 1.1
5-10	VL(pY)TAV	4.3 ± 0.5
5-10 Ala 5	AL(pY)TAV	57 ± 6
5-10 Ala 6	VA(pY)TAV	6.0 ± 0.8
5-10 Ala 8	VL(pY)AAV	14 ± 3
5-10 Ala 10	VL(pY)TAA	52 ± 10
5-10 Ala 5,10	AL(pY)TAA	>500
1-13 (nonphospho)	DTSSVLYTAVQPN	>1000
5-10 (nonphospho)	VLYTAV	>1000
PDGFR 1021	GDND(pY)IIPLPD	>500
PDGFR 1009/1021	DTSSVL(pY)TAVQPNE-GDND(pY)IIPLPD	9.2 ± 2.6
IRS-1 147	GEDLS(pY)GDVPP	>200
IRS-1 460	EELSN(pY)ICMGG	61 ± 7
IRS-1 608	HTDDG(pY)MPMSP	>200
IRS-1 895	KSPGE(pY)VNIEF	4.6 ± 1.8
IRS-1 939	TGTEE(pY)MKMDL	>200
IRS-1 1172	ENGLN(pY)IDLDL	0.6 ± 0.05
IRS-1 1172 PMP	ENGLN(PMP)IDLDL	48 ± 12
IRP 5	TRDI(pY)ETDYYRK	>200
IRP 9	TRDIYETD(pY)YRK	>200
phenyl phosphate		2500 ± 500
phosphotyrosine	pY	4300 ± 1100

[a] PDGFR 1009-derived peptides are numbered starting with the aspartic acid at position pY-6 as #1. In the naming of the PDGFR, IRS-1, and IRP peptides, the number refers to the position of the phosphotyrosine in the native protein. All phosphopeptides have carboxy amide at the C-terminal except for IRP-5 and IRP-9 which have carboxyl groups.

[b] The IC_{50} values were determined as described in the legend to Fig. 2.3b.

Phosphotyrosyl peptide library screening also identified the PDGFR 1021 peptide as a potential binding site for the SH2 domains of SHP-2 (11). When this peptide was tested in the competitive binding assay, the IC_{50} was >500 μM. A 25mer peptide encompassing the pY residues at positions 1009 and 1021 of PDGFR was only as effective as the PDGFR 1009 1-13 pY peptide, showing that there is no enhancement in binding with the second phosphotyrosine present (Table 2.2).

Stimulation of SHP-2 Activity by SH2 Domain Occupancy – The phosphatase activity of SHP-2 has been shown to be stimulated when its SH2 domains are bound to a pY-containing peptide (30-32). We decided to test a number of our PDGFR 1009-derived pY peptides for their ability to stimulate the activity of SHP-2, to ensure that this correlates with the binding data. ^{32}P-labeled RCM lysozyme was used as a substrate for the activity assays (22). The amount of stimulation with full-length SHP-2 increases rapidly with increasing concentration of exogenous PDGFR 1009 1-13 peptide, reaching a maximum six- to seven-fold increase at about 50 μM peptide (Fig. 2.6a, closed circles). SHP-2 without the SH2 domains (*i.e.*, the catalytic domain alone) shows no stimulation with exogenous peptide (Fig. 2.6a, open circles). In general, there was a very good correlation between a peptide's capacity to stimulate SHP-2 activity and its ability to bind to the SH2 domains (Fig. 2.6b). PDGFR 1009 peptides which bind well to the SH2 domains [as reflected by their IC_{50} values (Table 2.2)] stimulate the phosphatase activity by four- to five-fold at 50 μM peptide. Peptides that do not bind (7-14 and 1-9) show no stimulation, while peptides that are poor binding ligands (6-14, 5-10 Ala 5, and 5-10 Ala 10) show only about a two-fold stimulation. The activity of SHP-1 (another SH2 domain-

containing PTP) (33-36) was not affected by the above peptides with RCM lysozyme as a substrate (data not shown).

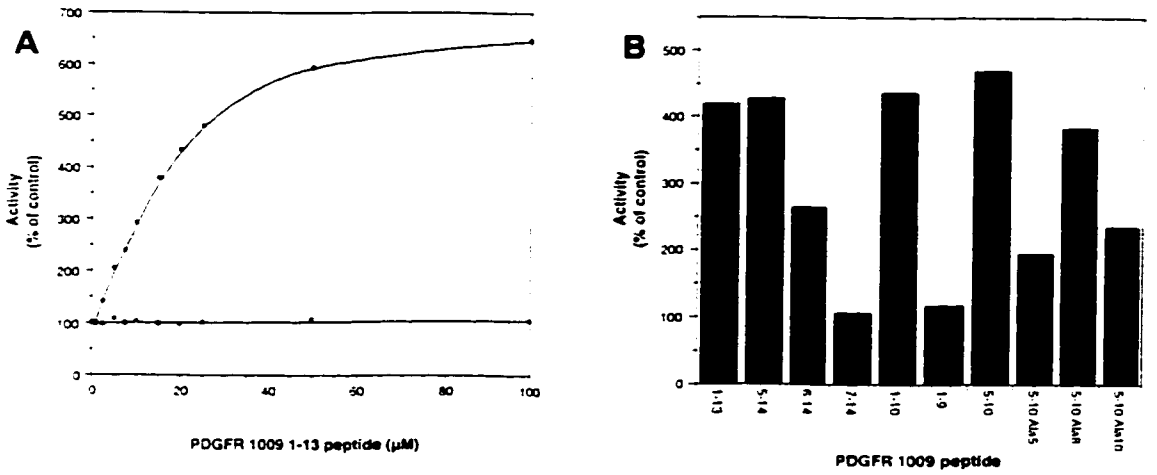


Fig. 2.6. **Stimulation of SHP-2 phosphatase activity by PDGFR 1009-derived peptides.** All assays were performed as described under "Experimental Procedures." (A) Stimulation of SHP-2 phosphatase activity as a function of concentration of exogenous peptide added. PDGFR 1009 1-13 pY peptide was added to the reaction mixture at various concentrations and the activities were determined from the amount of ^{32}P released. Closed circles, full-length SHP-2; open circles, [ΔSH2]SHP-2 (catalytic domain alone without the SH2 domains). Activities are expressed as % of control, which is the activity with no exogenous peptide added. (B) Effect of various PDGFR 1009-derived peptides on SHP-2 phosphatase activity. The indicated PDGFR 1009 pY peptides were added to the reaction mixture at 50 μM each and the activities were determined. Activities are expressed as % of control, which is the activity with no exogenous peptide added.

DISCUSSION

SH2 domains provide a means of bringing proteins together in a specific phosphotyrosine-dependent interaction, thereby playing an essential role in the transduction of signals generated both by receptor and non-receptor PTKs. An understanding of the factors that determine the specificity of this interaction will provide a

clearer picture of the nature of the associations and allow predictions of potential interaction partners for a particular SH2 domain. We chose to study the SH2 domains of the protein tyrosine phosphatase SHP-2 and to attempt to determine some of the factors that govern its associations with a known *in vivo* binding site, at phosphotyrosine 1009 of the PDGFR ß subunit.

SH2 domain binding studies were done using the BIAcore (surface plasmon resonance) and using a competitive binding assay. In the BIAcore assay the peptide was immobilized and free protein was injected and allowed to bind, while in the competitive assay the fusion proteins were immobilized and free peptide was allowed to bind. The fact that two different approaches gave very similar results for the relative affinities of various truncated peptides and PDGFR 1009 1-13 pY peptide for SH2 domain binding (Fig. 2.4) provided added validity to the two methods and their results. The SH2 domains of SHP-2 have a very high affinity for the PDGFR 1009 pY peptide, as indicated by the sub-micromolar dissociation constants (Fig. 2.2, Table 2.1). Interestingly, the full-length Cys_{459}→Ser mutant enzyme has a dissociation constant about 13-fold lower than that for the SH2 domains alone (*i.e.*, with the catalytic domain removed). This is not due to binding of the PDGFR 1009 pY peptide to the inactive catalytic domain, because the catalytically inactive mutant without the SH2 domains (GST–[ΔSH2]SHP-2•C→S) did not bind any peptide (Fig. 2.1). Similar differences in the behaviour of the full-length mutant enzyme *vs.* the SH2 domains alone were observed in the competitive assay. The time course for peptide binding to the SH2 domains (Fig. 2.3a) showed that binding to GST–SH2 was maximal within five minutes of the addition of peptide, while for the full-length

mutant (GST–SHP-2•C→S) it took up to 30 minutes for maximal binding to be reached. The amount of peptide bound to the full-length mutant at the concentrations used was also three- to four-fold greater than that for GST–SH2. In both types of experiment, the presence of the catalytic domain had an effect on the binding properties of the SH2 domains: the structure of the full-length enzyme was more stable or in a more favourable conformation for binding than the SH2 domains alone.

Structures of SH2 domains complexed with their binding peptides have been determined by X-ray crystallography and NMR (6-10). These studies have shown that most SH2 domains have a conserved Arg residue at the αA2 position in SH2 domains that interacts with the phenyl ring of the pY residue through the ε-NH_2 group. However, in SH2 domain-containing PTPs (SHP-1, SHP-2, and corkscrew), this Arg is replaced by Gly (12-17, 33-37). This suggests that the interactions with the target peptide for this class of SH2 domain may be different. To determine the important residues, we undertook a systematic approach using N- and C-terminal deletion peptides based on the PDGFR 1009 peptide sequence DTSSVL(pY)TAVQPNE (Fig. 2.4). These studies indicated that the minimum peptide sequence necessary and sufficient for binding to the SH2 domains was VL(pY)TAV. Structural data have shown that with SH2 domains of the Src family, the residues which make the most important contacts with the SH2 domain are the phosphotyrosine and the residues at positions pY+1 and pY+3 (6-9). In the case of the PLCγ SH2 domain, contacts have been shown to occur with residues at pY+4 to pY+6 in addition to the contacts shown for the Src family (10). None of these studies has implicated residues N-terminal to the phosphotyrosine as being important in recognition

and binding. In our study, though, it is clear that two residues N-terminal to the phosphotyrosine are necessary for binding to the SH2 domains of SHP-2 (Fig. 2.4). The Ala scan performed on this minimal peptide (Fig. 2.5) indicated that the Val residue at position pY-2, as well as the Val residue at pY+3, are critical for recognition and binding: replacement of either by Ala results in an approximately 12- to 13-fold increase in the IC_{50} value for the peptides, and replacement of both Val residues with Ala results in a >100-fold increase (Table 2.2). The Thr at pY+1 is also important, as replacement by Ala leads to a three-to four-fold increase in the IC_{50} value. Recently the crystal structure of the amino-terminal SH2 domain of SHP-2 complexed with various peptides was determined (38), in which there appear to be important interactions between the SH2 domain and the peptide at the pY residue and residues at pY+1, pY+3, and pY+5. Our results with the PDGFR 1009 deletion peptides do not suggest that the pY+5 position is necessary for high-affinity binding (Fig. 2.4, Table 2.2). While the crystal structure shows interactions at the pY+5 site, the contribution of this site to the total binding energy may be minor compared to that of the minimum peptide (pY-2 to pY+3). The importance of the contacts between the peptide at pY+5 and the SH2 domain may also depend on the peptide ligand, as the pY+5 residue (Phe) in the IRS-1 895 peptide interacts more extensively with the SH2 domain than does the pY+5 residue (Pro) in PDGFR 1009 (38). Interestingly, the crystal structure of the SH2 domain with the PDGFR 1009 peptide also shows that the Val at pY-2 covers the face of the phosphotyrosine ring. This position is occupied by the guanidinium group of the Arg at αA2 in other SH2 domains, but in SHP-2 this Arg is replaced by Gly (see above). Thus, it is not unreasonable from the structure

that this position in the PDGFR 1009 peptide should also have a role in maintaining a high-affinity interaction with the SH2 domains of SHP-2.

Results for the binding with the PDGFR 1021 peptide and other non-PDGFR peptides (Table 2.2) suggest some basic rules for peptide recognition and binding by the SH2 domains of SHP-2. In general, there appear to be three critical positions in addition to the phosphotyrosine: pY+1, pY+3, and pY-2. The Ala-substitutions in the PDGFR 1009 5-10 peptide that changed any of the residues at these positions resulted in a much poorer ligand for SH2 domain recognition and binding. By comparing the sequences of the three peptides that bind well to the SH2 domains (PDGFR 1009, IRS-1 1172, and IRS-1 895), some preliminary predictions can be made. First, a ß-branched residue is required at position pY+1: the three peptides have either Thr, Ile, or Val at this position. Second, a hydrophobic residue with an aliphatic side chain is required at position pY+3: the three peptides have either Val, Leu, or Ile at this position. Third, the pY-2 position is critical, with Val, Leu, and Gly present here in the high-affinity binding peptides. A general rule for what residues would be tolerated at this position is not, however, immediately apparent. The other non-PDGFR peptides tested that do not bind have deviations from these rules at the critical positions. Two peptides in particular are noteworthy. First, the IRS-1 147 peptide has a very similar sequence to the IRS-1 1172 peptide in the region from pY-2 to pY+3, including hydrophobic residues at pY-2 and pY+3. But IRS-1 147 has a Gly at pY+1, which is not a ß-branched residue: this difference appears to be enough so as not to make IRS-1 147 a binding ligand for the SH2 domains of SHP-2. Second, the IRS-1 460 peptide has a ß-branched residue (Ile) at pY+1

and a hydrophobic residue (Met) at pY+3, but has a polar residue (Ser) at pY-2. This peptide is poorly recognized by the SH2 domains, with an IC_{50} of 61 μM. These rules are derived from a small number of peptides: binding experiments with more peptides are required to see how broadly applicable these rules are.

A final obvious requirement for recognition and binding is the pY residue. Nonphosphorylated PDGFR 1009 1-13 and 5-10 peptides were not recognized by the SH2 domains of SHP-2, with IC_{50} values >1 mM (Table 2.2). The pY residue alone is not sufficient for binding, though, since both phosphotyrosine and phenyl phosphate are very poor competitors in the binding assay, with IC_{50} values of 4.3 mM and 2.5 mM, respectively (Table 2.2). Phosphonomethyl phenylalanine (PMP), a phosphotyrosine analogue, does not substitute well for phosphotyrosine in the IRS-1 1172 peptide: the IC_{50} value for this PMP peptide is 48 μM, or 80-fold greater than that for the pY peptide (Table 2.2). This difference may be because the pK_{a2} of the phosphate of PMP is 7.1 as compared to 5.7 for phosphotyrosine (39); as well, the substitution of the ester oxygen with a methylene bridge, and the resulting loss of hydrogen bonding between the oxygen and the SH2 domain, may also be important (6, 9).

With SH2 domain-containing PTPs, a stimulation in phosphatase activity is observed when the SH2 domain is complexed with its peptide binding target (30-32). When we tested a number of the PDGFR 1009-derived pY peptides for their ability to stimulate the PTP activity of SHP-2, a good functional correlation was observed with the *in vitro* binding data (Fig. 2.6b). In general, the peptides that were shown to bind to the SH2 domains in the BIAcore and competitive binding assays were also able to stimulate

the phosphatase activity of SHP-2 by as much as five- to seven-fold at the concentrations used in the experiments. This stimulation occurred through the SH2 domains, since the catalytic domain alone without the SH2 domains was not stimulated by the addition of exogenous peptides. In fact, removing the SH2 domains in itself results in a higher specific activity (20), suggesting that the SH2 domains, when not bound to a peptide target, inhibit the phosphatase activity. Therefore, the stimulation observed with the full-length enzyme is really a relief of inhibition. Another SH2 domain-containing PTP, SHP-1, was not stimulated by the addition of peptides that are bound by the SH2 domains of SHP-2, showing that the stimulation by these peptides is specific for SHP-2. This is not surprising since different SH2 domains generally recognize different peptide sequences: peptides that are bound by the SH2 domains of SHP-2 would not necessarily be bound by the SH2 domain of SHP-1. It is interesting to note that binding of peptides at the SH2 domains stimulates phosphatase activity of SHP-2, and that the presence of the catalytic domain enhances binding of target peptide to the SH2 domains (see above). Perhaps both the SH2 domain and the catalytic domain are able to allosterically modulate each other's function.

While this work was being completed, another study on the binding of the SH2 domains of SHP-2 was presented (40) in which a binding assay with ^{125}I-labeled PDGFR 1009 peptide was used, with competitor peptides derived from phosphotyrosine sites in PDGFR, EGFR, and IRS-1. The nature of this assay resulted in a signal to noise ratio of only about five or ten to one, while the assay described in this paper has a signal at least 40-fold above the background, resulting in a much higher sensitivity. The use here of a

^{32}P-labeled peptide also has the advantage of not needing to modify the peptide substrate for the assay, as the radioactive phosphate is incorporated exactly where the non-radioactive phosphate would be. ^{125}I labeling requires an ε-amino group for the best labeling, and if the peptide does not have a Lys residue in the native sequence, this must be added to the peptide substrate. The authors of the other study concluded that the sequence recognition requirements for the N-terminal SH2 domain were, in addition to the phosphotyrosine, (i) a ß-branched residue at pY+1, and (ii) a hydrophobic residue at pY+3. Our results using both SH2 domains together support this finding, but in addition to this, the pY-2 position is shown to be critical for recognition and binding. This is the first report of a specific requirement in the region N-terminal to the phosphotyrosine: all other reports have only identified residues C-terminal to the phosphotyrosine. From the truncated PDGFR 1009 peptide studies, it was determined that the minimum peptide necessary and sufficient for SH2 domain recognition and binding is the region from pY-2 to pY+3. This is clearly different from the SH2 domains of Src-family kinases and of PLCγ, and may suggest yet another class of binding mode. Similar studies with other SH2 domains, from both known proteins and ones yet to be identified, will help to clarify the rules that govern the recognition of SH2 domain binding targets and thus explain how specificity is maintained by a small structural element that is used in many different situations and pathways in a cell. Deciphering how this is accomplished may provide clues towards an understanding of the signaling mediated by this class of multifunctional PTPs.

ACKNOWLEDGMENTS

We thank F. Jirik for supplying the PDGFR 1021 and 1009/1021 peptides, D. Steiner for providing the cell line over-expressing the human insulin receptor, and J. Yergey for performing the mass spectroscopy on the biotinylated peptides. We also thank P. Brown for assistance with the BIAcore, and members of the department, especially Z. Huang, K. Abdullah, and M. Gresser, for helpful discussions and critical suggestions. This paper is dedicated to the memory of Sandra Huyer (1935-1992).

REFERENCES

1. Anderson, D., Koch, C. A., Grey, L., Ellis, C., Moran, M. F., and Pawson, T. (1990) *Science* **250,** 979-982

2. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R., and Soltoff, S. (1991) *Cell* **64,** 281-302

3. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991) *Science* **252,** 668-674

4. Pawson, T., and Gish, G. D. (1992) *Cell* **71,** 359-362

5. Birge, R. B., and Hanafusa, H. (1993) *Science* **262,** 1522-1524

6. Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. D., Silverman, L., and Kuriyan, J. (1992) *Nature* **358,** 646-653

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

8. Eck, M. J., Atwell, S. K., Shoelson, S. E., and Harrison, S. C. (1994) *Nature* **368,** 764-769

9. Eck, M. J., Shoelson, S. E., and Harrison, S. C. (1993) *Nature* **362,** 87-91

10. Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., and Forman-Kay, J. D. (1994) *Cell* **77,** 461-472

11. 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., and Cantley, L. C. (1993) *Cell* **72,** 767-778

12. Bastien, L., Ramachandran, C., Liu, S., and Adam, M. (1993) *Biochem. Biophys. Res. Comm.* **196,** 124-133

13. Freeman, R. M., Jr., Plutzky, J., & Neel, B. G. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 11239-11243

14. Ahmad, S., Banville, D., Zhao, Z., Fischer, E. H., and Shen, S. H. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 2197-2201

15. Adachi, M., Sekiya, M., Miyachi, T., Matsuno, K., Hinoda, Y., Imai, K., and Yachi, A. (1992) *FEBS Lett.* **314,** 335-339

16. Feng, G-S., Hui, C.-C., and Pawson, T. (1993) *Science* **259,** 1607-1611

17. Vogel, W., Lammers, R., Huang, J., and Ullrich, A. (1993) *Science* **259,** 1611-1614

18. Lechleider, R. J., Freeman, R. M., Jr., and Neel, B. G. (1993) *J. Biol. Chem.* **268,** 13434-13438

19. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248-254

20. Dechert, U., Adam, M., Harder, K. W., Clark-Lewis, I., and Jirik, F. (1994) *J. Biol. Chem.* **269,** 5602-5611

21. Laemmli, U. K. (1970) *Nature* **227,** 680-685

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

23. Yoshimasa, Y., Paul, J. I., Whittaker, J., and Steiner, D. F. (1990) *J. Biol. Chem.* **265,** 17230-17237

24. Felder, S., Zhou, M., Hu, P., Urena, J., Ullrich, A., Chaudhuri, M., White, M., Shoelson, S. E., and Schlessinger, J. (1993) *Mol. Cell. Biol.* **13,** 1449-1455

25. Panayotou, G., Gish, G., End, P., Truong, O., Gout, I., Dhand, R., Fry, M. J., Hiles, I., Pawson, T., and Waterfield, M. D. (1993) *Mol. Cell. Biol.* **13,** 3567-3576

26. Payne, G., Shoelson, S. E., Gish, G. D., Pawson, T., and Walsh, C. T. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 4902-4906

27. Kuhné, M. R., Pawson, T., Lienhard, G. E., and Feng, G. S. (1993) *J. Biol. Chem.* **268,** 11479-11481

28. Sun, X. J., Crimmins, D. L., Myers, M. G., Jr., Miralpeix, M., and White, M. F. (1993) *Mol. Cell. Biol.* **13,** 7418-7428

29. White, M. F., and Kahn, C. R. (1994) *J. Biol. Chem.* **269,** 1-4

30. Sugimoto, S., Lechleider, R. J., Shoelson, S. E., Neel, B. G., and Walsh, C. T. (1993) *J. Biol. Chem.* **268,** 22771-22776

31. Lechleider, R. J., Sugimoto, S., Bennett, A. M., Kashishian, A. S., Cooper, J. A., Shoelson, S. E., Walsh, C. T., and Neel, B. G. (1993) *J. Biol. Chem.* **268,** 21478-21481

32. Sugimoto, S., Wandless, T. J., Shoelson, S. E., Neel, B. G., and Walsh, C. T. (1994) *J. Biol. Chem.* **269,** 13614-13622

33. Plutzky, J., Neel, B. G., and Rosenberg, R. D. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 1123-1127

34. Matthews, R. J., Bowne, D. B., Flores, E., and Thomas, M. L. (1992) *Mol. Cell. Biol.* **12,** 2396-2405

35. Shen, S. H., Bastien, L., Posner, B. I., and Chrétien, P. (1991) *Nature* **352,** 736-739

36. Yi, T. L., Cleveland, J. L., and Ihle, J. N. (1992) *Mol. Cell. Biol.* **12,** 836-846

37. Perkins, L. A., Larsen, I., and Perrimon, N. (1992) *Cell* **70,** 225-236

38. Lee, C.-H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E., and Kuriyan, J. (1994) *Structure* **2,** 423-438

39. Domchek, S. M., Auger, K. R., Chatterjee, S., Burke, T. R., Jr., and Shoelson, S. E. (1992) *Biochemistry* **31,** 9865-9870

40. Case, R. D., Piccione, E., Wolf, G., Benett, A. M., Lechleider, R. J., Neel, B. G., and Shoelson, S. E. (1994) *J. Biol. Chem.* **269,** 10467-10474

CHAPTER 3

Affinity Selection from Peptide Libraries by the SH2 Domains of SHP-2

SUMMARY

SH2 domains direct the association of proteins by binding to phosphotyrosine (pY) residues in a sequence-specific manner. The SH2 domains of the protein tyrosine phosphatase SHP-2 bind *in vivo* to the platelet-derived growth factor receptor (PDGFR) ß at Tyr_{1009}. In Chapter 2, we showed that the minimum peptide sequence surrounding this pY necessary and sufficient for high-affinity binding is VL(pY)TAV, with critical interactions at the pY-2 and pY+3 positions, and to a lesser extent at pY+1. Interestingly, other SH2 domains do not appear to have a requirement for residues N-terminal to the pY residue for binding. In order to explore the peptide binding interactions in greater detail, we used peptide libraries in an affinity selection procedure. The SH2 domains were expressed as a GST fusion protein, immobilized on GT-Sepharose beads, and used to purify high-affinity peptides from a mixture of 18 peptides in which the pY-2 or pY+3 residue was replaced with all naturally-occurring L-amino acids except Cys and Trp. Analysis of high-affinity peptides by sequencing and by mass spectrometry indicated distinct amino acid preferences at both positions. To validate the analysis, individual peptides were synthesized and tested based on the predictions from the affinity selection with the pY-2 library. While additional work is required to optimize this procedure, the preliminary results demonstrate in principle that affinity selection from peptide libraries coupled with mass spectrometric analysis is feasible and that it should be possible to extend the process to other systems.

INTRODUCTION

The specific interactions of proteins in cellular signaling pathways is critical for the transmission of signals and for maintaining signal specificity. A number of conserved protein domains have been identified that are responsible for bringing proteins together in a specific, regulated fashion. One such domain is the Src homology 2 (SH2)[1] domain, a region of ~100 amino acids that binds to phosphotyrosine (pY) residues in proteins (1-4). The specificity of this association is controlled by the sequence surrounding the pY residue, and the interaction is regulated by the phosphorylation state of the tyrosine residue.

SH2 domains have been identified in as many as 100 proteins (3), including tyrosine kinases, tyrosine phosphatases, phospholipases, and structural (adapter) proteins. A number of crystal structures and NMR structures of SH2 domains, with and without bound peptide ligands, have been determined (for examples, 5-15). From these structures, it is apparent that all SH2 domains share a conserved overall 3° structure. The structures with bound peptides have suggested that the primary interactions between the peptide and the SH2 domain involve residues C-terminal to the pY residue of the peptide. In particular, the residues at pY+1 and pY+3 (*i.e.*, the first and third positions C-terminal to the pY) appear to form the major interactions and would therefore be expected to control the SH2 domain sequence specificity.

The sequence specificity of SH2 domains has been studied *in vitro* with individual pY peptides. A number of approaches have been employed, including the ability of peptides to block SH2 domain–pY protein interactions (16-18); radioactive competitive

[1] Abbreviations: SH2, Src homology 2; pY, phosphotyrosine; GT, glutathione; GST, glutathione *S*-transferase; PBS, phosphate-buffered saline; TFA, trifluoroacetic acid; PDGFR, platelet-derived growth factor receptor ß.

binding assays (19-21); and surface plasmon resonance to measure affinities directly (22-24). However, the use of individual peptides is a labourious, time-consuming, and expensive process. As an alternative, methods have been developed using degenerate peptide libraries; *i.e.*, mixtures of peptides derived from the same sequence in which one or more positions have been substituted with multiple amino acids. In this approach, the SH2 domain of interest is used as an affinity matrix to select high affinity binding peptides from such a library, making it possible to sample many sequence permutations and combinations at once. Since all the peptides bind reversibly at the same site on the SH2 domain, the highest affinity peptides will be selected as they compete out the lower affinity peptides. This technique has been used to define the binding specificities of a number of SH2 domains (25, 26), and consistent with the structural studies, the primary sites of specificity in the target peptide appear to be the pY+1 and pY+3 positions. In general, SH2 domain binding specificities fall into two main classes: (i) pY–hydrophilic–hydrophilic–hydrophobic, and (ii) pY–hydrophobic–X–hydrophobic, where X indicates any residue (25).

Our studies with the SH2 domains of the protein tyrosine phosphatase SHP-2 in Chapter 2 (21) suggested that in addition to the interactions at pY+1 and (more importantly) pY+3, critical interactions were also formed with the pY-2 residue (*i.e.*, the second residue N-terminal to the pY). In order to explore further the SH2 domain–peptide interactions, we adapted the affinity selection procedure developed by Songyang *et al.* (25). The SH2 domains of SHP-2 were expressed as a GST fusion protein, immobilized onto GT-Sepharose beads, and used to select high-affinity peptides from

peptide libraries based on the PDGFR sequence surrounding Tyr_{1009} [TSSVL(pY)TAVQP], in which the pY-2 or pY+3 Val residue was replaced with all naturally-occurring L-amino acids except Cys and Trp. Sequencing and mass spectrometry were used to analyze affinity-selected peptides. Our results suggest that there are distinct amino acids preferences at both positions, supporting our earlier observations that both positions are important. The advantages of mass spectrometry over sequencing for analyzing the affinity-selected peptides are also demonstrated, although the mass spectrometric analysis requires additional controls and optimization.

EXPERIMENTAL PROCEDURES

Materials – Glutathione-Sepharose was obtained from Pharmacia. Peptides and peptide libraries were synthesized by California Peptide Research, Inc. (Napa, CA). All chemicals were of reagent grade from Sigma. Peptide library samples were sequenced by Harvard Microchem (Cambridge, MA).

Table 3.1. **Peptides used in this study.**

peptide name	peptide sequence
PDGFR 1-13	DTSSVL(pY)TAVQPN
PDGFR 5-10	VL(pY)TAV
PDGFR 6-14	(pY)TAVQPNE
PDGFR 5-10 Ala10	VL(pY)TAA
PDGFR 2-12	TSSVL(pY)TAVQP
PDGFR 1-14	DTSSVL(pY)TAVQPNE
Lck Ala9	TEGQ(pY)QPQA

SH2 Affinity Column Protocol – The two SH2 domains of SHP-2 were expressed as a GST fusion protein in the vector pGEX-2T (Pharmacia) and purified as described in Chapter 2 (21). For affinity selection, the procedure developed by Songyang *et al.* (25) was followed, with some modifications. Bio-Spin disposable chromatography columns (Bio-Rad) were packed with 225 µl GT-Sepharose beads containing approximately 30 or 60 nmol of fusion protein and equilibrated with PBS. Samples were loaded onto the columns and allowed to sit for 10 min at room temperature. The columns were then washed with 500 µl ice-cold PBS containing 10 mg/ml blue dextran and 0.5% NP-40, followed by 500 µl ice-cold PBS. A Sep-Pak vacuum manifold (Waters) was used to accelerate the washing process to minimize dissociation of bound peptide. Bound peptides were eluted either by competition with 100 mM phenyl phosphate or by the addition of 5% TFA. As a background control, GT-Sepharose beads with GST alone immobilized at the same concentration were used.

HPLC Analysis of Peptides – HPLC was used in the development of the affinity column protocol to quantitate peptides in the column loads and eluates, as well as to desalt affinity-selected peptides prior to mass spectrometric and sequencing analyses. Peptide samples were acidified with TFA to 5% final concentration and filtered through a Millex GV4 0.22-µm syringe filter (Millipore) prior to injection on an Aquapore reversed-phase C_{18} column. To separate peptides for quantitation, a solvent gradient of 0 – 20% CH_3CN containing 0.07% TFA was applied over 40 min at a flow rate of 0.4 ml/min, and peptides were quantitated by the peak area at 218 nm. To desalt samples, peptides were loaded onto the column, washed with a gradient of 0 – 4% CH_3CN/0.07% TFA over 10 min, and

eluted with 45% CH_3CN/0.07% TFA over 10 min. A single fraction of 2 – 2.5 ml was collected and concentrated to ~400 µl in a Speed-Vac to remove the TFA and CH_3CN.

Mass Spectrometric Analysis of Peptides – The desalted peptides were analyzed by capillary HPLC/continuous flow–liquid secondary ion mass spectrometry. Peptide samples were diluted 5-fold with 0.1% TFA in H_2O containing 1.5% glycerol, and a 10-µl aliquot was injected onto a Valco injection loop on a Waters HPLC system. Peptides were separated on a BDS-Hypersil C_{18} capillary column (0.3 x 100 mm) (Keystone Scientific) at a flow rate of 3 µl/min (reduced from 900 µl/min by a splitter immediately before the column), with a solvent gradient of 10 – 50% CH_3CN containing 0.1% TFA and 1.5% glycerol over 25 min. The column eluate was directed to the frit probe of a JEOL HX110A double-focusing mass spectrometer (EB configuration; JEOL, Boston, MA) equipped with a 10 kV LSIMS source and a cesium ion gun. Analysis was in positive ion mode: ions were produced by bombardment with a beam of Cs^+ ions at 10 keV with the ion source accelerating voltage at 10 kV. The resolution was set at 1000, and the mass spectrometer was scanned at a rate of 10 s from m/z 1000 to 2000. The spectra were background-corrected by subtracting the signal from regions devoid of peptide ions. The peak heights of the singly-protonated peptide ions were used as a quantitative measure of the peptides.

Determination of Peptide Binding Affinities – Relative affinities of pY-peptides for the SH2 domains of SHP-2 were determined in the competitive binding assay described in Chapter 2 (21). A peptide's relative affinity is inversely proportional to its IC_{50} (the concentration required to reduce the binding of the ^{32}P-labeled PDGFR 1-13 peptide by 50%).

RESULTS

Development of Affinity Column Protocol – In general, the protocol used by Songyang *et al.* (18) was followed for the SH2 column binding. For the initial development of the system, the minimum high-affinity PDGFR peptide as determined in Chapter 2, VL(pY)TAV ("5-10") (Table 3.1) was used (21). Even though the GST fusion contained both SH2 domains of SHP-2, the two SH2 domains have different specificities (20, 31) and this PDGFR peptide binds strongly only to the N-terminal SH2 domain. Columns were prepared with 30 nmol GST-SH2 immobilized on 225 µl GT-Sepharose beads and loaded with increasing amounts of peptide to titrate the column capacity. Eluted peptide was run on HPLC and quantitated by determining the peak area at 218 nm and comparing to a standard curve. As shown in Fig. 3.1, saturation was observed at ≥60 nmol of peptide loaded; however, only ~8 nmol of peptide was eluted at the saturating level. The rest of the peptide was likely in the column washes, although some peptide may have been retained nonspecifically on the column. The fact that the column did not saturate at 30 nmol as would be expected from the 1:1 binding stoichiometry may have been because not all of the SH2 domains were available or competent for binding. As well, bound peptide may have been lost during washing. Dissociation during washing is a factor as increasing the number of washes did decrease the amount of peptide bound (data not shown). It is also possible that the phenyl phosphate elution was not quantitative. In theory, it should be possible to account for all of the peptide in the wash and eluate. However, the peptide in the wash could not be quantitated because of the large dilution and interference from wash buffer components.

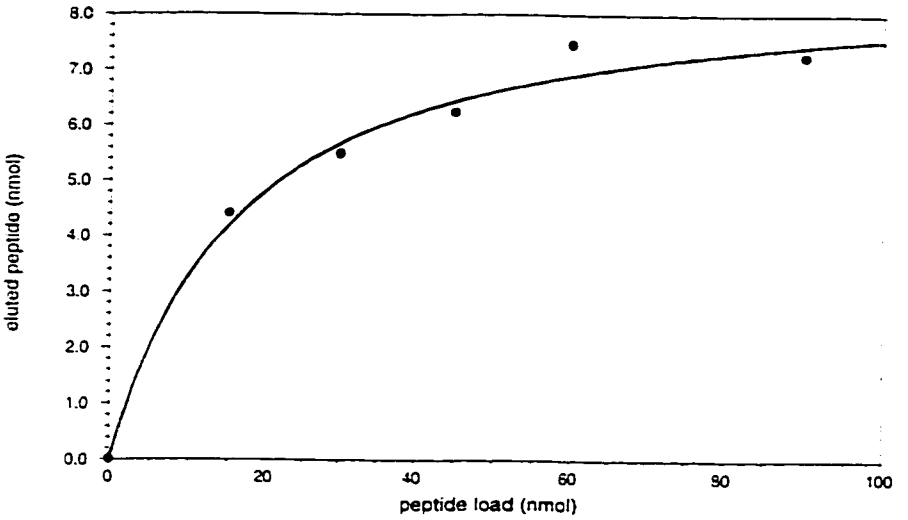


Fig. 3.1. **Titration of GST-SH2 column capacity.** A column with 30 nmol GST-SH2 fusion protein immobilized on GT-Sepharose beads was loaded with the indicated amounts of PDGFR 5-10 peptide as described under "Experimental Procedures." Eluted peptide was quantitated by HPLC from the peak area at 218 nm in comparison to injections of known amounts of peptide.

To determine if peptides were being retained specifically by the column, various mixtures of peptides, consisting of high- and low-affinity pY-peptides, were loaded onto the column. In the first case, a column consisting of 60 nmol of GST-SH2 on 225 µl GT-Sepharose beads was loaded with a mixture of 30 nmol each of two PDGFR peptides: 1-13 and 6-14 (Table 3.1). The 1-13 peptide is known to have a high affinity for the SH2 domains (IC_{50} = 7.5 µM), while the 6-14 peptide has a poor affinity (IC_{50} = 97 µM) (21). As expected, the 1-13 peptide was specifically retained by the column and eluted with phenyl phosphate, while none of the 6-14 peptide was in the column eluate (Fig. 3.2a). Only about 30% of the 1-13 peptide loaded was in the eluate (*i.e.*, ~9 nmol), consistent with what was observed with the column capacity titration (see above).

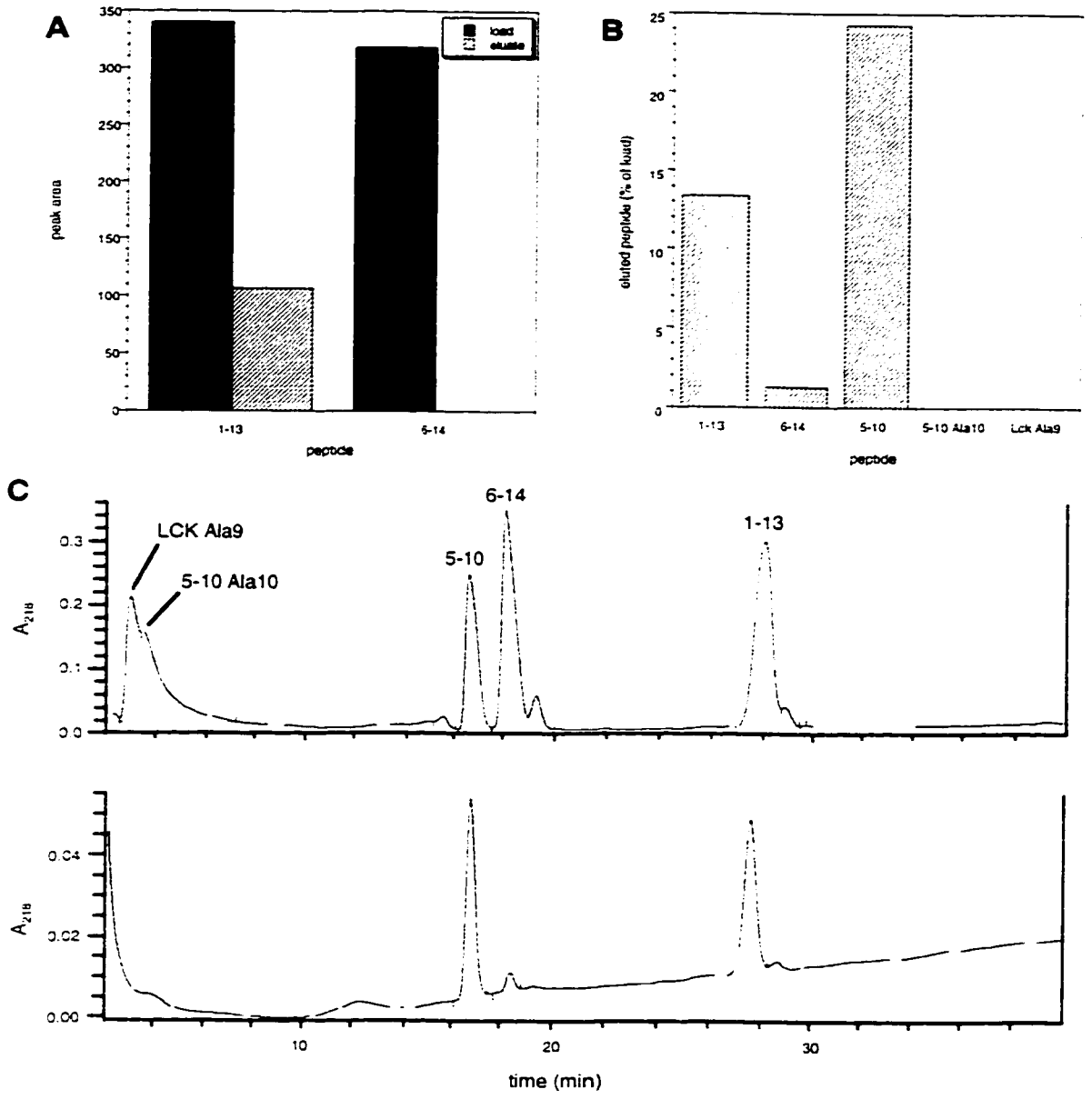


Fig. 3.2. **Controls for specificity of peptide binding to GST-SH2 column.** Columns with 60 nmol GST-SH2 fusion protein immobilized on GT-Sepharose beads were loaded with mixtures of peptides as described under "Experimental Procedures." Eluted peptides were quantitated by HPLC from the peak area and compared to the peak areas of the peptides in the load mixtures. (A) Load mixture consisted of 30 nmol each PDGFR 1-13 and 6-14 peptides. (B) Load mixture consisted of 30 nmol each PDGFR 1-13, 6-14, 5-10, 5-10 Ala10, and Lck Ala9 peptides. (C) HPLC traces of the peptide load (top) and column eluate (bottom) from (B). Peptides were run on HPLC as described under "Experimental Procedures." The traces represent ~67% of the column load and ~40% of the total eluate, and the data in (B) are adjusted for this difference.

In the second test for binding specificity, a mixture of five peptides at 30 nmol each was loaded onto the column as described above. In addition to the 1-13 and 6-14 peptides, the 5-10 peptide (IC_{50} = 4.3 μM), 5-10 Ala10 (IC_{50} = 52 μM), and an unrelated peptide from Lck of unknown affinity were used (21) (Table 3.1). Peptides in the load and eluate were resolved on HPLC and quantitated by peak area at 218 nm. As expected, the 5-10 and 1-13 peptides were bound by the column, with more 5-10 in the eluate than 1-13 reflecting the higher affinity of the 5-10 peptide (Fig. 3.2b). Only background amounts of the other three peptides were present in the eluate. Again, a total of only ~12 nmol of peptide was eluted from the column, consistent with what was observed previously. The results of the affinity selection from this mix of five peptides is readily apparent in the comparison of the HPLC traces of the column load and eluate (Fig. 3.2c).

Affinity Selection of Peptides from Libraries – In order to evaluate the effect of amino acid substitutions at the pY-2 and pY+3 positions, libraries were synthesized based on the PDGFR 2-12 sequence [TSSVL(pY)TAVQP], in which either the pY-2 or pY+3 position (underlined) was replaced with all naturally-occurring L-amino acids except Cys and Trp. Cys and Trp were omitted because of their potential for oxidation and decomposition in sequencing. Amino acid analysis and sequencing indicated that all amino acids were present in the expected amounts (data not shown). Columns were packed with 60 nmol GST-SH2 or GST alone as a control on 225 μl GT-Sepharose beads and loaded with 1 mg of peptide library in 100 μl PBS. The columns were washed, the peptides eluted as described above, and the eluates desalted on HPLC.

Two approaches were used to analyze the column eluates: peptide sequencing and mass spectrometry. Peptide sequencing was relatively straightforward: ~1000 pmol of sample (as estimated by amino acid analysis) was run on a Hewlett-Packard peptide sequencer, and amino acids at the degenerate position were quantitated as the PTH-derivatives compared to standards. As such, actual pmol amounts of amino acids could be determined. Data were expressed as the ratio of each amino acid in the GST-SH2 column eluate to the GST column eluate. This corrects for the peptide background and indicates the enrichment of specific peptides, with a relative selection >1.0 indicating a positive selection. The results of this analysis are shown in Fig. 3.3. For the pY+3 library, the only amino acid that seems to be selected is Val, the pY+3 residue in the PDGFR sequence. For the pY-2 library, a greater range of amino acids appears to be tolerated. The pY-2 PDGFR residue, Val, is selected, as well as Ile, Phe, Thr, and Tyr.

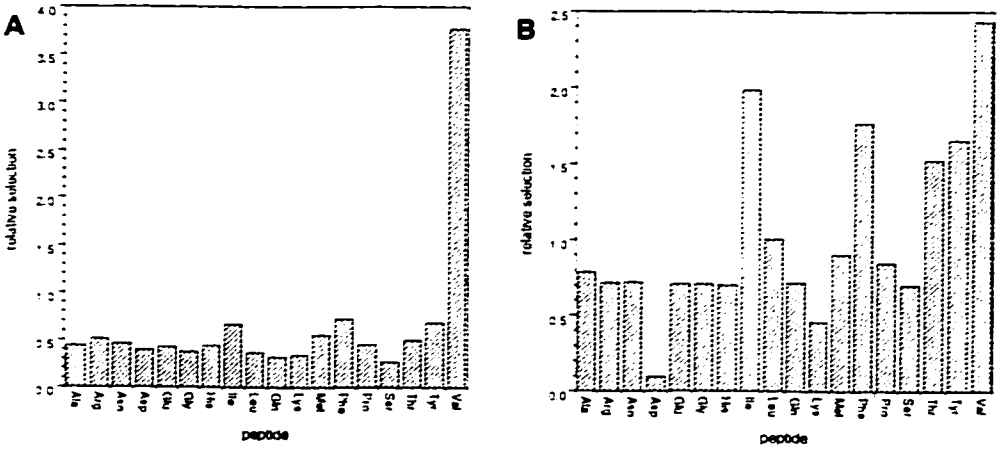


Fig. 3.3. **Affinity selection from peptide libraries, as analyzed by sequencing.** Columns with 60 nmol GST-SH2 fusion protein immobilized on GT-Sepharose beads were loaded with 1 mg of the pY+3 library (A) or pY-2 library (B) as described under "Experimental Procedures." Amino acids at the degenerate position were quantitated by sequencing. Data are expressed as the ratio of the pmol amino acid in the GST-SH2 column eluate over the pmol amino acid in the GST column eluate.

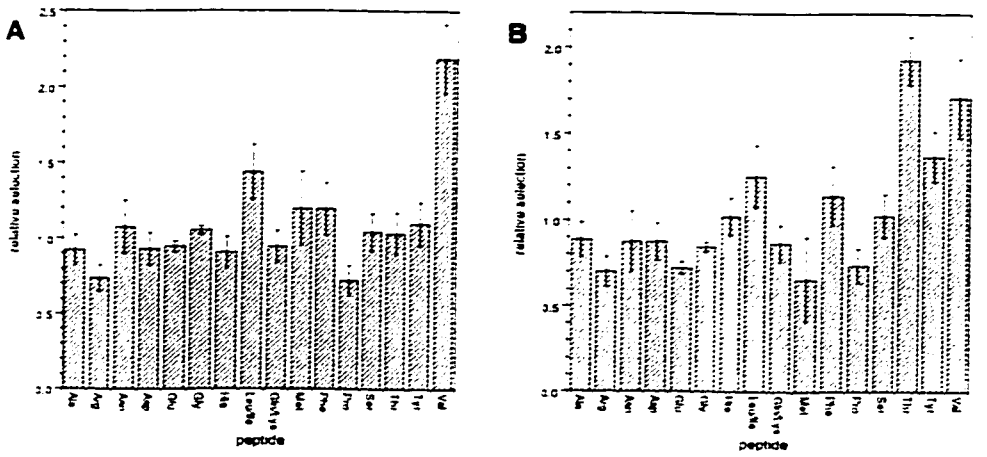


Fig. 3.4. **Affinity selection from peptide libraries, as analyzed by mass spectrometry.** Columns with 60 nmol GST-SH2 fusion protein immobilized on GT-Sepharose beads were loaded with 1 mg of the pY+3 library (A) or pY-2 library (B) as described under "Experimental Procedures." Amino acids at the degenerate position were quantitated by mass spectrometry peak height and normalized as a percentage of the sum of the peak heights of all the peptides. Data are expressed as the ratio of the normalized peak heights in the GST-SH2 column eluate over the normalized peak heights in the GST column eluate.

The other approach for analysis was mass spectrometry. Peptides were separated on a C_{18} capillary column connected on-line to a mass spectrometer. The peak heights of the ions corresponding to the singly-charged mass of each peptide were determined and used as a quantitative measure of the peptides. Without appropriate standards (*i.e.*, individual peptides corresponding to all the peptides in the libraries), it was not possible to convert the arbitrary units of peak height into actual mass or mole units. In some experiments, an internal standard peptide (PDGFR 1-13 or 1-14) was included before injection on the mass spectrometer to normalize the peptide ion peak heights for comparison between runs. However, it was found that this created more scatter in the data, perhaps because of irreproducibility in the addition of internal standard peptide. Ultimately the peak height for each peptide was expressed as a percentage of the sum of

all the peptide peak heights in the sample. It was assumed that the peak height *vs.* peptide concentration response was the same for all peptides since the peptides were very similar, being identical at 11 of 12 positions in the peptide. Thus, the sum of peak heights should be indicative of the total amount of sample injected on the mass spectrometer, and expressing individual peak heights as a percentage of the sum adjusts for differences in total peptide loaded on the mass spectrometer, eliminating the need for an internal standard.

As for the sequencing data, the mass spectrometry results were expressed as the ratio of GST-SH2 data over GST data (Fig. 3.4). The results are generally quite similar to the sequencing data. For the pY+3 library, Val was again very strongly selected. There also appeared to be a weak selection for Leu and/or Ile (these amino acids have the same mass; therefore, peptides containing these residues cannot be distinguished by mass). At the pY-2 position, Val was again positively selected, along with Leu/Ile, Phe, Thr, and Tyr. As compared to the sequencing results, the relative selection for Phe was not as strong while the selection for Thr was stronger. The Leu/Ile selection is midway between the Leu and Ile selections as determined by sequencing, as would be expected since the mass spectrometric results represent the average of these two peptides.

As an alternative to an active site-directed elution with the competitor phenyl phosphate, the use of TFA was explored. Acid elution has been described previously for characterizing peptides bound to MHC molecules (27, 28). TFA should denature the protein and free it from the column; as well, peptide bound both specifically at the SH2 binding site and nonspecifically to other sites on the protein is released. It was found that using 5% TFA led to the precipitation of the eluted GST-SH2 protein and the trace

amounts of blue dextran from the column wash, while leaving the peptides free in solution. Thus, protein did not need to be removed from the column eluates by ultrafiltration, and not using phenyl phosphate meant that samples did not need to be desalted by HPLC before analysis. The results from one preliminary experiment with mass spectrometric analysis suggested a similar profile of peptides was selected as compared to the phenyl phosphate elution (data not shown). The elimination of the HPLC desalting step removed an additional manipulation of the sample, reducing the potential for sample loss. The TFA elution is also more quantitative than phenyl phosphate; however, non-specifically bound peptides are also eluted, and denaturation of the immobilized protein makes re-use of the column impossible.

Determination of Binding Affinity of Individual pY-2 Substituted Peptides – In order to confirm the predictions from the affinity selection procedure, a number of individual peptides substituted at the pY-2 position were synthesized and tested in the competitive binding assay (21). Peptides representing residues selected for and against were chosen. All peptides were based on the minimum binding sequence, VL(pY)TAV. In general, there was a good agreement between the affinity selection results and the relative affinities of the peptides (Table 3.2). Deciding what constitutes high and low affinity from the IC_{50} data is somewhat arbitrary. From our previous work (21), the PDGFR 1-13 and 5-10 peptides were considered to be the references for high affinity binding. Peptides with an IC_{50} more than five-fold higher were considered to have weak affinity, while a three- to five-fold increase indicated moderate affinity. On this basis, as compared to the IC_{50} of the PDGFR 5-10 peptide (Table 3.2), the Ile and Leu peptides are of high affinity; Thr and

Phe, moderate affinity; and Met and Ser, poor affinity. The one clear correlation disagreement is with the Thr substitution: both mass spectrometry and sequencing predicted that this residue is strongly selected, but the peptide has only moderate affinity. The Phe peptide as well has a moderate affinity, in agreement with mass spectrometry; however, sequencing predicted that the peptide should have strong affinity. The Leu peptide was predicted by sequencing to have poor affinity, but in fact it is as good as the parent peptide. Ile and Leu were not differentiated by mass spectrometry as they have the same mass, and therefore it is not possible to tell if the mass spectrometric selection observed is for both peptides or only one.

Table 3.2. **Correlation of affinity selection results with competitive binding assay.**

peptide sequence	relative selection (sequencing)[a]	relative selection (mass spec)[a]	IC_{50}[b] (μM)	correlation[c]
VL(pY)TAV	+	+	3.1 ± 0.6	√
IL(pY)TAV	+	+[d]	3.1 ± 0.4	√
LL(pY)TAV	–	+[d]	4.4 ± 1.0	?
TL(pY)TAV	+	+	12.0 ± 3.4	X
FL(pY)TAV	+	+/–	13.1 ± 2.5	?
ML(pY)TAV	–	–	23.7 ± 10.0	√
SL(pY)TAV	–	–	46.0 ± 12.1	√

[a] Relative selection data is taken from Figs. 3.3 and 3.4: "+", > 1.0; "–", ≤ 1.0.

[b] IC_{50} values were determined in the competitive binding assay as described under "Experimental Procedures." From the IC_{50}'s, the pY-2 Val, Ile, and Leu peptides were considered to have high affinity; Thr and Phe, moderate affinity; and Met and Ser, low affinity (see "Results" for details).

[c] A "√" indicates agreement between the affinity as determined from the IC_{50} with both mass spectrometry and sequencing; a "?" indicates agreement with either mass spectrometry or sequencing; and an "X" indicates no agreement with either technique.

[d] These two peptides cannot be distinguished by mass spectrometry; thus, the positive selection observed may apply to one or both of the peptides.

DISCUSSION

The use of peptide libraries in an affinity selection approach is a powerful means to define binding specificities of SH2 domains. The biggest challenge in this approach is the identification of the high affinity peptides bound to the SH2 domains. Sequencing, which is one method that has been used (25), indicates the relative abundance of amino acids at each position in the peptide. As such, all library peptides must be derived from the same sequence, but multiple positions in the peptide can be degenerate. However, sequencing has a number of limitations. First, it is not strictly quantitative: the recoveries vary for the different amino acids because of the PTH derivitization and hydrolysis of the amino acids. Second, a large amount of sample is required (~1000 pmol total peptide was required in this analysis) in order to be able to quantitate the amino acids at the degenerate positions with certainty. Finally, sequencing is expensive and time-consuming, limiting the number of samples that can be analyzed.

As an alternative to sequencing, we chose to explore mass spectrometry as a means of quantitating the peptides. Mass spectrometry requires less sample than sequencing because of its high sensitivity: samples with less than 100 pmol total peptide were analyzed in this study. As well, it is much faster and cheaper than sequencing: for example, the samples in this study were analyzed in about a half hour. Furthermore, since the peptides are analyzed directly without hydrolysis or modification, the analysis should be more quantitative. As well, the library peptides do not necessarily need to be derived from the same sequence as long as all the peptides have unique masses to permit their identification by mass spectrometry. However, this requirement for unique masses limits

the potential complexities of the libraries. Even in the relatively simple libraries with one degenerate position used in this study, Leu/Ile and Gln/Lys could not be resolved since these pairs of residues have the same masses.

To evaluate mass spectrometric detection as an alternative to sequencing, we used both techniques in our study with the SH2 domains of SHP-2. In general, there was a good agreement between the results obtained from sequencing (Fig. 3.3) and from mass spectrometry (Fig. 3.4). For the pY+3 library, both techniques showed a strong selection for Val. Mass spectrometry also suggested a weaker selection for Leu/Ile which was not apparent by sequencing. For the pY-2 library, the same residues were selected by both techniques. However, the relative selections varied; for example, Phe had a strong relative selection by sequencing but a weak relative selection by mass spectrometry. Ile was strongly selected by sequencing, while the relative selection for Leu was much less. Mass spectrometry cannot differentiate Leu and Ile, and thus the relative selection represents an average of the two residues. Therefore, it is conceivable that this difference is real and masked in the mass spectrometric detection. The quality of data from mass spectrometry is likely better than that from sequencing, as the mass spectrometric data represent an average of four (pY+3 library) or six (pY-2 library) runs, while only one sequencing run was performed for each library.

The ultimate test for the affinity selection process is to determine if the selection predictions do in fact correlate with the relative affinities of the individual peptides. To this end, six peptides substituted at the pY-2 position were synthesized and their IC_{50}'s determined (Table 3.2). In general, there was a very good correlation between relative

selection and relative affinity, with a few exceptions. The IC_{50}'s for the Leu and Phe peptides were consistent with mass spectrometry but did not agree with the sequencing results, which may simply reflect the quality of the sequencing data as compared to mass spectrometry (see above). It is also somewhat difficult to imagine that there would be such a strong selection for Ile *vs.* Leu as suggested by sequencing since these two amino acids are very similar in structure and the binding pocket is clearly flexible enough to tolerate other substitutions. The moderate affinity of the Thr peptide as indicated by its IC_{50} was not consistent with either the sequencing or mass spectrometry results and is difficult to rationalize. However, it should be noted that that the peptide libraries were based on the PDGFR 2-12 sequence, while the peptides used for IC_{50} determinations were based on the minimum binding sequence, PDGFR 5-10. While it is unlikely that amino acids outside of this minimum sequence would have any effect on the affinity of the peptide based on our earlier data (21), for completeness a PDGFR 2-12 peptide with a pY-2 Thr substitution should be tested.

The fact that there are sequence preferences at the pY-2 position supports the hypothesis that this position is important for peptide interactions with the SH2 domains of SHP-2. This result goes against the general consensus that in SH2 domain–peptide interactions the specificity is determined by residues C-terminal to the target pY and that N-terminal residues are of little or no importance (1-4). Another published report using a binding assay to assess relative peptide affinities for the N-terminal SH2 domain of SHP-2 (20) is consistent with our results. In that study, PDGFR 3-14 bound tightly, while PDGFR 6-11, in which the pY-2 residue is eliminated, bound poorly. Similarly, a peptide

derived from the insulin receptor substrate-1 (IRS-1), LN(pY)IDLDLV, bound with high affinity, while a truncated version without the pY-2 residue, N(pY)IDLD, had poor affinity. The authors suggested that the loss of the pY+5 residues accounts for the drop in affinity; however, in light of our studies, it is more likely due to the removal of the pY-2 residues. Other reports have also suggested a role for residues N-terminal to pY for binding by the SH2 domains of Nck (18) and Src (29), and recently, a study of the SH2 domains of the PTP SHP-1 revealed a requirement for a hydrophobic amino acid at the pY-2 position for high-affinity binding (30).

The importance of the pY-2 position may be a result of a Gly residue at the αA2 position of the SH2 domains of SHP-2 and SHP-1. In almost all other SH2 domains, the αA2 residue is normally Arg and occasionally Lys. The crystal structure of the N-terminal SH2 domain of SHP-2 bound to the PDGFR 4-14 peptide (9) reveals a role for the pY-2 residue in covering the pY phenyl ring, filling a gap created by the αA2 Arg→Gly substitution. Replacing the Gly with Arg may eliminate the pY-2 peptide interaction and make the SH2 domain specificity of SHP-2 more like that of other SH2 domains. However, because of the Arg→Gly substitution, the pY phosphate of the peptide is rotated 180° such that, although the plane of the phenyl ring is the same, the phosphate is pointing in the opposite direction as compared to the Src and Lck SH2 domain–peptide complexes (5, 7). Therefore, a Gly→Arg substitution in the SH2 domains of SHP-2 may have other secondary effects on peptide binding, affecting pY interactions.

A number of details regarding the affinity selection technique and the analysis need to be explored and optimized to improve the analysis. For example, it is likely that bound

peptide is dissociating from the column during the wash steps. This may result in weakly-bound peptides being preferentially washed off, making the relative selection of the higher-affinity peptides appear stronger. However, if bound peptides have significantly different off rates, then washing will affect their recoveries unequally. Furthermore, incomplete elution of peptides from the column may be a problem. Using TFA as a denaturing eluent may improve recovery, but this needs to be further optimized. The HPLC desalting step is an additional manipulation that may contribute to the loss of peptide. Again, from preliminary experiments a TFA elution may obviate the need for HPLC desalting and should be explored further. Radiolabeled PDGFR peptide would be useful as a tracer for optimizing the affinity column binding and elution.

Besides addressing problems with the affinity column, there are a number of mass spectrometric analysis issues that need to be addressed as well. For example, the peak heights of the peptide ions were used as a relative measure of peptide quantity, and the individual peak heights were normalized to the sum of all the peak heights to correct for differences in peptide load. However, this assumes a linear peak height–concentration response for all peptides on the mass spectrometer. To establish whether or not this is the case, peptide standards need to be analyzed in a dilution series to determine the concentration *vs.* peak height profile. In theory the best way to normalize the data would be to use an internal peptide standard added to the sample immediately before mass spectrometric analysis. A preliminary experiment with an internal standard did not give reproducible results (data not shown); perhaps the addition of the standard to samples was not consistent, or the standard did not generate a reproducible signal by mass

spectrometry. The use of an internal standard should be explored further and optimized to improve the data analysis. It is also possible that liquid secondary ion mass spectrometry employed here may not be optimal for peptide quantitation, and other mass spectrometric methods like electrospray ionization could be explored.

The affinity selection approach described in this study with SH2 domains should in principle be applicable to any pY-binding domain, or indeed any protein or peptide binding domain. It will, however, be more complicated to extend this approach to enzymes. For example, protein tyrosine phosphatases bind pY peptides but also hydrolyze the phosphate, releasing the peptide. In such a case, either a mutant enzyme that binds but does not hydrolyze the peptide, or peptides containing a non-hydrolyzable pY analogue, would have to be used. The application of mass spectrometry as described here has the potential to facilitate greatly the analysis of peptides selected from degenerate libraries because of the sensitivity and speed of the analysis. However, more experiments are needed to ensure the quantitative nature of the mass spectrometric results. In general, the results obtained in this study, while not definitive and optimized, do suggest the validity of a peptide library affinity selection approach coupled with mass spectrometric analysis and suggest that it should be possible to apply this method to other systems successfully.

ACKNOWLEDGMENTS

We thank C. Li for performing the mass spectrometric analyses and C. Ramachandran for much patient assistance with the HPLC.

REFERENCES

1. Cohen, G. B., Ren, R., and Baltimore, D. (1995) *Cell* **80,** 237-248

2. Pawson, T. (1995) *Nature* **373,** 573-580

3. Schaffhausen, B. (1995) *Biochim. Biophys. Acta* **1242,** 61-75

4. Songyang, Z., and Cantley, L. C. (1995) *Trends Biochem. Sci.* **20,** 470-475

5. Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. D., Silverman, L., and Kuriyan, J. (1992) *Nature* **358,** 646-653

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

7. Eck, M. J., Shoelson, S. E., and Harrison, S. C. (1993) *Nature* **362,** 87-91

8. Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., and Forman-Kay, J. D. (1994) *Cell* **77,** 461-472

9. Lee, C.-H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E., and Kuriyan, J. (1994) *Structure* **2,** 423-438

10. Mikol, V., Baumann, G., Keller, T. H., Manning, U., and Zurini, M. G. M. (1995) *J. Mol. Biol.* **246,** 344-355

11. Nolte, R. T., Eck, M. J., Schlessinger, J., Shoelson, S. E., and Harrison, S. C. (1996) *Nature Struct. Biol.* **3,** 364-374

12. Tong, L., Warren, T. C., King, J., Betageri, R., Rose, J., and Jakes, S. (1996) *J. Mol. Biol.* **256,** 601-610

13. Metzler, W. J., Leiting, B., Pryor, K., Mueller, L., and Farmer, B. T. (1996) *Biochemistry* **35,** 6201-6211

14 Breeze, A. L., Kara, B. V., Barratt, D. G., Anderson, M., Smith, J. C., Luke, R. W., Best, J. R., and Cartlidge, S. A. (1996) *EMBO J.* **15,** 3579-3589

15. Thorton, K. H., Mueller, W. T., McConnell, P., Zhu, G., Saltiel, A. R.,and Thanabal, V. (1996) *Biochemistry* **35,** 11852-11864

16. Escobedo, J. A., Kaplan, D. R., Kavanaugh, W. M., Turck, C. W., and Williams, L. T. (1991) *Mol. Cell. Biol.* **11,** 1125-1132

17. Fantl, W. J., Escobedo, J. A., Martin, G. A., Turck, C. W., del Rosario, M., McCormick, F., and Williams, L. T. (1992) *Cell* **69,** 413-423

18. Nishimura, R., Li, W., Kashishian, A., Mondino, A., Zhou, M., Cooper, J., and Schlessinger, J. (1993) *Mol. Cell. Biol.* **13,** 6889-6896

19. Piccione, E., Case, R. D., Domchek, S. M., Hu, P., Chaudhuri, M., Backer, J. M., Schlessinger, J., and Shoelson, S. E. (1993) *Biochemistry* **32,** 3197-3202

20. Case, R. D., Piccione, E., Wolf, G., Benett, A. M., Lechleider, R. J., Neel, B. G., and Shoelson, S. E. (1994) *J. Biol. Chem.* **269,** 10467-10474

21. Huyer, G., Li, Z. M., Adam, M., Huckle, W. R., and Ramachandran, C. (1995) *Biochemistry* **34,** 1040-1049

22. Felder, S., Zhou, M., Hu, P., Urena, J., Ullrich, A., Chaudhuri, M., White, M., Shoelson, S. E., and Schlessinger, J. (1993) *Mol. Cell. Biol.* **13,** 1449-1455

23. Panayotou, G., Gish, G., End, P., Truong, O., Gout, I., Dhand, R., Fry, M. J., Hiles, I., Pawson, T., and Waterfield, M. D. (1993) *Mol. Cell. Biol.* **13,** 3567-3576

24. Payne, G., Shoelson, S. E., Gish, G. D., Pawson, T., and Walsh, C. T. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 4902-4906

25. 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., and Cantley, L. C. (1993) *Cell* **72,** 767-778

26. Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R. A., and Cantley, L. C. (1994) *Mol. Cell. Biol.* **14,** 2777-2785

27. Hunt, D. F., Henderson, R. A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilir, N., Cox, A. L., Appella, E., and Engelhard, V. H. (1992) *Science* **255,** 1261-1263

28. Slingluff, C. L., Jr., Cox, A. L., Henderson, R. A., Hunt, D. F., and Engelhard, V. H. (1993) *J. Immunol.* **150,** 2955-2963

29. Bibbins, K. B., Boeuf, H., and Varmus, H. E. (1993) *Mol. Cell. Biol.* **13,** 7278-7287

30. Burshtyn, D. N., Yang, W., Yi, T., and Long, E. O. (1997) *J. Biol. Chem.* **272,** 13066-13072

31. Sugimoto, S., Wandless, T. J., Shoelson, S. E., Neel, B. G., and Walsh, C. T. (1994) *J. Biol. Chem.* **269,** 13614-13622

CHAPTER 4

Specificity of the N-terminal SH2 Domain of SHP-2 is Modified by a Single Point Mutation

SUMMARY

SH2 domains are small protein domains of ~100 amino acids that bind to phosphotyrosine (pY) in the context of a specific sequence surrounding the target pY. In general, the residues C-terminal to the pY of the binding target are considered most important for defining the binding specificity, and in particular the pY+1 and pY+3 residues (*i.e.*, the first and third amino acids C-terminal to the pY). However, our previous studies with the SH2 domains of the protein tyrosine phosphatase SHP-2 [Huyer, G., Li, Z. M., Adam, M., Huckle, W. R., and Ramachandran, C. (1995) *Biochemistry* **34,** 1040-1049] indicated important interactions with the pY-2 residue as well. In the SH2 domains of SHP-2, the highly-conserved αA2 Arg is replaced by Gly. A comparison of the published crystal structures of the Src SH2 domain and the N-terminal SH2 domain of SHP-2 complexed with high-affinity peptides suggested that the αA2 Gly of SHP-2 creates a gap which is filled by the side chain of the pY-2 residue of the bound peptide. It was predicted that replacing this Gly with Arg would alter or eliminate the involvement of the pY-2 residue in binding. The αA2 Gly→Arg mutant was constructed, and indeed this mutant no longer required residues N-terminal to the target pY for high affinity binding, making its specificity more like that of other SH2 domains. The αA2 Gly is clearly involved in

directing the unusual requirement for the pY-2 residue in the binding sequence of this SH2 domain, which has important implications for its *in vivo* targeting and specificity.

INTRODUCTION

SH2 domains (for Src homology 2)[1] are discrete protein domains of ~100 amino acids that bind to phosphotyrosine (pY) in the context of a specific sequence surrounding the pY (1-4). SH2 domains have been identified in as many as 100 proteins, including two mammalian protein tyrosine phosphatases (PTPs), SHP-1 and SHP-2. These PTPs share approximately 58% identity and have two tandem SH2 domains in the N-terminal portion and a single C-terminal catalytic domain. SHP-1 and SHP-2 play important roles in regulating many signaling pathways, including hematopoietic cell signal transduction and growth factor signaling (5, 6).

The involvement of SH2 domain-containing proteins in a particular cell signaling pathway is directed by their SH2 domains; thus, knowledge of SH2 domain binding specificity is important for predicting and identifying their *in vivo* targets. For most SH2 domains, interactions with residues C-terminal to the target pY, and in particular the pY+1 and pY+3 residues (*i.e.*, the first and third positions C-terminal to pY), are responsible for defining specificity. Using affinity selection of peptides from degenerate libraries, Songyang *et al.* (7) defined the binding sequence specificities of a number of SH2 domains which fell into two main classes: (i) pY–hydrophilic–hydrophilic–hydrophobic; and (ii) pY–hydrophobic–X–hydrophobic, where "X" indicates any amino acid. Site-directed

[1] Abbreviations: SH2, Src homology 2; pY, phosphotyrosine; PDGFR, platelet-derived growth factor receptor ß; GST, glutathione *S*-transferase; GT, glutathione; PBS, phosphate buffered saline; DTT, dithiothreitol; TBS, Tris buffered saline; HBS, HEPES buffered saline; PTP, protein tyrosine phosphatase.

mutagenesis has demonstrated the involvement of particular residues of SH2 domains in defining binding specificity (8-10). In some cases a single amino acid change in the SH2 domain lead to dramatic changes in binding sequence specificity.

We have previously shown that the SH2 domains of SHP-2 have an unusual specificity in that residues N-terminal to the target pY, in particular the pY-2 residue, influence binding (11). Recent studies with SHP-1 also revealed a similar involvement of N-terminal residues (12). The crystal structure of the N-terminal SH2 domain of SHP-2 bound to a peptide corresponding to its *in vivo* target (13) suggests a structural reason for this pY-2 interaction. Almost all SH2 domains have a highly-conserved Arg (or occasionally Lys) residue at the αA2 position; however, in the SH2 domains of SHP-2 (and SHP-1), this position is occupied by Gly. As a result, a gap is created in the SH2 domain that appears to be filled by the pY-2 side chain of the target peptide and is a likely reason for the requirement for N-terminal residues. Thus, changing this Gly to Arg was predicted to eliminate this pY-2 interaction. This change was introduced, along with five other αA2 substitutions which molecular modeling suggested may have a similar effect. The Gly→Arg substitution resulted in an SH2 domain with peptide binding affinity comparable to wild-type. As predicted the requirement for residues N-terminal to the target pY was eliminated in the αA2 Arg mutant, resulting in a specificity more typical of other SH2 domains. The αA2 Gly is clearly involved in the unusual specificity of the N-terminal SH2 domain of SHP-2 and reveals an additional mode of SH2 domain binding that involves residues N-terminal to the target pY.

EXPERIMENTAL PROCEDURES

Materials – Phosphotyrosine peptides were obtained from California Peptide Research (Napa, CA), and PCR primers were synthesized by Research Genetics. Anti-phosphotyrosine antibody was from Upstate Biotechnology Inc., and GT-Sepharose was from Pharmacia. All other chemicals were of reagent grade from Sigma.

Cloning and Mutagenesis of N-terminal SH2 Domain of SHP-2 – The N-terminal SH2 domain of SHP-2 was cloned by PCR from a full-length clone of SHP-2 in the vector pBluescript (14). The PCR primer used at the 5' end was GGGGGATCCATGACATCGCGGAGATGGTTT, and at the 3' end, AGGAATTCTATGCACAGTTCAGAGGATATTT. The primers were designed to amplify residues 1-105 of SHP-2 with the addition of a *Bam*HI site at the 5' end and a stop codon and *Eco*RI site at the 3' end. The amplification conditions were 30 s at 95°C, 30 s at 52°C, and 30 s at 72°C, for 20 cycles. The PCR products were purified using a PCR purification kit (QIAgen), digested with *Bam*HI and *Eco*RI, gel purified, and ligated into pBluescript SK+. Sequencing ensured that no errors had been introduced by PCR. The SH2 domain insert was subsequently cut out from pBluescript by digestion with *Bam*HI and *Eco*RI, gel purified, and ligated in-frame into the GST fusion vector pGEX-2T (Pharmacia) that had been digested with the same enzymes.

The αA2 substitutions were created using a PCR mutagenesis technique (15). To introduce the change at the αA2 position (the 13th codon), 5' primers of 58-60 bases were used that included the 5' start and *Bam*HI site plus 10-12 bases after the mutated codon; the same 3' primer from the original subcloning of the SH2 domain (see above) was used.

The SH2 domain cloned into pGEX-2T was used as the template DNA. The PCR conditions were 40 s at 95°C, 30 s at 55°C, and 30 s at 72°C, for 20 cycles, and the PCR products were purified, digested, and ligated into pGEX-2T as described above. Sequencing verified that the correct mutations had been made.

Expression and Purification of GST-SH2 Constructs – Escherichia coli DH5α cells containing the GST-SH2 constructs were grown at 37°C in 700 ml LB with 100 μg/ml ampicillin to $A_{600} = 0.6 - 0.8$. Cultures were chilled on ice to ~27°C and expression was induced by the addition of 50 μM isopropylthio-ß-D-galactoside (IPTG) with incubation overnight at 27°C. All subsequent steps were performed at 4°C. Cells were harvested by centrifugation at 4000 x *g* for 10 min, resuspended in 15 ml lysis buffer [PBS containing 5 mM DTT, 1% Triton X-100, and Complete protease inhibitor cocktail (Boehringer Mannheim)], and lysed by sonication. Lysates were cleared by centrifugation at 23,000 x *g* for 15 min, and the fusion proteins were purified by incubating with 2 ml GT-Sepharose beads for 1 h with end-over-end mixing. The beads were washed extensively with lysis buffer and then with TBS (50 mM Tris pH 8.0, 150 mM NaCl) containing 5 mM DTT. Beads were stored at -20°C after the addition of an equal volume of 100% glycerol to a 50% suspension of the beads in TBS/DTT. The fusion proteins were >95% pure as estimated by SDS–PAGE.

To elute the GST-SH2 fusion proteins from the beads, 200 μl of beads were washed with TBS containing 5 mM DTT and 1 mM EDTA. Then 500 μl of the same buffer containing 10 mM GT were added and the mixture incubated on ice for 15 min with occasional mixing. The beads were pelleted by centrifugation and the supernatant was

removed. Typically 2.5 – 3 mg of fusion protein were recovered from 200 µl of beads. The fusion proteins had the expected mass as determined by electrospray ionization mass spectrometry.

BIAcore Assay for SH2 Domain Binding – All experiments were carried out in HBS buffer (10 mM HEPES pH 7.4, 0.15 M NaCl, 3 mM EDTA, and 0.005% Surfactant P20). The PDGFR 1009 peptide [DTSSVL(pY)TAVQPN] was biotinylated as described (11) and immobilized on a streptavidin-coated sensor chip (Pharmacia) by injecting 10 µl of 8.3 ng/ml peptide (flow rate of 5 µl/min). For all binding experiments, a flow rate of 10 µl/min was used, and samples were also injected over a flow cell with no immobilized peptide to provide a blank sensogram that was subtracted from the binding sensogram. Control experiments involving binding of anti-pY monoclonal antibody (injected at 10 µg/ml) reproducibly resulted in a 400-450 change in resonance units. The sensor surface was regenerated at the end of each experiment with 100 mM HCl which did not disrupt the high-affinity biotin–streptavidin interaction but did remove the bound proteins.

RESULTS

Modeling of SH2 Domain–Peptide Interactions – A unique feature of the SH2 domains of SHP-2 and SHP-1 is that Gly occupies the αA2 position instead of a normally highly-conserved Arg (Fig. 4.1). In the crystal structures of the Src (16, 17) and Lck (18) SH2 domains complexed with peptides, an amino nitrogen of the αA2 Arg side chain makes an amino–aromatic interaction with the pY ring; in addition, the αA2 Arg forms several hydrogen bonds with the phosphate of the pY and with the pY-1 residue. In the structure

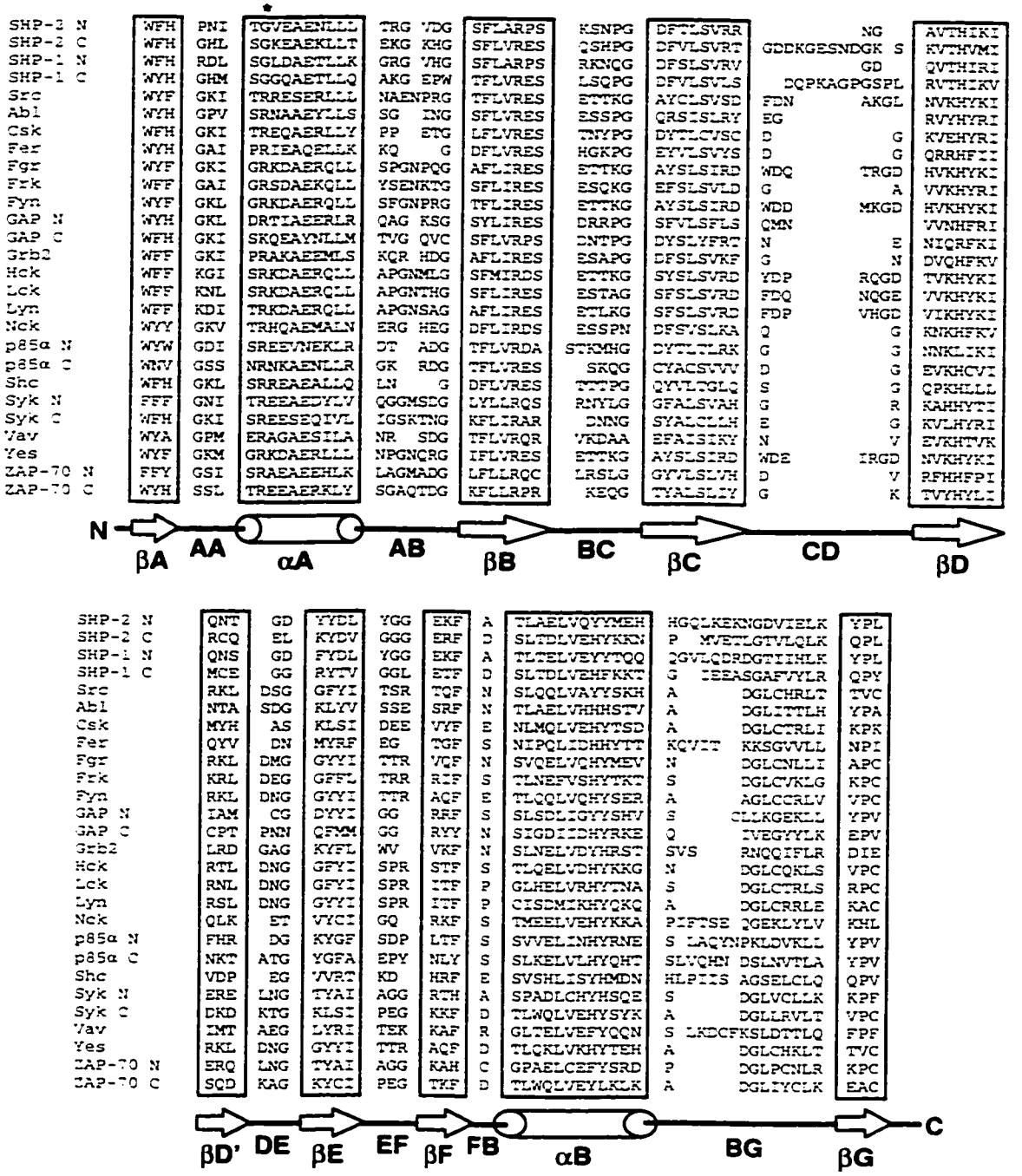


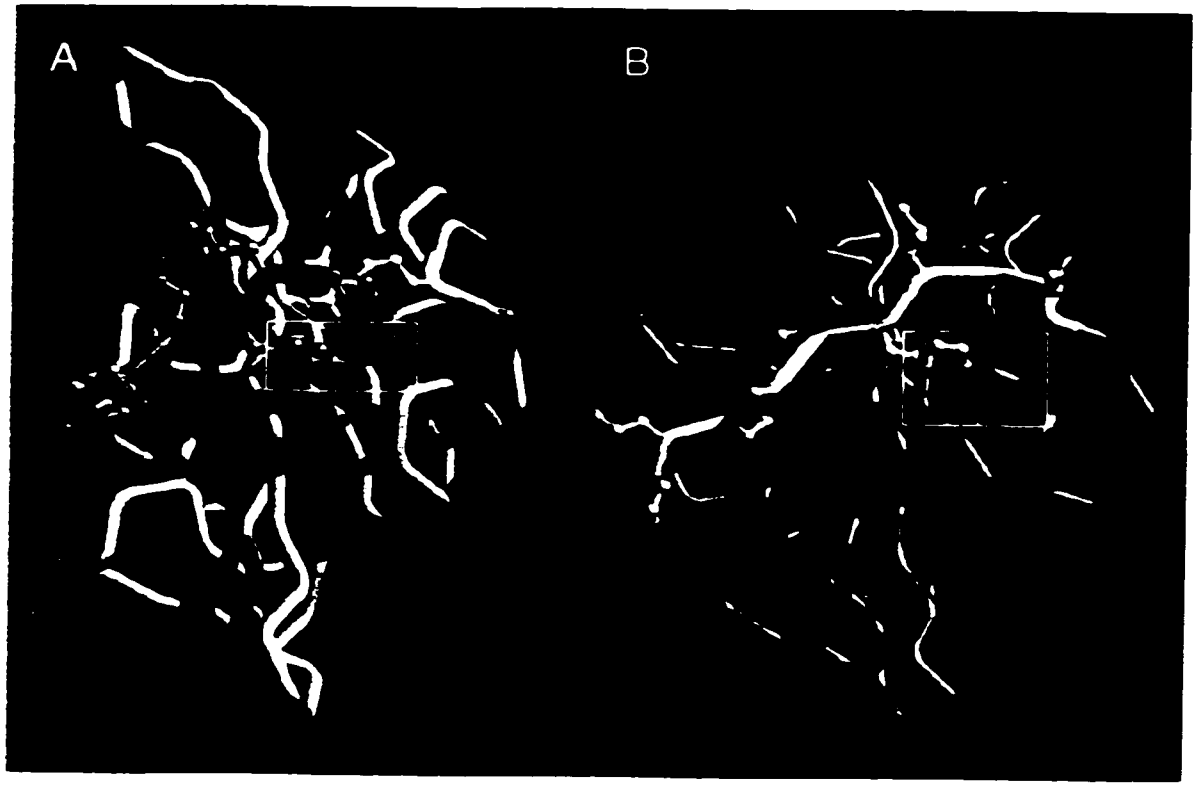
Fig. 4.1. **Alignment of human SH2 domain sequences.** A subset of human SH2 domain-containing proteins was chosen from the Prosite SH2 domain profile and their SH2 domains aligned using GCG PileUp. The boxes show the boundaries of secondary structure elements as defined by Eck *et al.* (18) and are indicated schematically at the bottom along with the notation for these elements. The αA2 residue is marked by an asterisk. For proteins with two SH2 domains (*e.g.*, SHP-2), the SH2 domains are indicated as N or C, for N-terminal or C-terminal.

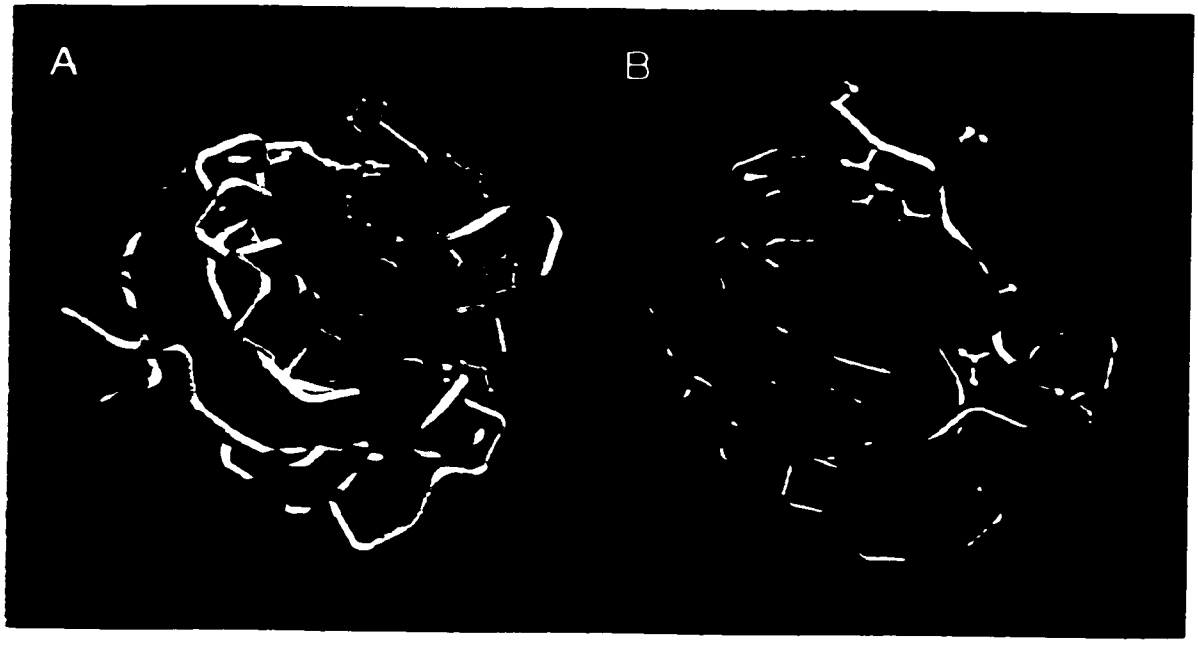
of the N-terminal SH2 domain of SHP-2 complexed with peptide (13), the replacement of this Arg by Gly is not compensated by any structural rearrangement of the SH2 domain relative to Src or Lck. However, while the phenyl ring of the pY residue is bound in a very similar conformation in SHP-2 and Src, the phosphate is rotated ~180° relative to that in Src (Fig. 4.2). This rotation results in additional interactions with the βB5 Arg residue, and overall the same number of hydrogen bonds are formed as compared to those in the Src SH2 domain (13).

An additional consequence of the αA2 Gly of SHP-2 is that a gap is created in the SH2 domain that appears in the crystal structure to be filled by the side chain of the pY-2 Val of the PDGFR peptide, helping to cover the phenyl ring of the pY (Fig. 4.3). We rationalized that replacing this Gly with Arg would eliminate the involvement of the pY-2 residue in the SH2 binding specificity. In addition to the Arg mutant, five other substitutions were made: Lys, another basic residue that is occasionally observed at this

Fig. 4.2 (*opposite, top*). **Comparison of the pY-binding sites of Src and SHP-2.** (A) Src SH2 domain (orange) with bound PQ(pY)EEI peptide (green); (B) N-terminal SH2 domain of SHP-2 (cyan) with bound SVL(pY)TAVQPNE peptide (yellow). In both SH2 domain structures, the αA2 residue is shown in white. The orientations were matched by overlaying the structures, with an RMS deviation of 0.72 Å between 40 residues in conserved structural elements. The phosphotyrosine of the bound peptide is within the white box. While the orientation of the phenyl rings is similar, the phosphate in the SHP-2 structure is rotated ~180° relative to that in the Src structure. The images were produced using the MidasPlus software system from the Computer Graphics Laboratory, UCSF (40). Coordinates were obtained from the Brookhaven Protein Data Bank (Src, ID # 1SPS; SHP-2, ID # 1AYA).

Fig. 4.3 (*opposite, bottom*). **Comparison of the αA2 residue–pY peptide interactions of Src and SHP-2.** (A) Src SH2 domain (orange) with bound PQ(pY)EEI peptide (green); (B) N-terminal SH2 domain of SHP-2 (cyan) with bound SVL(pY)TAVQPNE peptide (yellow). In both SH2 domain structures, the αA2 residue is shown in white. Note the gap created by the replacement of the αA2 Arg with Gly in SHP-2, which is filled by the Val side chain of the pY-2 residue.


A
B


A
B

position in other SH2 domains; His, the only other basic residue, for completeness; Gln, as it appeared by modeling to be able to reach far enough to interact favorably with the pY; Met, to occupy the space created by Gly at αA2; and Ser, because modeling suggested that it would eliminate the need for a bound water molecule that hydrogen bonds with the peptide.

Binding Studies with Mutant SH2 Domains – The ability of the SH2 domains to bind to the PDGFR 1009 peptide was assessed by surface plasmon resonance, using a BIAcore system. The SH2 domains were eluted from the Sepharose beads by competition with glutathione, diluted to 500 nM in HBS buffer, and injected over immobilized PDGFR 1009 peptide on a sensor chip. As apparent from the overlay sensograms (Fig. 4.4), the wild-type SH2 domain caused the largest change in resonance units upon binding, followed by the αA2 Arg mutant. The other five mutants displayed no detectable binding to the peptide under similar conditions. From the sensograms, the kinetic binding constants were calculated for the wild-type and αA2 Arg SH2 domains and are summarized in Table 4.1. The dissociation constants for the two SH2 domains are essentially identical, with an affinity of ~40 nM.

Binding Sequence Specificities of Wild-type and αA2 Arg SH2 Domains – We previously demonstrated (11) that the minimum sequence surrounding Tyr_{1009} of the PDGFR for high-affinity binding to the SH2 domains of SHP-2 was VL(pY)TAV. In order to determine if the αA2 Arg mutation changed the binding specificity of the SH2 domain, a number of peptides were tested for their ability to compete in the BIAcore assay. In referring to the peptides, the 13mer PDGFR 1009 peptide immobilized on the sensor chip

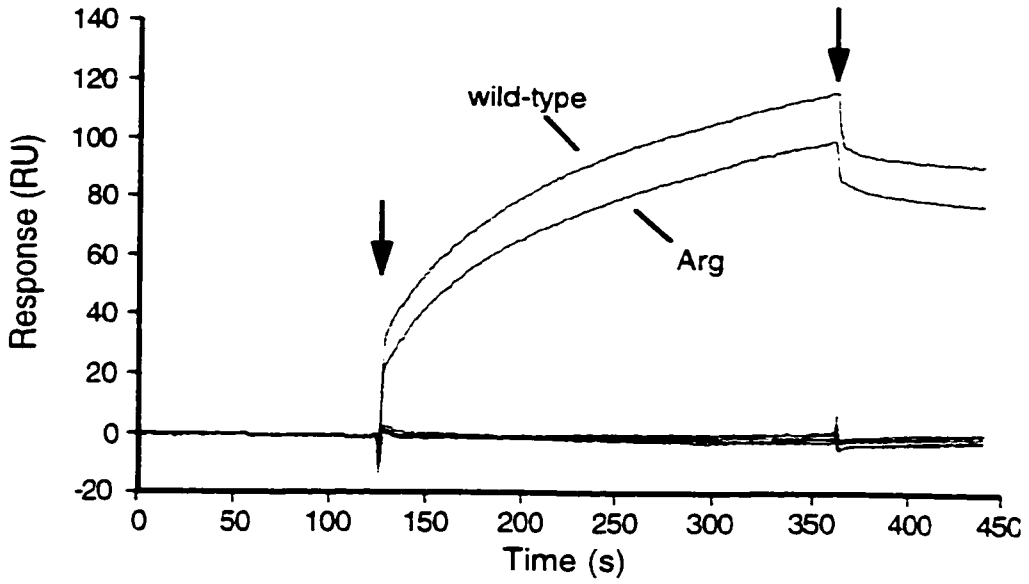


Figure 4.4. **Overlay sensograms for the binding of wild-type and αA2 substituted SH2 domains to immobilized PDGFR 1009 peptide.** The various SH2 domains (indicated by the αA2 residue, or "w.t." for wild-type) were diluted to 500 nM in HBS buffer and injected over the sensor chip surface with immobilized PDGFR 1009 peptide, and the background (*i.e.*, injection over the sensor surface without immobilized peptide) was subtracted. First arrow, injection of SH2 domain; second arrow, wash with HBS buffer.

Table 4.1. **Association and dissociation rate constants for the interaction of wild-type and αA2 Arg SH2 domains with PDGFR 1009 peptide.**

SH2 domain	$k_a \pm$ S.E. (10^{-4} $M^{-1}s^{-1}$)[a]	$k_d \pm$ S.E. (10^{-4} s^{-1})[a]	$K_d \pm$ S.E. (nM)[b]
wild-type	2.10 ± 0.42	8.87 ± 1.17	42.9 ± 5.7
αA2 Arg	2.19 ± 0.77	7.66 ± 0.29	38.3 ± 12.9

[a] Kinetic parameters for binding were calculated from the sensograms in Fig. 4.4 and at three other SH2 domain concentrations, after blank subtraction using the BIAevaluation software package. The values represent the average and standard deviation of the four determinations.

[b] The K_d was calculated for each of the four SH2 domain concentrations from the ratio of the k_d over the k_a, and the values were averaged.

[DTSSVL(pY)TAVQPN] is designated as "1-13" and the numbering of all truncated peptides is based on this numbering. First, the concentration of free 1-13 peptide that reduced the binding by 50% (*i.e.*, the IC_{50}) was determined, which was ~1.5 μM for both the wild-type and αA2 Arg SH2 domains (Table 4.2). The various peptide competitors were tested at this concentration: if they competed as well as the 1-13 peptide then binding would similarly be reduced by 50%; if they were poorer competitors, little or no reduction in binding would be observed.

The specificity for the pY-2 residue was examined by testing a series of peptides based on the minimum high affinity sequence [VL(pY)TAV, "5-10"] with substitutions at the pY-2 position. The pY-2 residue was an important determinant of binding to the wild-type SH2 domain. Peptides with Val, Ile, and Leu at this position competed strongly; Met, Phe, and Thr were reasonable competitors; and Ser and Ala were poor competitors (Fig. 4.5a). For the αA2 Arg SH2 domain, all peptides competed equally well (Fig. 4.5b), indicating that the side chain at this position was unimportant. The IC_{50}'s of a subset of the peptides were determined (Table 4.2), further emphasizing the difference between the wild-type and αA2 Arg SH2 domains. For the wild-type SH2 domain the IC_{50}'s varied over a 35-fold range, while for the αA2 Arg SH2 domain less than a two-fold range of IC_{50}'s was observed. The specificity for the pY+3 residue was unaffected by the αA2 Arg substitution: replacing the pY+3 residue with Ala increased the IC_{50} ten-fold relative to Val for both the wild-type and αA2 Arg SH2 domains (Table 4.2).

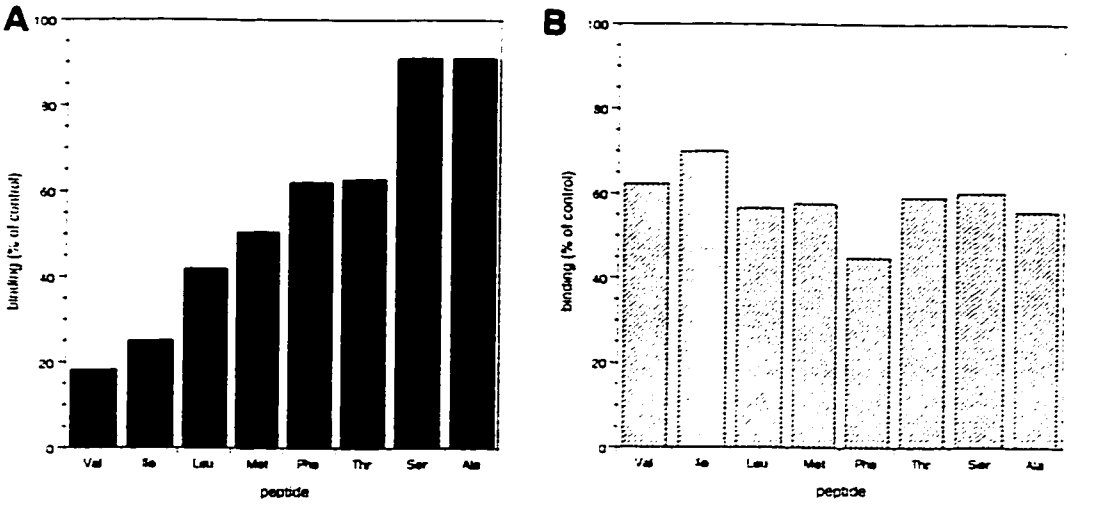


Fig. 4.5. **Effect of pY-2 residue in peptide binding competitions.** Wild-type (A) or αA2 Arg mutant (B) SH2 domains (500 nM each) were incubated with the indicated peptides at 1.5 μM in HBS buffer. The increase in resonance units at steady state was determined and the background (*i.e.*, injection over the sensor surface without immobilized peptide) was subtracted. Binding is expressed as a percent of the SH2 domain binding in the absence of competitor to the 1-13 peptide immobilized on the sensor chip. All peptides have the sequence XL(pY)TAV, with the residue at the pY-2 position ("X") indicated.

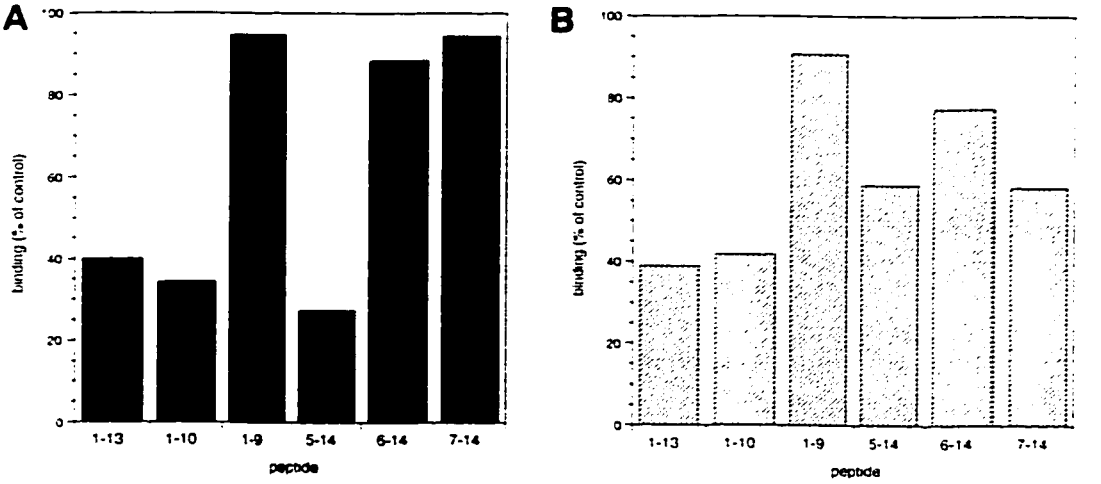


Fig. 4.6. **Effect of peptide truncations in peptide binding competitions.** Wild-type (A) or αA2 Arg mutant (B) SH2 domains (500 nM each) were incubated with the indicated peptides at 1.5 μM in HBS buffer. Binding is expressed as described for Fig. 4.5.

Table 4.2. **IC_{50} values for peptide competitors with wild-type and αA2 Arg SH2 domains.**

peptide competitor	sequence[a]	IC_{50} (wild-type) ± S.E. (μM)[b]	IC_{50} (αA2 Arg) ± S.E. (μM)[b]
1-13	DTSSVL(pY)TAVQPN	1.3 ± 0.1	1.2 ± 0.1
5-10	VL(pY)TAV	0.56 ± 0.04	2.5 ± 0.3
5-10 Ala5	AL(pY)TAV	11.3 ± 0.7	1.5 ± 0.2
5-10 Thr5	TL(pY)TAV	1.7 ± 0.5	1.8 ± 0.2
5-10 Ser5	SL(pY)TAV	19.2 ± 1.5	2.2 ± 0.3
5-10 Leu5	LL(pY)TAV	1.0 ± 0.1	2.7 ± 0.1
5-10 Ala10	VL(pY)TAA	5.3 ± 0.6	24.8 ± 1.5
5-14	VL(pY)TAVQPNE	0.90 ± 0.03	2.5 ± 0.4
6-14	L(pY)TAVQPNE	24.3 ± 2.6	5.4 ± 0.7
7-14	Ac-(pY)TAVQPNE	35.1 ± 2.0	2.2 ± 0.2

[a] All peptides have carboxyamide at the C-terminus and a free amino group at the N-terminus, except the 7-14 peptide which is N-acetylated (indicated by "Ac").

[b] Values were determined from BIAcore binding assay, based on the amount of peptide required to reduce the maximal SH2 domain binding to the immobilized 1-13 peptide in the absence of peptide competitors by 50%. The IC_{50} values were determined from a graphical fit as described previously (11).

To explore further the peptide binding determinants, truncated peptides were tested for their ability to compete for binding. The results with the wild-type SH2 domain were as expected: peptides that contained the minimum 5-10 sequence [VL(pY)TAV] competed as well as the 1-13 peptide, while truncations that extended into this sequence (*i.e.*, the 1-9, 6-14, and 7-14 peptides) did not compete (Fig. 4.6a). With the αA2 Arg SH2 domain, the same results was observed with C-terminal truncated peptides: the 1-10 peptide competed well while the 1-9 peptide was a poor competitor (Fig. 4.6b). However, substantial differences were observed with the N-terminal truncated peptides:

removal of N-terminal residues had little effect on the ability of the peptides to compete for binding (Fig. 4.6b). The differences are clearly apparent by comparing the IC_{50}'s of the N-terminal truncated peptides (Table 4.2): removal of the N-terminal residues increased the IC_{50} 40-fold for the wild-type SH2 domains but had no effect on the IC_{50} with the αA2 Arg SH2 domain. The higher IC_{50} of the 6-14 peptide is presumably due to the fact that it is not N-acetylated, since non-acetylated 7-14 peptide was a poorer competitor than N-acetylated 7-14 peptide (data not shown).

Comparison of in Vivo Binding Target Sequences – The involvement of the pY-2 residue in directing high-affinity binding to the SH2 domains of SHP-1 has also been demonstrated (12), which by analogy to SHP-2 is likely a consequence of their having Gly residues at the αA2 position (Fig. 4.1). A number of *in vivo* binding targets have been identified for SHP-2 and SHP-1 and the sites of association mapped to particular pY residues (Table 4.3). A comparison of the sequences surrounding the target pY residues further supports the importance of the pY-2 residue, as all of these sequences have a hydrophobic residue at this position. The consensus binding sequence for these SH2 domains can thus be extended to hydrophobic–X–(pY)–hydrophobic–X–hydrophobic, with a preference for Val/Leu/Ile at the pY-2 and pY+3 positions.

SHP-1 has also been shown to associate with c-Kit (30), but the actual pY residue to which the SH2 domains bind has not been determined. Based on this extended consensus binding sequence, a likely site for SHP-1 association with c-Kit is at Tyr_{730} (VSYVVP), assuming this Tyr represents an *in vivo* phosphorylation site. Similarly, SHP-2 associates with the platelet endothelial cell adhesion molecule-1 (PECAM-1) at an

unknown site (31). Of the five intracellular Tyr residues, only the sequence surrounding Tyr_{663} (VQYTEV) conforms to the high-affinity binding sequence (32) and represents the likely site of association if it is phosphorylated *in vivo*. Interestingly, there is also a Tyr at position 686 which may provide a binding target for the C-terminal SH2 domain of SHP-2.

Table 4.3. **Sequences of identified *in vivo* binding sites of SHP-2 and SHP-1.**

Enzyme	*In vivo* binding target[a]	Binding sequence[b]	Reference
SHP-2	IRS-1	SLN(pY$_{1172}$)IDLDLV	20
SHP-2	PDGFR	SVL(pY$_{1009}$)TAVQPN	21
SHP-2	BIT	DIT(pY$_{436}$)ADLNLP	22
SHP-2	BIT	TLT(pY$_{477}$)ADLDMV	22
SHP-2	IR	HIP(pY$_{1322}$)THMNGG	23
SHP-2	EpoR[c]	SFE(pY$_{427}$)TILDPS	24
SHP-2	IL-3 ßc	SLE(pY$_{612}$)LCLPAG	25
SHP-1	CD22	TVS(pY$_{783}$)AILRFP	26
SHP-1	CD22	SIH(pY$_{843}$)SELVQF	26
SHP-1	CD22	DVD(pY$_{863}$)VTLKH	26
SHP-1	FCγR$_{IIB1}$	TIT(pY$_{309}$)SLLKHP	27
SHP-1	EpoR[c]	HLK(pY$_{455}$)LYLVVS	28
SHP-1	IL-3 ßc	SLE(pY$_{612}$)LCLPAG	25

[a] IRS-1, insulin receptor substrate 1; BIT, brain immunoglobulin-like molecule with tyrosine-based activation motifs; IR, insulin receptor; IL-3 ßc, interleukin 3 receptor ßc subunit; FCγR$_{IIB1}$, B-cell Fc receptor for IgG; EpoR, erythropoietin receptor.

[b] The number indicates the location of the pY residue in the target protein sequence. The pY-2, pY+1, and pY+3 residues are highlighted in bold.

[c] The amino acid sequence numbering for EpoR is taken from Yi *et al.* (29).

DISCUSSION

The PTPs SHP-2 and SHP-1 play essential roles in several tyrosine phosphorylation signaling pathways, as evidenced from the number of *in vivo* binding targets that have been identified (Table 4.3). Furthermore, mice carrying mutations in these PTPs display severe developmental defects (41, 42). The SH2 domains of these PTPs direct their association with the appropriate targets to form specific, regulated signaling complexes, which may serve to bring SHP-1 and SHP-2 in proximity of their substrates. As well, the SH2 domains modulate the catalytic activity of these PTPs (6). Thus, knowledge of the SH2 domain binding specificity is clearly important to determine the interaction partners of SHP-1 and SHP-2 and place them in the appropriate pathways as well as to understand their regulation.

The SH2 domains of SHP-2 and SHP-1 have an unusual requirement for residues N-terminal to the target pY for high-affinity binding (11, 12). We have shown here that for the N-terminal SH2 domain of SHP-2, this requirement arises from the αA2 Gly residue in place of a highly-conserved Arg. The wild-type and αA2 Arg mutant SH2 domains have distinct binding specificities as demonstrated by the peptide binding competitions (Figs. 4.5 and 4.6 and Table 4.2). In particular, interactions between the SH2 domain and residues N-terminal to the target pY were eliminated in the αA2 Arg mutant. The αA2 Arg mutant showed no selectivity towards a variety of pY-2 substituted peptides as compared to wild-type (Fig. 4.5b *vs.* Fig. 4.5a, and Table 4.2). Furthermore, removal of all residues N-terminal to the pY had no appreciable effect on binding to the αA2 Arg SH2 domain (Table 4.2; compare "5-14" and "7-14" peptides), while affinity for the wild-type SH2 domain was decreased 40-fold.

Changing the αA2 Gly of the N-terminal SH2 domain of SHP-2 to Arg had little effect on binding to the PDGFR 1009 peptide; in fact, the binding kinetic constants were almost identical for the wild-type and αA2 Arg SH2 domains (Table 4.1). The αA2 position has previously been shown to be relatively tolerant to changes, as mutations at this position in the N-terminal SH2 domain of Gap had only minor effects on binding (19). However, in this study only the αA2 Arg-substituted SH2 domain was competent for high-affinity binding: substitution with five other residues at this position resulted in weaker binding that was not detectable in the BIAcore assay (Fig. 4.4). Because of the αA2 Gly, there is a major change in the pY binding as the phosphate of the pY residue is rotated ~180° relative to the Src and Lck complexes (Fig. 4.2). It is likely that in the Gly→Arg mutant, the pY phosphate binds in an orientation similar to that for the Src and Lck SH2 domains, although a structural determination is required to confirm this. Thus, while the interaction with the pY-2 residue is eliminated, compensatory amino–aromatic interactions between the αA2 Arg and the phenyl ring of the pY may be created similar to that observed in the Src and Lck SH2 domains, accounting for the similar binding affinities of the wild-type and αA2 Arg SH2 domains. Such interactions would not be formed in the other SH2 domain mutants made here, consistent with these SH2 domains not binding efficiently to the PDGFR 1009 peptide. These data suggest that the αA2 Arg is directly involved in maintaining a high-affinity interaction and is responsible for the change in specificity.

From Fig. 4.5a and Table 4.3, the preferred residues at the pY-2 position for SHP-2 are Val, Leu, and Ile. A more systematic study along the lines of the affinity selection

studies of Songyang *et al.* (7) is required to define the precise selectivity at this position. Even though the C-terminal SH2 domain of SHP-2 has a Gly residue at the αA2 position, its pY-2 specificity may be different. A study of the two SH2 domains suggested that amongst peptides derived from IRS-1, the C-terminal SH2 domain has the highest affinity for the sequence surrounding Tyr_{895}, SPGE(pY)VNIEFGS (36), which has Gly at the pY-2 position. Similarly, SHP-2 has been reported to associate through its SH2 domains with CTLA-4 at Tyr_{201} [TTGV(pY)VKMPPT] (37), again with a pY-2 Gly residue; this association may be directed primarily by the C-terminal SH2 domain. SHP-2 has also been shown to associate with the EGF receptor (38, 39) at an unknown pY site. None of the sequences surrounding Tyr residues in the cytosolic domain of the receptor conforms to the predicted high-affinity consensus binding sequence for the N-terminal SH2 domain; thus, the association may be directed by the C-terminal SH2 domain or by an unknown mechanism. A more detailed comparative study of the binding sequence preferences of the two SH2 domains, and structural studies of the C-terminal SH2 domain, would be informative in elucidating the basis for these apparent differences and the role of the αA2 Gly residue.

The fact that the αA2 Gly is responsible for the pY-2 interaction reinforces the involvement of N-terminal residues in the binding of the SH2 domains of SHP-2 (11). Similarly, the SH2 domains of SHP-1, the other mammalian SH2 domain-containing PTP, have Gly at the αA2 position (Fig. 4.1), and an interaction with the pY-2 residue has been demonstrated (12). An αA2 Gly is also found in the SH2 domain of Ctk (33) [also known as Matk (34) and Hyl (35)], a member of the Csk family of PTKs. While there is no

information on the binding specificity of the Ctk SH2 domain, the presence of the αA2 Gly suggests by analogy to SHP-2 and SHP-1 that residues N-terminal to the target pY are involved. Thus, these αA2 Gly-containing SH2 domains may define another class with specificity for residues both N- and C-terminal to the target pY. From the results presented in this study, the consensus binding sequence for high-affinity binding appears to be hydrophobic–X–(pY)–hydrophobic–X–hydrophobic, with a preference for aliphatic amino acids at the pY-2 and pY+3 positions. Indeed, almost every *in vivo* binding target for the SH2 domains of SHP-2 and SHP-1 has Val, Leu, or Ile at the pY-2 position (Table 4.3), further supporting these observations. This binding mode has important implications for identifying *in vivo* binding targets for these critical PTPs in cell signaling and reinforces the fact that SH2 binding specificity is defined by more than the three residues C-terminal to the target pY.

ACKNOWLEDGMENTS

We thank M. Abramovitz for assistance with the mutagenesis strategies and techniques, C. Bayly for molecular modeling work and many stimulating and helpful discussions, L. Stolz for assistance with the BIAcore analysis, and A. Filosa for mass spectrometry. We also thank M. Gresser for his support and A. English for critical reading of the manuscript.

REFERENCES

1. Cohen, G. B., Ren, R., and Baltimore, D. (1995) *Cell* **80,** 237-248

2. Pawson, T. (1995) *Nature* **373,** 573-580

3. Schaffhausen, B. (1995) *Biochim. Biophys. Acta* **1242,** 61-75

4. Songyang, Z., and Cantley, L. C. (1995) *Trends Biochem. Sci.* **20,** 470-475

5. Neel, B. G., and Tonks, N. K. (1997) *Curr. Opin. Cell Biol.* **9,** 193-204

6. Frearson, J. A., and Alexander, D. R. (1997) *BioEssays* **19,** 417-427

7. 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., and Cantley, L. C. (1993) *Cell* **72,** 767-778

8. von Bonin, A., Wienands, J., Manning, U., Zuber, J. F., and Baumann, G. (1994) *J. Biol. Chem.* **269,** 33035-33041

9. Songyang, Z., Gish, G. D., Mbamalu, G., Pawson, T., and Cantley, L. C. (1995) *J. Biol. Chem.* **270,** 26029-26032

10. Marengere, L. E. M., Songyang, Z., Gish, G. D., Schaller, M. D., Parsons, J. T., Stern, M. J., Cantley, L. C., and Pawson, T. (1994) *Nature* **369,** 502-505

11. Huyer, G., Li, Z. M., Adam, M., Huckle, W. R., and Ramachandran, C. (1995) *Biochemistry* **34,** 1040-1049

12. Burshtyn, D. N., Yang, W., Yi, T., and Long, E. O. (1997) *J. Biol. Chem.* **272,** 13066-13072

13. Lee, C.-H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E., and Kuriyan, J. (1994) *Structure* **2,** 423-438

14. Bastien, L., Ramachandran, C., Liu, S., and Adam, M. (1993) *Biochem. Biophys. Res. Comm.* **196,** 124-133

15. Horton, R. M., Cai, Z., Ho, S. N., and Pease, L. R. (1990) *BioTechniques* **8,** 528-535

16. Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. D., Silverman, L., and Kuriyan, J. (1992) *Nature* **358,** 646-653

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

18. Eck, M. J., Shoelson, S. E., and Harrison, S. C. (1993) *Nature* **362,** 87-91

19. Marengere, L. E. M., and Pawson, T. (1992) *J. Biol. Chem.* **267,** 22779-22786

20. Sun, X. J., Crimmins, D. L., Myers, M. G., Jr., Miralpeix, M., and White, M. F. (1993) *Mol. Cell. Biol.* **13,** 7418-7428

21. Lechleider, R. J., Sugimoto, S., Bennett, A. M., Kashishian, A. S., Cooper, J. A., Shoelson, S. E., Walsh, C. T., and Neel, B. G. (1993) *J. Biol. Chem.* **268,** 21478-21481

22. Ohnishi, H., Kubota, M., Ohtake, A., Sato, K., and Sano, S. (1996) *J. Biol. Chem.* **271,** 25569-25574

23. Staubs, P. A., Reichart, D. R., Saltiel, A. R., Milarski, K. L., Maegawa, H., Berhanu, P., Olefsky, J. M., and Seely, B. L. (1994) *J. Biol. Chem.* **269,** 27186-27192

24. Tauchi, T., Damen, J. E., Toyama, K., Feng, G.-S., Broxmeyer, H. E., and Krystal, G. (1996) *Blood* **87,** 4495-4501

25. Bone, H., Dechert, U., Jirik, F., Schrader, J. W., and Welham, M. J. (1997) *J. Biol. Chem.* **272,** 14470-14476

26. Doody, G. M., Justement, L. B., Delibrias, C. C., Matthews, R. J., Lin, J., Thomas, M. L., and Fearon, D. T. (1995) *Science* **269,** 242-244

27. D'Ambrosio, D., Hippen, K. L., Minskoff, S. A., Mellman, I., Pani, G., Siminovitch, K. A., and Cambier, J. C. (1995) *Science* **268,** 293-297

28. Klingmuller, U., Lorenz, U., Cantley, L. C., Neel, B. G., and Lodish, H. F. (1995) *Cell* **80,** 729-738

29. Yi, T., Zhang, J., Miura, O., and Ihle, J. N. (1995) *Blood* **85,** 87-95

30. Yi, T., and Ihle, J. N. (1993) *Mol. Cell. Biol.* **13,** 3350-3358

31. Masuda, M., Osawa, M., Shigematsu, H., Harada, N., and Fujiwara, K. (1997) *FEBS Lett.* **408,** 331-336

32. Newman, P. J., Berndt, M. C., Gorski, J., White, G. C., Lyman, S., Paddock, C., and Muller, W. A. (1990) *Science* **247,** 1219-1222

33. Klages, S., Adam, D., Class, K., Fargnoli, J., Bolen, J. B., and Penhallow, R. C. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 2597-2601

34. Bennet, B. D., Cowley, S., Jiang, S., London, R., Deng, B., Grabarek, J., Groopman, J. E., Goeddel, D. V., and Avarham, H. (1994) *J. Biol. Chem.* **269,** 1068-1074

35. Sakano, S., Iwama, A., Inazawa, J., Ariyama, T., Ohno, M., and Suda, T. (1994) *Oncogene* **9,** 1155-1161

36. Sugimoto, S., Wandless, T. J., Shoelson, S. E., Neel, B. G., and Walsh, C. T. (1994) *J. Biol. Chem.* **269,** 13614-13622

37. Marengere, L. E. M., Waterhouse, P., Duncan, G. S., Mittrucker, H.-W., Feng, G.-S., and Mak, T. W. (1996) *Science* **272,** 1170-1173

38. Feng, G.-S., Hui, C.-C., and Pawson, T. (1993) *Science* **259,** 1607-1611

39. Vogel, W., Lammers, R., Huang, J., and Ullrich, A. (1993) *Science* **259,** 1611-1614

40. Ferrin, T. E., Huang, C. C., Jarvis, L. E., and Langridge, R. (1988) *J. Mol. Graphics* **6,** 13-27

41. Saxton, T. M., Henkemeyer, M., Gasca, S., Shen, R., Rossi, D. J., Shalaby, F., Feng, G. S., and Pawson, T. (1997) *EMBO J.* **16,** 2352-2364

42. Tsui, H. W., Siminovitch, K. A., de Souza, L., and Tsui, F. W. L. (1993) *Nature Genet.* **4,** 124-129

CHAPTER 5

Mechanism of Inhibition of Protein Tyrosine Phosphatases by Vanadate and Pervanadate*

SUMMARY

Vanadate and pervanadate (the complexes of vanadate with hydrogen peroxide) are two commonly-used general protein tyrosine phosphatase (PTP) inhibitors. These compounds also have insulin mimetic properties, an observation that has generated a great deal of interest and study. As a careful kinetic study of the two inhibitors has been lacking, we sought to analyze their mechanisms of inhibition. Our results show that vanadate is a competitive inhibitor for the protein tyrosine phosphatase PTP1B, with a K_i of 0.38 ± 0.02 μM. EDTA, which is known to chelate vanadate, causes an immediate and complete reversal of the inhibition due to vanadate when added to an enzyme assay. Pervanadate, by contrast, inhibits by irreversibly oxidizing the catalytic cysteine of PTP1B, as determined by mass spectrometry. Reducing agents such as DTT that are used in PTP assays to keep the catalytic cysteine reduced and active were found to convert pervanadate rapidly to vanadate. Under certain conditions, slow time-dependent inactivation by vanadate was observed; since catalase blocked this inactivation, it was ascribed to *in situ* generation of hydrogen peroxide and subsequent formation of pervanadate. Implications for the use of these compounds as inhibitors and rationalization for some of their *in vivo* effects are considered.

---

* Reprinted with permission from *The Journal of Biological Chemistry* 1997, 272, 843-851.

INTRODUCTION

Protein tyrosine phosphorylation plays a central role in regulating a variety of fundamental cellular processes (1–3). The tyrosyl phosphorylation state of a protein in the cell reflects the balance between the competing activities of the protein tyrosine kinases (PTKs)[1] and the protein tyrosine phosphatases (PTPs). Substantial progress has been made in understanding the function of PTKs while increased attention has only been recently focused on the PTPs. The identification of a large family of PTPs (4), comprising two distinct groups, arose from the purification and sequencing of human placental PTP1B (5–7). The first group of PTPs has been described as receptor-like with a single transmembrane domain, one or two intracellular catalytic domains and a unique extracellular domain of variable length. Several PTPs in this group also contain regions representing putative ligand binding domain motifs. The second group consists of cytoplasmic enzymes, having a single catalytic domain and a variable amino or carboxyl terminal regulatory domain.

PTPs are specific for the dephosphorylation of phosphotyrosyl residues of proteins and peptides (reviewed in 8–11). The hallmark of PTPs is an essential cysteine residue at the catalytic site which forms a thiol-phosphate intermediate during catalysis (12, 13). Since reducing conditions usually maintained by thiol reagents are necessary for keeping the enzyme active, thiol oxidizing agents are potent PTP inhibitors. Vanadate (VO_4^{3-}) is a general PTP inhibitor (14), whose mechanisms of action have not been investigated in

---

[1] Abbreviations: PTK, protein tyrosine kinase; PTP, protein tyrosine phosphatase; EDTA, ethylenediaminetetraacetic acid; IPTG, isopropylthio-ß-D-galactoside; DTT, dithiothreitol; PBS, phosphate-buffered saline; HEPES, N-(hydroxyethyl) piperazine-N'-(2-ethane sulfonic acid); BSA, bovine serum albumin; FDP, fluorescein diphosphate; FMP, fluorescein monophosphate; MS, mass spectrometry; MS/MS, tandem mass spectrometry.

detail. Vanadate is a phosphate analog and is generally thought to bind as a transition state analog to the phosphoryl transfer enzymes which it inhibits, as it can easily adopt a trigonal bipyramidal structure (15). Vanadate has a wide variety of effects on biological systems (16). It is insulin-mimetic (17) and has been shown in human clinical trials to be potentially useful in treating both insulin- and noninsulin-dependent diabetes mellitus (18). It has been suggested that part of vanadate's insulin-mimetic effect may be due to its inhibition of PTPs (19). Another inhibitor on which considerable attention has been recently focused is pervanadate (a general term for the variety of complexes formed between vanadate and hydrogen peroxide). Pervanadate is also insulin mimetic and appears to be more effective than vanadate in increasing the level of cellular tyrosine phosphorylation (20–25). Peroxovanadium complexes containing one or more chelating ligands in addition to the oxo and peroxo ligands are generally more stable and exhibit even more potent insulin mimetic effects (26–28).

While vanadate and pervanadate have been widely studied for their insulin mimetic effects and are commonly used as general inhibitors of PTPs, their mechanisms of inhibition have not been carefully studied. Here we show that vanadate and pervanadate inhibit PTPs by completely different mechanisms. Furthermore, common buffer components such as EDTA and reducing agents can interact with these inhibitors, greatly affecting their potencies. These medium effects and the different modes of inhibition of vanadate and pervanadate have important implications for their use and provide a rationalization for their different effects both *in vitro* and *in vivo*.

EXPERIMENTAL PROCEDURES

Materials – Vanadium (V) oxide (99.99%) and hydrogen peroxide (30%) were from Aldrich. Bovine liver catalase was obtained from Sigma, and sequencing grade trypsin was from Boehringer Mannheim. All other chemicals were of reagent grade from Sigma.

Expression and Purification of PTP1B – The catalytic domain of PTP1B, consisting of amino acids 1–321, was inserted into the *Eco*RI site of the pFLAG-2 vector (Kodak). Bacterial cultures were grown in Terrific Broth (29) containing 100 μg/ml ampicillin and 0.4% glucose at 37°C, 225 rpm, for 2 – 3 hours (OD_{600} = 0.4 – 0.8). The culture was then induced with the addition of IPTG to 0.5 mM and grown overnight at 27°C, 225 rpm. The induced culture (500 ml) was centrifuged at 4°C for 20 min at 6000 x *g*. All subsequent steps were carried out at 4°C. The cell pellet was resuspended in 25 ml buffer A (10 mM NaH_2PO_4 pH 7.4, 150 mM NaCl, 0.67 mg/ml lysozyme, 2 mg/ml each pepstatin and aprotinin, 1 mM DTT) and then 5 ml of buffer B (1.5 M NaCl, 0.1 M $MgCl_2$, 0.1 M $CaCl_2$, 250 mg/ml DNase I) were added. The resuspended cell pellet was sonicated (5 x 10 s bursts), incubated on ice 10 – 15 min or until no longer viscous, and centrifuged for 15 min at 43,000 x *g*. The supernatant was then loaded onto a 17.5 ml M2 FLAG monoclonal antibody affinity column (Interscience), the column was washed with PBS (10 column volumes), and the FLAG-PTP1B fusion protein was eluted in PBS by competition with the FLAG peptide as described by the manufacturer. Approximately 5 – 7 mg of PTP1B (>95% pure as estimated by SDS–PAGE) were recovered from a 500 ml culture.

Preparation of Vanadate and Pervanadate – Vanadate stock solution was prepared by dissolving vanadium (V) oxide in 2.1 molar equivalents of 1 N NaOH, stirring the solution

until the yellow color had essentially disappeared (2 – 3 days), and then diluting with H_2O to a final concentration of 100 mM (30). Pervanadate stock solution (1 mM) was made by adding 10 µl of 100 mM vanadate and 50 µl of 100 mM hydrogen peroxide (diluted from a 30% stock in 20 mM HEPES pH 7.3) to 940 µl of H_2O. Excess hydrogen peroxide was removed by adding catalase (100 µg/ml final concentration = 260 units/ml) 5 min after mixing the vanadate and hydrogen peroxide. The pervanadate solutions were used within 5 min to minimize decomposition of the vanadate–hydrogen peroxide complex (31).

FDP Assay for PTP1B Activity – PTP1B was assayed in a buffer containing final concentrations of 25 mM HEPES pH 7.3, 5 mM DTT, and 10 µg/ml BSA using FDP as a substrate. Activity was measured by following the increase in absorbance at 450 nm due to FMP production[2] using a Hewlett Packard 8452A diode array spectrophotometer (ε_{450} for FMP = 27,500 $M^{-1}cm^{-1}$).

Mass Spectrometric Analysis – PTP1B samples were prepared for MS by incubating 9.4 µM enzyme on ice in 25 mM HEPES pH 7.3, 5 mM NaH_2PO_4, and 75 mM NaCl in the presence and absence of 47.6 µM pervanadate. The reaction with pervanadate was quenched after 15 min with the addition of DTT to a final concentration of 10 mM. The protein samples were analyzed by capillary HPLC-electrospray ionization mass spectrometry (capillary LC-MS). The capillary column used was a 254 µm i.d. x 120 mm length of PEEK tubing packed with 10 µm POROS® R2 stationary phase (PerSeptive Biosystems Inc., Framingham, MA). A Waters 600-MS HPLC pump (Waters, Milford, MA) was used to supply a mobile phase gradient (10 to 80% aqueous acetonitrile, 0.5%

[2] Huang, Z., Wang, Q., Ly, H. D., Govindarajan, A., Scheigetz, J., Zamboni, R., and Ramachandran, C., manuscript in preparation.

glacial acetic acid in 20 min). The capillary column was attached directly to the exit port of a Valco 0.5 µl internal loop injector (Valco Instrument Co. Inc., Houston, TX). A pre-column splitter was installed between the pump and the injector to adjust the flow through the column to 15 µl/min (from a pump flow of 900 µl/min). The column eluate was fed directly to the electrospray ionization source of a TSQ 7000 triple quadrupole mass spectrometer (Finnigan Mat, San Jose, CA). The electrospray ionization voltage was +4.2 kV, the nebulization gas pressure was 30 psi and the inlet capillary temperature was 200°C. Full scan mass spectra (m/z 600–2000) were acquired in profile mode using the third quadrupole (Q3) as the mass analyzer. The protein mass spectra were analyzed and deconvoluted using the Bioworks™ application software (Finnigan Mat, San Jose, CA).

Tryptic peptides were prepared by digesting the samples with approximately 1/20 (w/w) sequencing-grade trypsin for 18 h at 37°C. The digests were analyzed by capillary LC-MS using the same instrument arrangement as above. In this case, a 300 µm i.d. x 150 mm Hypersil C_{18} (3 µm particle size) capillary column was used (Keystone Scientific Inc., Bellefonte, PA). This column was fitted with Slipfree fittings which allowed direct attachment to the exit port of the injector. Tryptic peptides were resolved using a gradient of aqueous isopropyl alcohol, 0.5% glacial acetic acid (2 to 10% in 2 min, 10 to 80% in 28 min) at 6 µl/min. The column eluate was fed directly to the electrospray ionization source which was operated under the same conditions as described above. Full scan mass spectra (m/z 300–2000) were acquired in centroid mode using Q3 as the mass analyzer.

For tandem mass spectrometric analysis of tryptic peptides, the digests were first resolved by capillary LC-MS as described above. Fragmentation of the selected precursor

ions was achieved by collisional activation with argon in the RF-only quadrupole (Q2). The collision gas pressure was 3.0 mTorr and the collision energy was 75 V (laboratory frame of reference). The fragment ion spectra were acquired in profile mode at an acquisition rate of 3.5 s per scan.

RESULTS

Inhibition of PTP1B by Vanadate and Pervanadate – When assaying inhibition of PTPs by vanadate or pervanadate, the choice of assay conditions is very important because vanadate is known to interact with many buffer salts and organic compounds. As HEPES is one of the few buffers that does not complex with vanadate (34, 35), it was used here. A reducing agent is essential under aerobic conditions to prevent oxidation of the catalytic Cys residue, which would inactivate the enzyme. DTT was chosen for these studies, and while it is known that DTT does complex with vanadate, these complexes are binuclear in vanadium and are not formed to a significant extent at the vanadium concentrations used in this study[3]. A small amount of BSA was also included as a carrier protein, and while BSA will complex with vanadate, the formation constant is ~1 X 10^3 M^{-1} (34), making the amount of complexation negligible at these concentrations. EDTA is commonly included in PTP assays; however, it is known to form a complex with vanadate (32–35), with a K_{eff} at pH 8.0 of 1.4 X 10^4 M^{-1} (34), and was thus excluded here.

In a side-by-side comparison of vanadate and pervanadate inhibition, the order of addition of enzyme and inhibitor was observed to be very important. When vanadate or pervanadate was added to the assay mixture before initiating the reaction with enzyme, no

[3] Tracey, A., personal communication.

difference in inhibition was observed (Fig. 5.1a), with 500 nM vanadate or pervanadate causing approximately 50% inhibition. Furthermore, the inhibition by both was completely reversible upon the addition of EDTA to the assay (Fig. 5.1a) or by dilution (data not shown). When vanadate or pervanadate was added to the assay mixture after the enzyme, clear differences between the two were observed (Fig. 5.1b). The extent of inhibition due to vanadate was the same as in Fig. 5.1a, while for pervanadate a greater amount of inhibition was observed, and it was only partially reversible with EDTA.

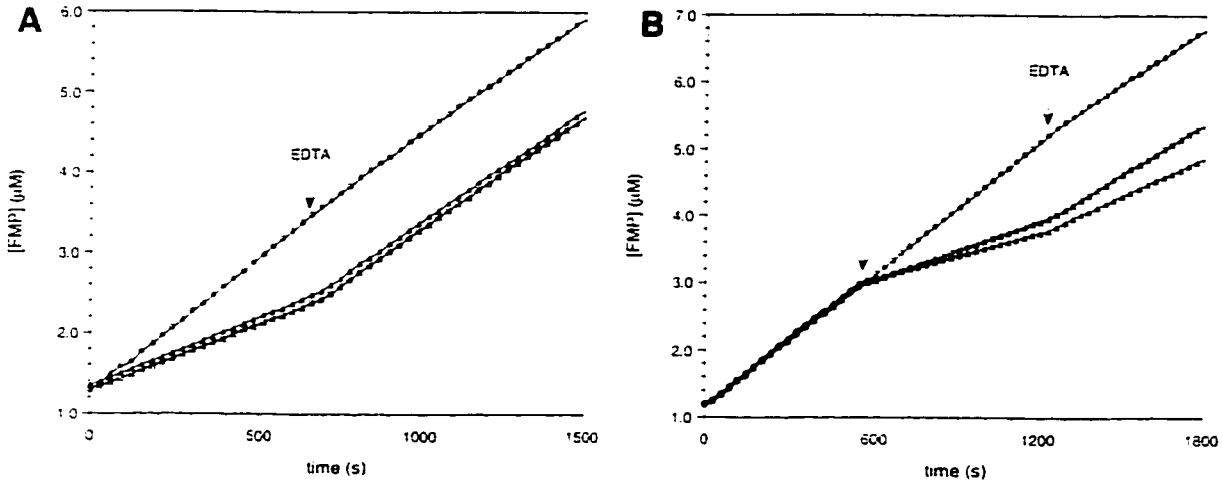


Fig. 5.1. **Inhibition of PTP1B by vanadate and pervanadate.** PTP1B (200 ng/ml final assay concentration) was assayed with 20 μM FDP as described under "Experimental Procedures." The time courses show the inhibition by 500 nM vanadate (squares) or pervanadate (triangles) relative to the control (circles). The inhibitors were added from 100 μM stock solutions 5 min before (A) or 9 min after (B, first arrow) enzyme. EDTA was added from a 0.5 M stock solution to a final concentration of 1 mM as indicated by the arrow.

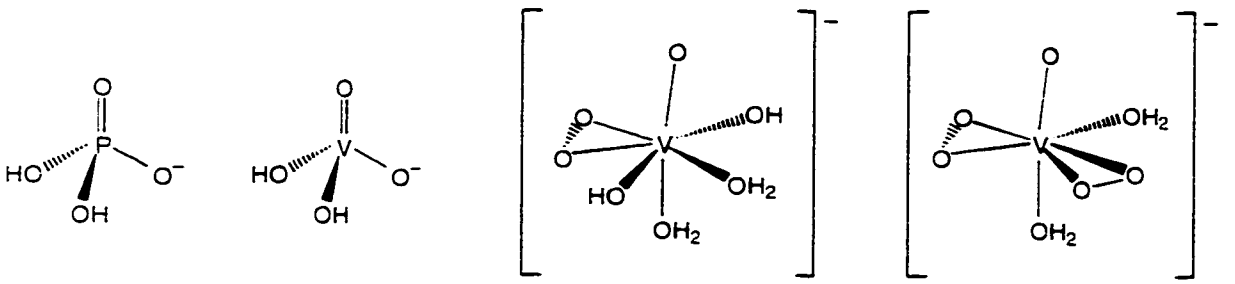


Fig. 5.2. **Structures of phosphate, vanadate, and pervanadate.** Vanadate in the +5 oxidation state adopts a tetrahedral structure similar to that of phosphate. Pervanadate exists as a number of complexes and is shown here in the monoperoxo (VL) and diperoxo (VL_2) forms, the major species present at the concentrations of vanadate and hydrogen peroxide used in these experiments. The pervanadate species are shown as hepta-coordinate structures with two water molecule ligands. However, the exact coordination geometry and the number and arrangement of water ligands is uncertain, and other structures may exist in solution (for further details, see refs. 31 & 36).

Mode of Vanadate Inhibition – Vanadate is generally considered as a phosphate analog, as it can adopt a similar structure to inorganic phosphate (Fig. 5.2) as well as a five-coordinate trigonal bipyramidal structure that resembles the transition state of many phosphoryl transfer reactions (15). The x-ray crystal structure of the *Yersinia* PTP complexed with vanadate shows that the vanadium molecule occupies the active site within covalent distance of the thiol of the catalytic Cys residue, forming a thiol-vanadyl ester linkage that resembles the covalent thiol-phosphate linkage formed during catalysis (37). These data would suggest that vanadate is a competitive inhibitor of PTPs. A Lineweaver-Burk analysis of vanadate inhibition was carried out (Fig. 5.3), and the pattern of inhibition corresponds to that expected for competitive inhibition, with a calculated K_i

of 0.38 ± 0.02 μM. The calculated K_m and k_{cat} for PTP1B of 21.7 ± 1.1 μM and 4.85 ± 0.07 s^{-1}, respectively, are in good agreement with previously determined values[2].

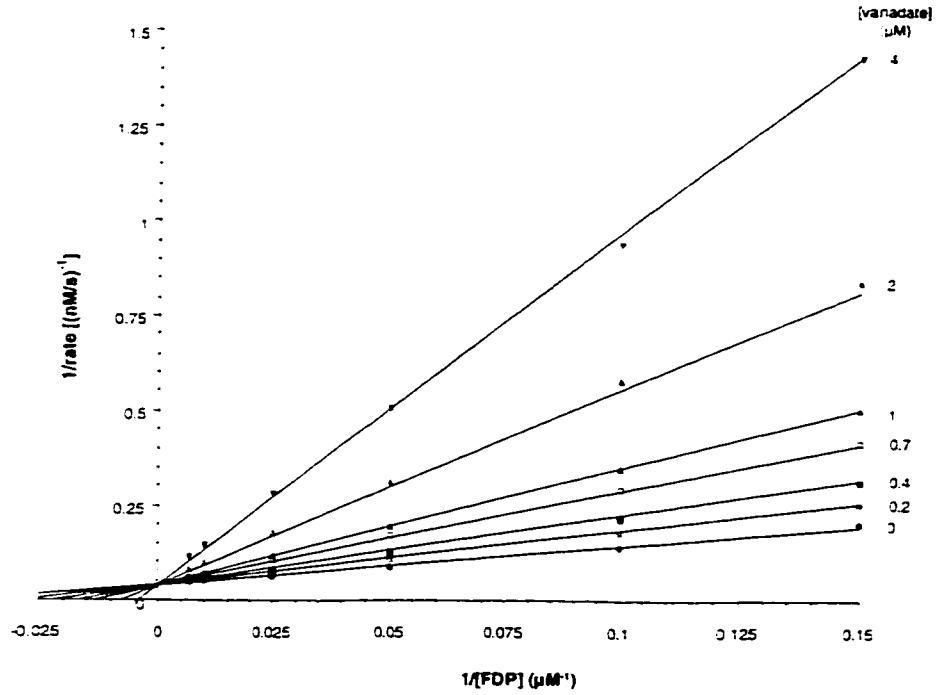


Fig. 5.3. **Lineweaver-Burk analysis of vanadate inhibition.** PTP1B (400 ng/ml final assay concentration) was assayed in the presence of the indicated vanadate concentrations while varying the FDP concentration, as described under "Experimental Procedures." Catalase (10 μg/ml) was included in the assay buffer as a precaution against the *in situ* formation of pervanadate (see "Results"). All the data points were fit simultaneously to the Lineweaver-Burk equation for competitive inhibition using the program GraFit. The values reported in the text for K_m, k_{cat}, and K_i were calculated from a non-linear fit of the data to the Michaelis-Menten equation for competitive inhibition using the same program.

Mode of Pervanadate Inhibition – From Fig. 5.1b, it is apparent that pervanadate is inhibiting PTP1B by a different mechanism than vanadate. When pervanadate is added to the enzyme assay last, the inhibition is only partially reversible with EDTA (Fig. 5.1b) or by dilution (data not shown), and the amount of irreversible inhibition increases with increasing pervanadate. The fact that no difference was observed between vanadate and

pervanadate when the two are added to the assay mixture before enzyme suggested that perhaps the pervanadate is being converted to vanadate. In fact, if pervanadate is pre-incubated in DTT alone before addition to the PTP assay solution, it behaves exactly like vanadate (data not shown), consistent with pervanadate being converted to vanadate by the presence of excess DTT; this was confirmed by NMR studies[3]. However, when pervanadate is added to the assay solution after enzyme, there is a portion of the inhibition that cannot be reversed with EDTA or by dilution, suggesting that the pervanadate is modifying or inactivating PTP1B before it is converted to vanadate by the DTT.

Vanadium peroxide complexes are known to be potent oxidizing agents (38), and the irreversible inhibition described above would be consistent with oxidation of the enzyme. In fact, Shaver *et al.* (39) have speculated that peroxovanadium compounds may be inhibiting PTPs by oxidizing the catalytic Cys residue. There are four oxidation states of cysteine that can be generated: disulfide (–S–S–), sulfenic acid (–SOH), sulfinic acid ($-SO_2H$), and sulfonic (or cysteic) acid ($-SO_3H$). Failure of DTT (or other reducing agents; data not shown) to reactivate the pervanadate-inactivated enzyme is consistent with one of the higher oxidation states. The latter three oxidations would result in a change in mass of PTP1B that could be measured by mass spectrometry. To test for this, PTP1B was incubated on ice for 15 min with an approximately five-fold molar excess of catalase-treated pervanadate, after which DTT was added to a final concentration of 10 mM to destroy the remaining pervanadate. Under these conditions essentially all of the enzyme was inactivated as determined by assaying the enzyme with FDP and comparing to a control incubation of PTP1B without pervanadate. The deconvoluted mass spectrum

obtained for the control protein by capillary LC-MS (inset in Fig. 5.4a) demonstrated that this sample consists primarily of a protein with a mass of 39,024 ± 1.3 Da. The reconstructed molecular weight profile of the pervanadate-treated protein (inset in Fig. 5.4b) indicated that the mass of the protein had increased by 47 Da, to 39,071 ± 0.6 Da. This could correspond to the addition of three oxygen atoms to the protein. As further controls, PTP1B was also incubated under the same conditions with the same concentrations of either vanadate or hydrogen peroxide used in the preparation of the pervanadate, resulting in no loss of activity in either case (data not shown). Thus, the inactivation of PTP1B by pervanadate and the corresponding increase in mass is clearly not due to the presence of vanadate nor free hydrogen peroxide.

Further examination of the trypsin-digested enzymes by capillary LC-MS indicated that the mass of the tryptic peptide containing the active-site cysteine (ESGSLSPEHGPVVVHCSAGIGR, mass = 2174.1 Da) had increased by 48 mass units after pervanadate treatment relative to the native digest (data not shown). In several replicate control digests, no modified active-site peptides were observed. Other Cys- and Met-containing peptides in the pervanadate-treated PTP1B tryptic digest were examined, and none was found to be modified. The predicted mass of FLAG-PTP1B from the sequence is 38,713.8, which is ~310 mass units less than the MS mass. Peptides corresponding to the entire protein sequence could be identified in the native tryptic digest, except for the extreme C-terminal 28 residues. Perhaps some post-translational modification on this portion of the enzyme gives rise to the increased mass, an hypothesis currently under investigation.

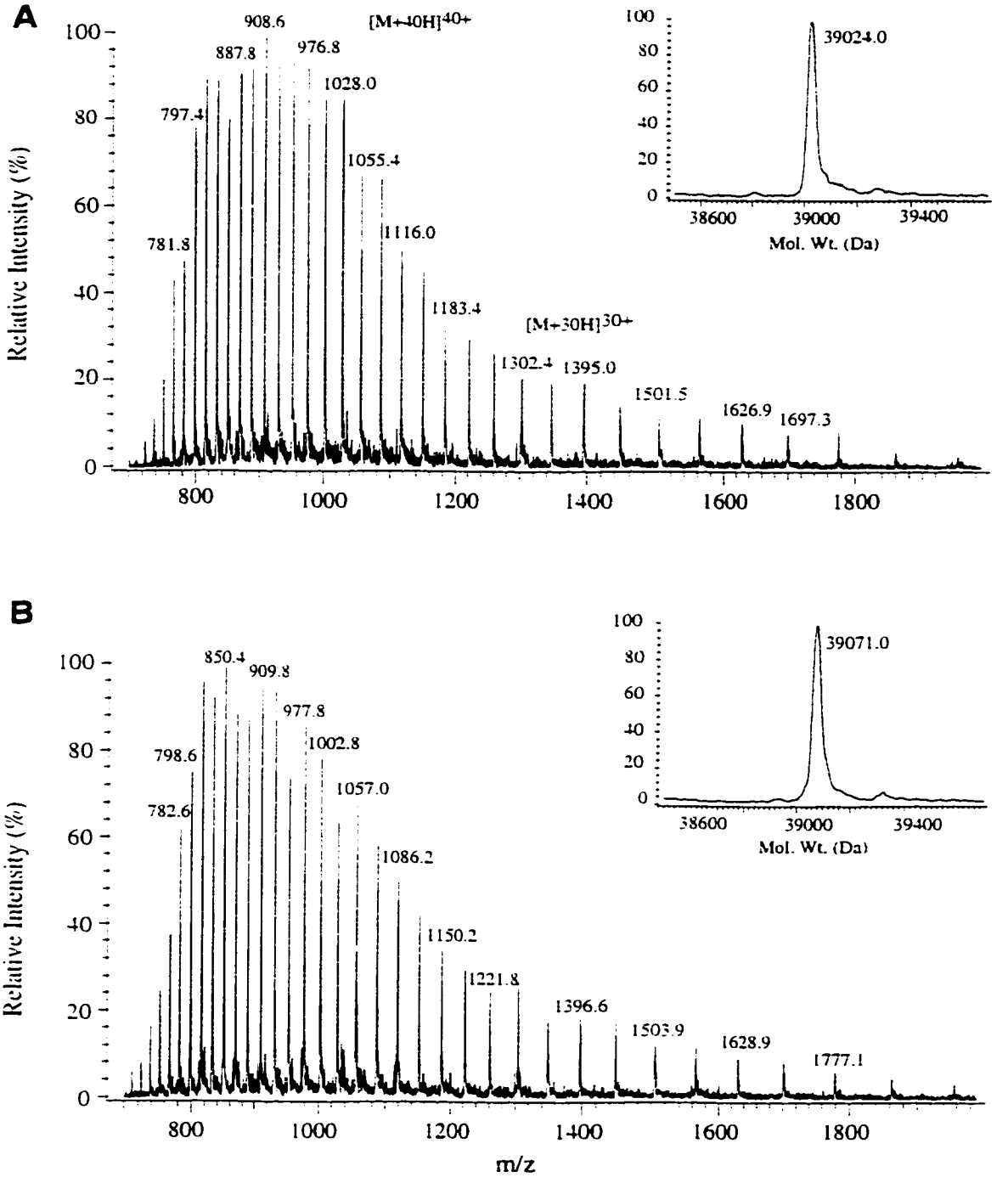


Fig. 5.4. **Mass spectrometry of native and pervanadate-inhibited PTP1B.** Capillary LC-MS spectra of native (A) and pervanadate-treated (B) PTP1B protein. Samples were prepared and analyzed as described under "Experimental Procedures." Approximately 250 ng (6.5 pmol) of protein sample were injected on-column in both cases. Each spectra is the average of 4 scans taken across the protein LC peak (m/z 600 – 2000, 4 s per scan). The insets represent the reconstructed molecular weight profiles obtained from these spectra.

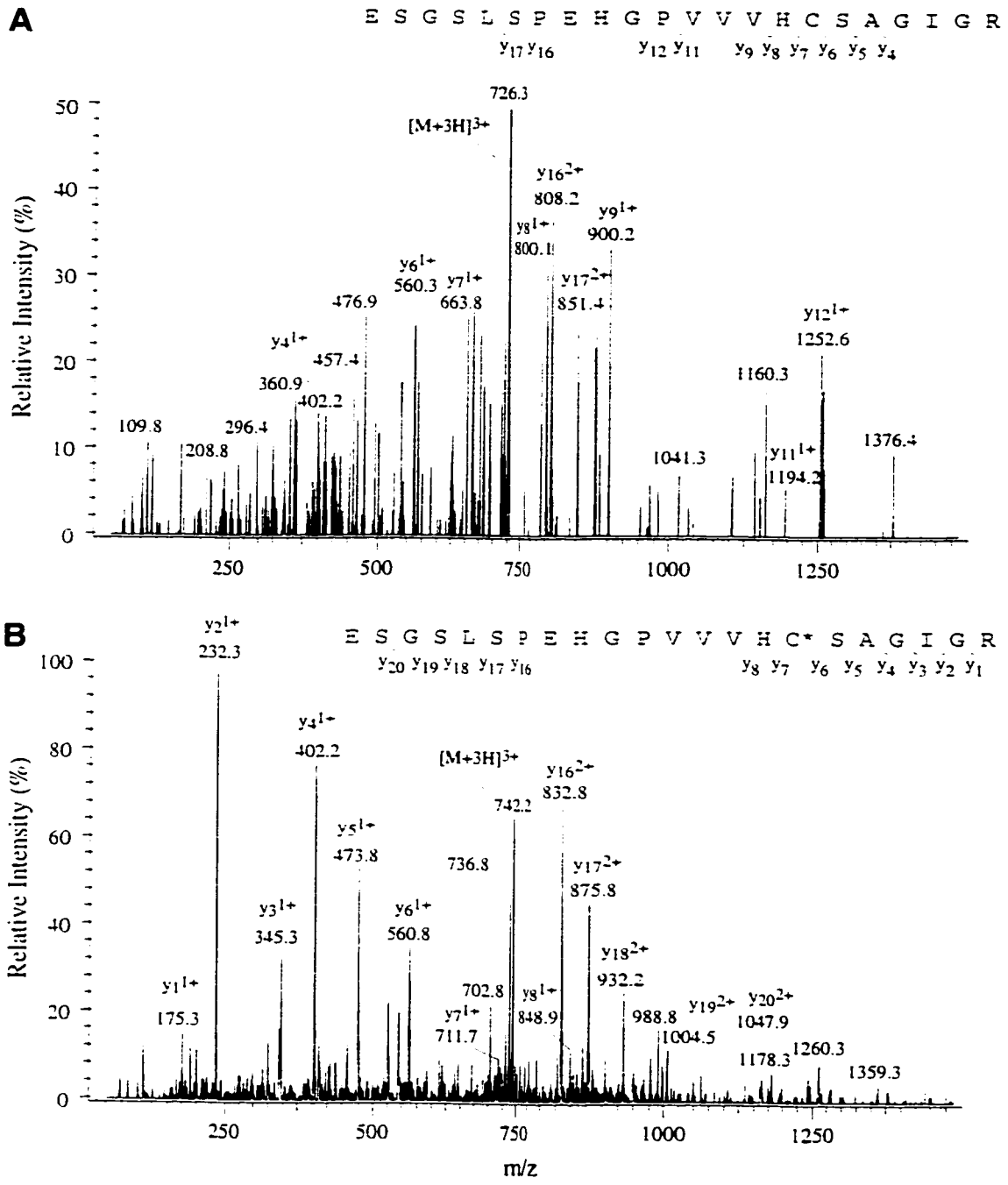


Fig. 5.5. **MS/MS analysis of active-site peptides.** Capillary LC-MS/MS spectra of the active-site peptide for the native (A) and pervanadate-treated (B) PTP1B protein. Approximately 250 ng of protein digest were injected on-column in each case. The precursor ions were m/z 726.3 (A) and m/z 742.3 (B). Each spectrum represents the average of two scans (m/z 50 – 2200, 3.5 s per scan) taken across the apex of the peptide LC peak. The sequence of the active-site containing tryptic peptide is shown, with the oxidized cysteine (cysteic acid) marked by an asterisk (C*) and the sites of Y-ion fragmentation indicated.

The active-site containing peptide (in both native and modified form) was observed principally as a triply-protonated ion. The MS/MS spectrum of the triply-protonated modified active-site peptide (Fig. 5.5b) is dominated by a series of singly- and doubly-charged Y ions (fragment ions containing the C-terminal residue) (40, 41). Although the ion intensity of the singly-charged Y ions decreases with increasing mass, it is evident that the extra 48 mass units were added to the Cys residue, confirming that the catalytic Cys is oxidized to the cysteic acid. Interestingly, the Y fragment ions are not as prominent in the MS/MS spectrum of the unoxidized active-site peptide (Fig. 5.5a), but they do confirm that the cysteine residue is not oxidized. A similar phenomenon was observed by Burlett *et al.* (42, 43) in the FAB-MS/MS spectra of peptides containing a cysteic acid close to a C-terminal arginine. They attributed this phenomenon to a gas phase interaction between these two residues leading to delocalization of the site of protonation and thus an increased yield of Y-series ions.

Kinetics of Pervanadate Inhibition – The irreversible inhibition observed when pervanadate was added to the enzyme assay last (after PTP1B) is consistent with the oxidation of the catalytic Cys residue by pervanadate. Time-dependent inhibition is normally anticipated in the case of an irreversible inactivator; however, pervanadate has a short half-life in the DTT-containing buffer. Upon addition of pervanadate, then, there are at least two competing reactions going on: conversion of pervanadate to vanadate by DTT, and oxidation of the catalytic Cys by pervanadate. Thus, once the pervanadate is destroyed by the DTT (which appears to occur too quickly to be measured on our kinetic time scale), there is no inhibitor available for any further inactivation. Increasing the

concentration of DTT in the enzyme assay was predicted to decrease the amount of inhibition observed upon addition of pervanadate to the enzyme assay by favoring the reaction of pervanadate with DTT over that with the enzyme: this was in fact observed (Fig. 5.6). Similar results were observed with the reducing agents ß-mercaptoethanol and reduced glutathione (data not shown), indicating that this effect is not exclusive to DTT. The assay mixtures in Fig. 5.6 contained EDTA to chelate vanadate and catalase to consume hydrogen peroxide, ensuring that the observed inhibition was due exclusively to inactivation by pervanadate.

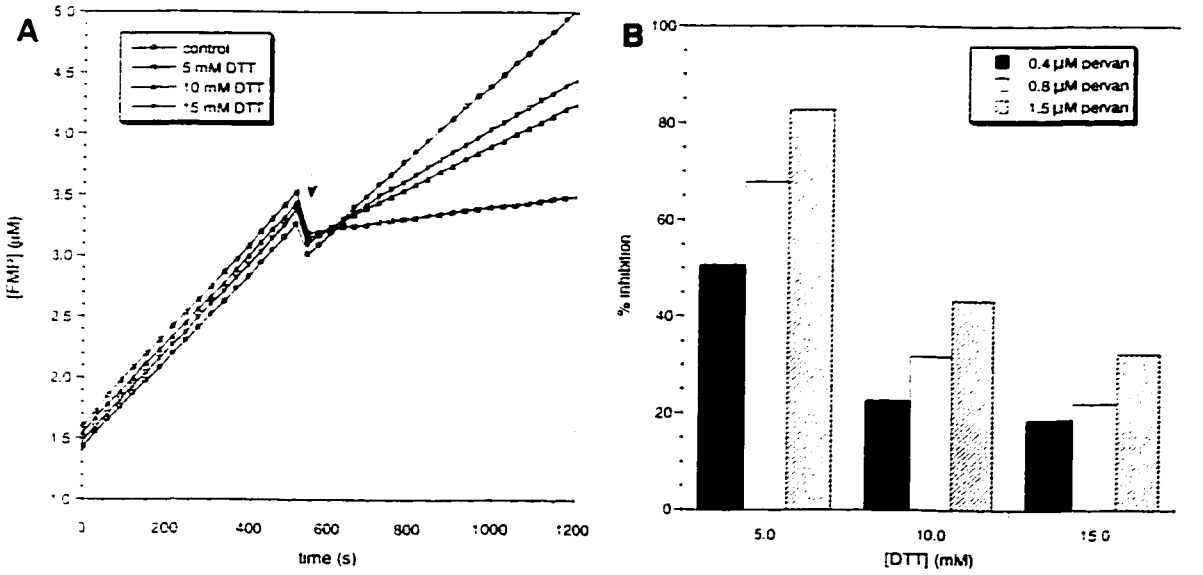


Fig. 5.6. **Inhibition of PTP1B by pervanadate.** PTP1B (200 ng/ml final assay concentration) was assayed with 20 µM FDP as described under "Experimental Procedures," with 1 mM EDTA and 10 µg/ml catalase included in the assay buffer. (A) Time courses showing the addition of 100 µl of 15 µM pervanadate to a 900 µl assay containing enzyme and DTT at the indicated concentrations. For the control, H_2O was added, and the rate was the same at all DTT concentrations used. (B) Bar graph showing the % inactivation relative to the control calculated from the time courses in (A) and at additional pervanadate concentrations as indicated.

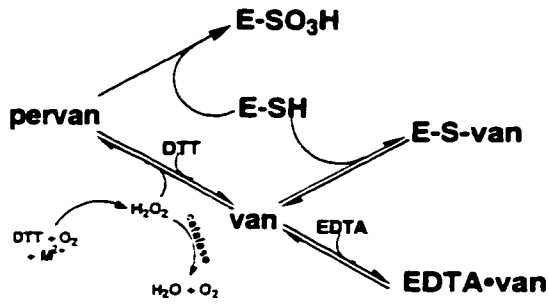


Fig. 5.7. **Scheme for vanadate and pervanadate inhibition of PTPs.** See "Results" for details. Pervan, pervanadate; van, vanadate; E-SH, catalytically active PTP (reduced catalytic Cys); E-SO_3H, PTP oxidized at the catalytic Cys residue.

The overall reaction scheme proposed for the inhibition by vanadate and pervanadate, and the effects of the thiol reductant on pervanadate and EDTA on vanadate, is outlined in Fig. 5.7. As similar inhibition kinetics for vanadate and pervanadate were observed with the PTP CD45 (data not shown), this scheme may be applicable to PTPs in general. A pathway by which pervanadate can re-form in the assay reaction is also shown in Fig. 5.7. Under certain assay conditions (*e.g.*, when imidazole buffer was used for the enzyme assays), a slow time-dependent inactivation of the enzyme that increased with increasing vanadate concentration was observed (data not shown). It is known that DTT auto-oxidation is catalyzed by transition metals, and superoxide is formed on the reduction of O_2 by a DTT thiol radical or by the reduced transition metal; the superoxide can then disproportionate to hydrogen peroxide and O_2 (44–46). Hence, the slow inactivation observed with vanadate is consistent with the generation of pervanadate *in situ* through the complexation of hydrogen peroxide with vanadate and subsequent oxidation of the catalytic Cys of the enzyme. The time-dependent inhibition was blocked by catalase and

stimulated by superoxide dismutase and Cu^{2+} (data not shown), supporting the mechanism in Fig. 5.7. Hydrogen peroxide itself is a much poorer inhibitor than pervanadate, as concentrations >100 μM were required for inhibition in our standard conditions (data not shown). However, the presence of vanadate in the enzyme assay greatly potentiates the effect of hydrogen peroxide (data not shown), consistent with the *in situ* formation of pervanadate as described above.

It is important to note that pervanadate is a mixture of several different complexes of vanadate with hydrogen peroxide (Fig. 5.2). At the concentrations of vanadate and hydrogen peroxide used here, the predominant species are monoperoxovanadate (VL) and diperoxovanadate (VL_2), as calculated from the equilibrium constants determined by Jaswal and Tracey (31). The literature equilibrium constants were obtained in 1.0 M ionic strength maintained with KCl, so the actual values here may be different but should follow the same trend. In order to determine the species active in oxidizing the enzyme, pervanadate solutions were prepared with different ratios of VL and VL_2 but a constant amount of VL + VL_2 (Table 5.1). These mixtures were added to PTP1B assays at a final concentration of VL + VL_2 = 400 nM and the extent of inhibition determined. As shown in Fig. 5.8, the amount of inhibition increases as the ratio of VL:VL_2 decreases, suggesting that VL_2 is a more active or potent inhibitor than VL. The concentrations of VL_3 and V_2L_4 present in the assay solution were estimated to be two to three orders of magnitude less than the enzyme concentration and for all practical purposes can be ignored. Little or no (≤5%) inhibition was observed when vanadate or free hydrogen peroxide was added separately to the PTP1B assay at the same concentrations used in the preparation of the

pervanadate solutions. This result was expected, since EDTA and catalase were present in the assay solution to respectively chelate the vanadate and remove the hydrogen peroxide.

Table 5.1. **Pervanadate species present in mixtures prepared for Figure 5.8.** Vanadate (V) and hydrogen peroxide (L) were mixed in distilled water at the indicated concentrations ($[V]_t$ and $[L]_t$) and allowed to equilibrate for 5 min before use. The equilibrium concentrations of V, L, and the various pervanadate species (VL, VL_2, VL_3, and V_2L_4) are indicated, as calculated from the published equilibrium constants (31).

$[V]_t$ (μM)	$[L]_t$ (μM)	[V] (μM)	[L] (μM)	[VL] (μM)	$[VL_2]$ (μM)	$[VL_3]$ (nM)	$[V_2L_4]$ (nM)	$[VL]+[VL_2]$ (μM)	$\frac{[VL]}{[VL_2]}$
25	22	21.0	15.3	1.32	2.70	0.68	0.47	4.03	0.49
100	11.5	96.0	5.73	2.27	1.75	0.17	0.19	4.02	1.30
225	8.3	221.0	3.11	2.83	1.18	0.060	0.089	4.01	2.40
360	7.0	356.0	2.12	3.11	0.88	0.031	0.050	4.00	3.52
500	6.3	496.0	1.60	3.28	0.71	0.019	0.032	3.99	4.65

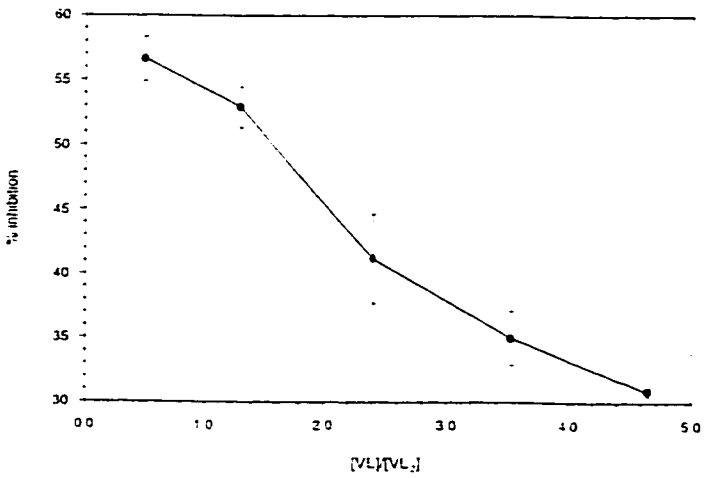


Fig. 5.8. **Dependence of pervanadate inhibition on the relative proportion of pervanadate species.** Pervanadate solutions were prepared with varying proportions of vanadate (V) and hydrogen peroxide (L) such that the major species generated were VL and VL_2 and the total concentration of VL + VL_2 remained constant at 4.0 μM (see Table 5.1). PTP1B (200 ng/ml final assay concentration) was assayed with 20 μM FDP as described under "Experimental Procedures" with 2 mM EDTA and 10 μg/ml catalase present, and 100 μl of the pervanadate solutions were added after 9 min (final concentration of VL + VL_2 = 400 nM) as in Fig. 5.6. The % inhibition relative to a control (addition of 100 μl of water) was determined and plotted vs. the ratio of VL:VL_2 added. The plot shows the average of data from two separate experiments.

DISCUSSION

Enzyme inhibitors are very important as research tools and for their pharmacological potential. Few inhibitors are known for PTPs, and inhibitors that are selective for a particular PTP have been elusive. Two general PTP inhibitors, vanadate and pervanadate, are commonly used, and their insulin mimetic properties have been the subject of much research. In spite of their routine use, a clear understanding of their inhibition mechanisms is still lacking. Furthermore, many researchers often seem to treat vanadate as a simple phosphate analog without considering its rich redox and coordination chemistry, including varied oxidation states and coordination with many reagents. In this study, we set out to elucidate the mechanisms of PTP inhibition by vanadate and pervanadate, and also to establish the reactivity of these inhibitors under various conditions relevant to their general use.

An analysis of vanadate inhibition showed that vanadate behaves as a competitive inhibitor of PTP1B (Fig. 5.3), with a K_i of 0.38 ± 0.02 μM. This is not unexpected, considering the fact that vanadate is structurally a phosphate analog (Fig. 5.2) that mimics the transition state of phosphoryl transfer reactions (15), and the crystal structure of the *Yersinia* PTP complexed with vanadate shows that vanadate occupies the active site within covalent bonding distance (2.5 Å) of the thiol of the catalytic Cys residue (37). While the mechanism of vanadate inhibition may be simple, many researchers appear to neglect the interaction of buffers and assay components with vanadate (34, 35). Vanadate will reversibly coordinate free hydroxyl and thiol groups, and is also chelated by many organic molecules. HEPES is one of the few buffers that does not interact appreciably

with vanadate (34, 35) and was thus chosen for the studies undertaken here. More important, though, is the interaction of vanadate with EDTA. This interaction has been well documented (32–35) but appears to be ignored by many researchers, both those studying PTPs (for examples, see 47–52) and those analyzing phosphotyrosine signaling pathways (for examples, see 53–61). The addition of EDTA in excess over vanadate to a PTP assay containing vanadate will immediately and completely reverse the inhibition due to vanadate (see Fig. 5.1). The presence of EDTA in an assay buffer can dramatically reduce the free concentration of vanadate; for example, 1 mM EDTA can decrease the potency of vanadate for PTP1B over 1000-fold (data not shown). This EDTA effect has important implications for the use of vanadate in cell lysis buffers or immunoprecipitations of phosphotyrosyl proteins in which endogenous cellular PTPs must be inhibited. For example, when 100 μM vanadate is used in the presence of 2 mM EDTA during cell lysis, a much lower level of total phosphotyrosyl proteins is observed as compared to a lysate in which EDTA is excluded (data not shown). The fact that EDTA and vanadate are often included together in cell lysis buffers suggests that in such procedures the endogenous PTPs are not being completely inhibited, which may affect the quality of the results obtained. EGTA, on the other hand, forms a weaker complex with vanadate (34) and may therefore be a practical alternative when divalent cations must be chelated.

Pervanadate (the complexes of vanadate with hydrogen peroxide) was found to be an irreversible inhibitor of PTPs, unlike vanadate whose inhibition is completely reversible. Mass spectrometry showed that when PTP1B was incubated with five molar equivalents of pervanadate, an increase of 48 mass units relative to the untreated enzyme was

observed which was localized to the catalytic Cys (Figs. 5.4 and 5.5). This modification is consistent with the addition of three oxygen atoms. Only the triply-oxidized cysteine (*i.e.*, cysteic acid: $-SO_3H$) was observed. However, it is possible that intermediate oxidized forms exist (*i.e.*, sulfenic and sulfinic acids: –SOH and $-SO_2H$ respectively), and the cysteic acid represents the stable endpoint in the presence of excess pervanadate. No other sites on the enzyme were observed to be oxidized. This specificity may be because the catalytic Cys exists as a thiolate anion (8–11), making it more reactive and susceptible to oxidation; furthermore, pervanadate may have a strong affinity for the active site. Oxidation of the catalytic Cys of PTPs may actually be an *in vivo* mechanism of down-regulating PTP activity. Endogenous reactive oxygen intermediates (ROI), which may act as signaling molecules, have been observed to increase cellular phosphotyrosine levels, and it has been proposed that at least part of this effect may be due to inactivation of PTPs (62). For example, platelet-derived growth factor (PDGF) signal transduction has been reported to cause a transient increase in intracellular hydrogen peroxide concentration which may prolong phosphotyrosine-dependent signaling by oxidizing and inactivating PTPs (63). The fact that the effect of ROI is potentiated by added vanadate (64, 65) would be consistent with the formation in the cell of pervanadate, which is a much more effective inhibitor than ROI alone. This is supported by the *in vitro* kinetic data showing a potentiation of hydrogen peroxide inhibition by vanadate (data not shown).

As was seen with vanadate, assay components can also interact with pervanadate and affect its inhibition. For example, it was observed that pervanadate is rapidly converted to vanadate by thiol reductants such as DTT that are included in PTP assay

buffers to protect the catalytic Cys residue from oxidation. When pervanadate is added to an assay solution containing enzyme and DTT, there appears to be a partitioning of pervanadate between oxidation of the enzyme and oxidation of DTT (Figs. 5.6 and 5.7). Considering that the DTT concentration in Fig. 6 is six orders of magnitude greater than the PTP1B concentration (5 – 15 mM vs. ~5 nM, respectively), pervanadate must have a much higher relative affinity for the enzyme active site. Even though pervanadate is an irreversible inactivator, time-dependent inhibition was not observed. This is explained by the fact that when the pervanadate is added to the enzyme assay, there are at least two competing reactions: oxidation of the enzyme by pervanadate and reduction of pervanadate to vanadate by DTT. No further inactivation of the enzyme can occur after the destruction of the pervanadate, and as these reactions occur faster than the time scale of our measurements, no time-dependent inhibition is observed. It is impossible, therefore, to determine the intrinsic potency of pervanadate under aerobic conditions because of the need to include a thiol reductant to prevent inactivation of the enzyme in air. It is possible, though, to make qualitative conclusions about the relative potencies of the pervanadate species. Vanadate can form several different complexes with hydrogen peroxide (31), with the major species at the concentrations used here being the mono- and di-peroxovanadates, or VL and VL_2 (Fig. 5.2). When the total amount of VL and VL_2 is kept constant, but the ratio of VL:VL_2 is varied (Table 5.1), the amount of inhibition observed increases with increasing VL_2 (Fig. 5.8). Thus, VL_2 appears to be a more potent inactivator of PTP1B than VL. It is possible that VL is not inhibitory at all; however, the

individual potencies of VL and VL_2 cannot be assessed in this experiment as both complexes are always present together.

The different modes of vanadate and pervanadate inhibition provide a possible rationale for the differences observed in their *in vivo* effects. Pervanadate has been reported to be more potent than vanadate both in increasing the level of phosphotyrosine-containing proteins in intact cells (23–25) and in insulin mimesis (21, 22). Vanadate can easily enter cells, after which it is reported to be reduced to vanadium (IV) (66–68). While the inhibition of PTPs by vanadium (IV) has not been determined, vanadium (IV) is likely to be a much less potent inhibitor than vanadium (V) as it is not a phosphate analog nor a PTP transition state mimic. In fact, kinetic studies suggest that vanadium (IV) is a much poorer inhibitor of the (Na,K)-ATPase than vanadium (V) (66, 68). Since any inhibition of intracellular PTPs by vanadate that does occur is reversible, it will be eliminated as the vanadate dissociates from the PTPs and is reduced. Pervanadate, by contrast, may partition in the cell between conversion to vanadate and oxidative inactivation of PTPs, analogous to that observed in the *in vitro* kinetic assays. Irreversible inactivation would result in a lower level of total intracellular PTP activity than in the case with vanadate, resulting in a greater overall level of protein tyrosine phosphorylation. Clearly an understanding of the inhibition mechanisms of vanadate and pervanadate and their potential reactions and interactions are key to explaining their *in vivo* and *in vitro* effects and must be taken into consideration when these inhibitors are used.

ACKNOWLEDGMENTS

We thank Dr. R. L. Erikson for providing the full-length PTP1B cDNA clone, Dr. R. Zamboni for synthesizing the FDP substrate, and Dr. A. Tracey for sharing unpublished NMR data and for helpful suggestions. We are also grateful to J. Davidow and A. Govindarajan for their technical assistance. Special thanks to the members of the Merck Frosst phosphatase group, especially Dr. Z. Huang, and to Dr. A. English and her laboratory at Concordia, for their support and critical insights. G. H. was supported by a Natural Sciences and Engineering Research Council of Canada 1967 Science and Engineering Studentship.

REFERENCES

1. Ullrich, A., and Schlessinger, J. (1990) *Cell* **61,** 203-212
2. Glenney, J. R., Jr. (1992) *Biochim. Biophys. Acta* **1134,** 113-127
3. Brautigan, D. L. (1992) *Biochim. Biophys. Acta* **1114,** 63-77
4. Walton, K. M., and Dixon, J. E. (1993) *Ann. Rev. Biochem.* **62,** 101-120
5. Tonks, N. K., Diltz, C. D., and Fischer, E. H. (1988) *J. Biol. Chem.* **263,** 6722-6730
6. Tonks, N. K., Diltz, C. D., and Fischer, E. H. (1988) *J. Biol. Chem.* **263,** 6731-6737
7. Charbonneau, H., Tonks, N. K., Walsh, K. A., and Fischer, E. H. (1988) *Proc. Natl. Acad. Sci.* **85,** 7182-7186
8. Stone, R. L., and Dixon, J. E. (1994) *J. Biol. Chem.* **269,** 31323-31326
9. Zhang, Z.-Y., and Dixon, J. E. (1994) *Adv. Enzymol.* **68,** 1-36
10. Barford, D. (1995) *Curr. Opin. Struct. Biol.* **5,** 728-734
11. Barford, D., Jia, Z., and Tonks, N. K. (1995) *Nature Struct. Biol.* **2,** 1043-1053
12. Guan, K. L., and Dixon, J. E. (1991) *J. Biol. Chem.* **266,** 17026-17030

13. Cho, H., Krishnaraj, R., Kitas, E., Walsh, C. T., and Anderson, K. S. (1992) *J. Am. Chem. Soc.* **114,** 7296-7298

14. Swarup, G., Cohen, S., and Garbers, D. L. (1982) *Biochem. Biophys. Res. Comm.* **107,** 1104-1109

15. Gresser, M. J., and Tracey, A. S. (1990) in *Vanadium in Biological Systems* (Chasteen, N. D., ed), pp. 63-79, Kluwer Academic Publishers, Netherlands

16. Shechter, Y., Meyerovitch, J., Farfel, Z., Sack, J., Bruck, R., Bar-Meir, S., Amir, S., Degani, H. and Karlish, S. J. D. (1990) in *Vanadium in Biological Systems* (Chasteen, N. D., ed), pp. 129-142, Kluwer Academic Publishers, Netherlands

17. Heyliger, C. E., Tahiliani, A. G., and McNeil, J. H. (1985) *Science* **227,** 1474-1477

18. Goldfine, A. B., Simonson, D. C., Folli, F., Patti, M.-E., and Kahn, C. R. (1995) *J. Clin. Endo. Met.* **80,** 3311-3320

19. Shechter, Y. (1990) *Diabetes* **39,** 1-5

20. Kadota, S., Fantus, I. G., Deragon, G., Guyda, H. J., Hersh, B., and Posner, B. I. (1987) *Biochem. Biophys. Res. Commun.* **147,** 259-266

21. Fantus, I. G., Kadota, S., Deragon, G., Foster, B. and Posner, B. I. (1989) *Biochemistry* **28,** 8864-8871

22. Leighton, B., Cooper, G. J. S., DaCosta, C., and Foot, E. A. (1991) *Biochem. J.* **276,** 289-292

23. Heffetz, D., Bushkin, I., Dror, R., and Zick, Y. (1990) *J. Biol. Chem.* **265,** 2896-2902

24. Zick, Y., and Eisenberg, R. S. (1990) *Biochemistry* **29,** 10240-10245

25. Secrist, J. P., Burns, L. A., Karnitz, L., Koretzky, G. A., and Abraham, R. T. (1993) *J. Biol. Chem.* **268,** 5886-5893

26. Posner, B. I., Faure, R., Burgess, J. W., Bevan, A. P., Lachance, D., Zhang-Sun, G., Fantus, I. G., Ng, J. B., Hall, D. A., Soo Lum, B., and Shaver, A. (1994) *J. Biol. Chem.* **269,** 4596-4604

27. Bevan, A. P., Drake, P. G., Yale, J.-F., Shaver, A., and Posner, B. I. (1995) *Mol. Cell. Biochem.* **153,** 49-58

28. Yale, J.-F., Vigeant, C., Nardolillo, C., Chu, Q., Yu, J.-Z., Shaver, A., and Posner, B. I. (1995) *Mol. Cell. Biochem.* **153,** 181-190

29. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

30. Tracey, A. S., Gresser, M. J., and Liu, S. (1988) *J. Am. Chem. Soc.* **110,** 5869-5874

31. Jaswal, J. S., and Tracey, A. S. (1991) *Inorg. Chem.* **30,** 3718-3722

32. Przyborowski, L., Schwarzenbach, G., and Zimmermann, T. (1965) *Helv. Chim. Acta* **48,** 1556-1565

33. Kustin, K., and Toppen, D. L. (1973) *J. Am. Chem. Soc.* **95,** 3564-3563

34. Crans, D. C., Bunch, R. L., and Theisen, L. A. (1989) *J. Am. Chem. Soc.* **111,** 7597-7607

35. Crans, D. C. (1994) *Comments Inorg. Chem.* **16,** 1-33

36. Harrison, A. T., and Howarth, O. W. (1985) *J. Chem. Soc. Dalton Trans.* **1985,** 1173-1177

37. Denu, J. M., Lohse, D. L., Vijayalakshmi, J., Saper, M. A., and Dixon, J. E. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93,** 2493-2498

38. Butler, A., Clague, M. J., and Meister, G. E. (1994) *Chem. Rev.* **94,** 625-638

39. Shaver, A., Ng, J. B., Hall, D. A., Soo Lum, B., and Posner, B. I. (1993) *Inorg. Chem.* **32,** 3109-3113

40. Biemann, K. (1988) *Biomed. Env. Mass Spectrom.* **16,** 99-111

41. Biemann, K. (1990) in *Biological Mass Spectrometry* (Burlingame, A. and McCloskey, J., eds), pp. 179-196, Elsevier, Amsterdam

42. Burlet, O., Yang, C.-Y., and Gaskell, S. J. (1992) *J. Am. Soc. Mass Spectrom.* **3,** 337-344

43. Burlet, O., Yang, C.-Y., Guyton, J. R., and Gaskell, S. J. (1995) *J. Am. Soc. Mass Spectrom.* **6,** 242-247

44. Trotta, P. P., Pinkus, L. M., and Meister, A. (1974) *J. Biol. Chem.* **249,** 1915-1921

45. Misra, H. P. (1974) *J. Biol. Chem.* **249,** 2151-2155

46. Costa, M., Pecci, L., Pensa, B., and Cannella, C. (1977) *Biochem. Biophys. Res. Comm.* **78,** 596-603

47. Pot, D. A., Woodford, T. A., Remboutsika, E., Haun, R. S., and Dixon, J. E. (1991) *J. Biol. Chem.* **266,** 19688-19696

48. Cho, H., Raemer, S., Itho, M., Winkler, D. G., Kitas, E., Bannwarth, W., Burn, P., Saito, H., and Walsh, C. T. (1991) *Biochemistry* **30,** 6210-6216

49. Liao, K., Hoffman, R. D., and Lane, M. D. (1991) *J. Biol. Chem.* **266,** 6544-6553

50. Itoh, M., Streuli, M., Krueger, N. X., and Saito, H. (1992) *J. Biol. Chem.* **267,** 12356-12363

51. Mei, L., and Huganir, R. L. (1991) *J. Biol. Chem.* **266,** 16063-16072

52. Zhao, Z., Bouchard, P., Diltz, C. D., Shen, S.-H., and Fischer, E. (1993) *J. Biol. Chem.* **268,** 2816-2820

53. Milarski, K. L., and Saltiel, A. R. (1994) *J. Biol. Chem.* **269,** 21239-21243

54. Zhao, Z., Tan, Z., Wright, J. H., Diltz, C. D., Shen, S.-H., Krebs, E. G., and Fischer, E. H. (1995) *J. Biol. Chem.* **270,** 11765-11769

55. Milarski, K. L., Lazar, D. F., Wiese, R. J., and Saltiel, A. R. (1995) *Biochem. J.* **308,** 579-583

56. Ezumi, Y., Takayama, H., and Okuma, M. (1995) *J. Biol. Chem.* **270,** 11927-11934

57. Yamauchi, K., and Pessin, J. E. (1995) *J. Biol. Chem.* **270,** 14871-14874

58. Kohn, A. D., Kovacina, K. S., and Roth, R. A. (1995) *EMBO J.* **14,** 4288-4295

59. Yetter, A., Uddin, S., Krolewski, J. J., Jiao, H., Yi, T., and Platanias, L. C. (1995) *J. Biol. Chem.* **270,** 18179-18182

60. Marengere, L. E. M., Waterhouse, P., Duncan, G. S., Mittrucker, H.-W., Feng, G.-S., and Mak, T. W. (1996) *Science* **272,** 1170-1173

61. Plas, D. R., Johnson, R., Pingel, J. T., Matthews, R. J., Dalton, M., Roy, G., Chan, A. C., and Thomas, M. L. (1996) *Science* **272,** 1173-1176

62. Fialkow, L., Chan, C. K., Grinstein, S., and Downey, G. P. (1993) *J. Biol. Chem.* **268,** 17131-17137

63. Sundaresan, M., Yu, Z.-X., Ferrans, V. J., Irani, K., and Finkel, T. (1995) *Science* **270,** 296-299

64. Schieven, G. L., Kirihara, J. M., Myers, D. E., Ledbetter, J. A., and Uckun, F. M. (1993) *Blood* **82,** 1212-1220

65. Zor, U., Ferber, E., Gergely, P., Szucs, K., Dombradi, V., and Goldman, R. (1993) *Biochem. J.* **295,** 879-888

66. Cantley, L. C., and Aisen, P. (1979) *J. Biol. Chem.* **254,** 1781-1784

67. Degani, H., Grochin, M., Karlish, S. J. D., and Shechter, Y. (1981) *Biochemistry* **20,** 5795-5799

68. Wilsky, G. R., White, D. A., and McCabe, B. C. (1984) *J. Biol. Chem.* **259,** 13273-13281

CHAPTER 5 – ADDENDUM

Effect of Assay Buffer on Vanadate Inhibition

SUMMARY

Initial studies with pervanadate suggested that it was a slow time-dependent inhibitor of PTP1B. However, it was subsequently discovered that DTT in the assay buffer was reducing pervanadate to vanadate before enzyme addition, and therefore the observed inhibition was due to vanadate. This time-dependent inhibition, though, was not consistent with that described for vanadate in Chapter 5 (1). As pervanadate was assayed in imidazole buffer and vanadate in HEPES buffer, it was possible that the buffer was responsible for the difference. In fact, when tested in imidazole buffer, vanadate inhibited PTP1B in an irreversible time-dependent manner, as opposed to the reversible time-independent inhibition observed in HEPES buffer. Catalase eliminated the time-dependent inhibition, suggesting that the inhibition was due to generation of hydrogen peroxide by the buffer system. Hydrogen peroxide was found to potentiate the inhibition of PTP1B by vanadate, consistent with this model. These results further demonstrate the need to evaluate carefully the buffer conditions for enzyme assays.

EXPERIMENTAL PROCEDURES

PTP Kinetic Assays – Flag-PTP1B was expressed and purified as previously described (1). Activity was measured using 10 μM FDP as substrate, and appearance of FMP product was followed by fluorescence using a Perkin-Elmer luminescence spectrometer (LS50B)

with excitation at 450 nm and emission at 515 nm. The reaction volume was 1.2 ml, consisting of ~125 ng/ml PTP1B in 50 mM imidazole pH 7.2 or 100 mM HEPES pH 7.3, 10 mM DTT, and 10 μg/ml BSA (except for Fig. 5.9, 100 μg/ml BSA). For Fig. 5.13, 25 mM HEPES pH 7.3, 5 mM DTT, and 10 μg/ml BSA was used.

RESULTS

Studies of Pervanadate Inhibition in Imidazole Buffer – Initial studies showed the inhibition of PTP1B by pervanadate was time- and concentration-dependent (Fig. 5.9a), and the time courses of inactivation could be fit to the first-order rate equation. Several experiments suggested that the inhibition by pervanadate was irreversible. First, pre-incubation of PTP1B with pervanadate on ice for 30 min, followed by dilution into assay buffer plus FDP such that the final concentration of pervanadate was 100 nM, resulted in no recovery of activity. Second, a 5-fold dilution of an assay mixture containing PTP1B, FDP, assay buffer, and 250 nM pervanadate after 10 or 30 min incubation did not result in any recovery of activity, but did significantly slow further inhibition. This inhibition is in contrast to vanadate, which we have previously shown (1) to be a completely reversible, time-independent inhibitor of PTPs.

From the curve fits in Fig. 5.9a, the k_{obs} for inactivation was calculated for each pervanadate concentration. The replot of the k_{obs} values *vs.* pervanadate concentration is hyperbolic (Fig. 5.9b), suggesting that there is a reversible association of pervanadate with the enzyme and that the rate-limiting step is the irreversible inactivation step, as detailed in the scheme below:

$$E + I \underset{}{\overset{K_i}{\rightleftharpoons}} EI \xrightarrow{k_{inact}} E^*$$

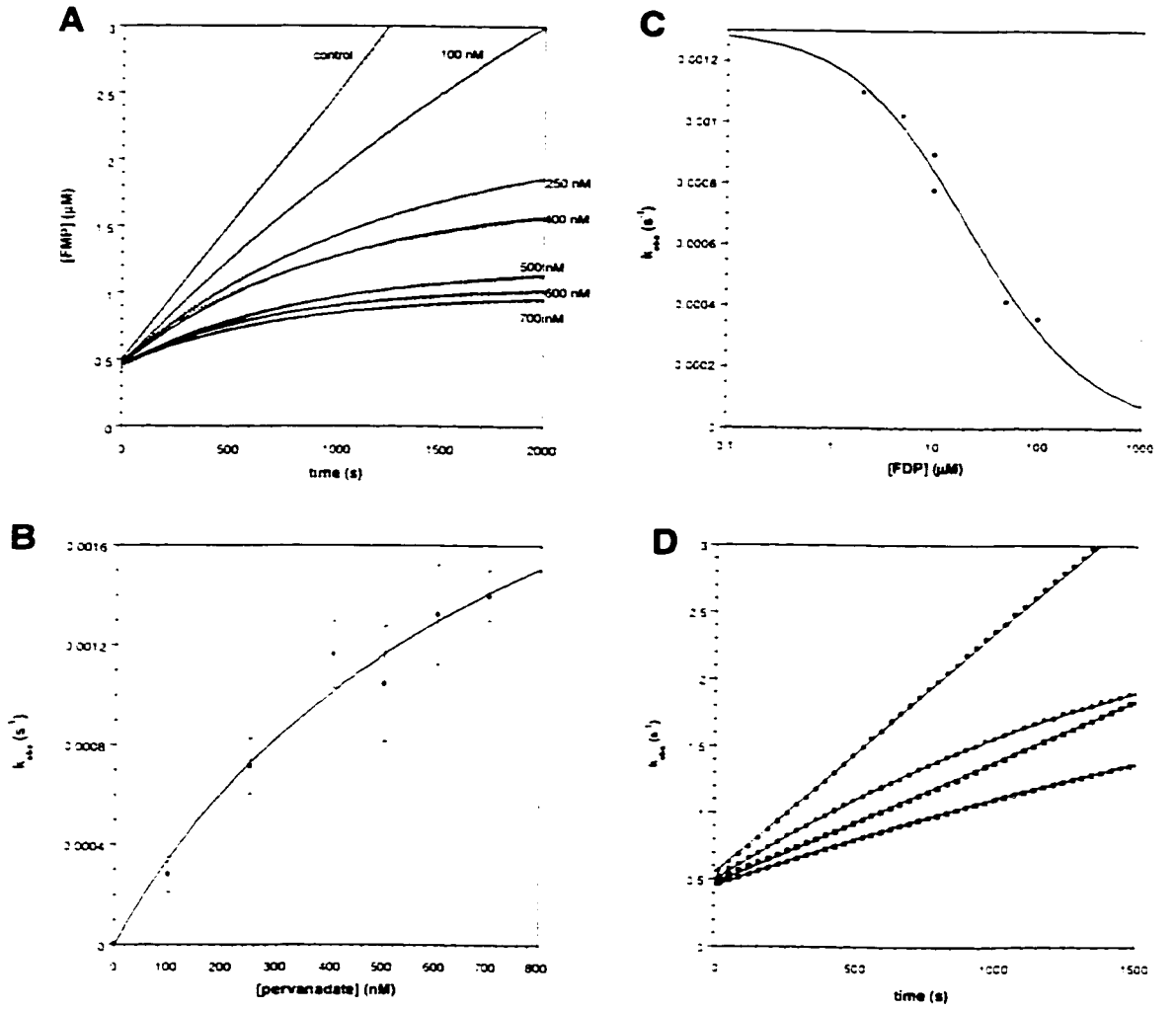


Fig. 5.9. **Kinetics of pervanadate inhibition of PTP1B.** PTP1B (125 ng/ml final assay concentration) was assayed with 10 μM FDP as described under "Experimental Procedures." (A) Time-dependent inhibition by pervanadate. PTP1B was assayed with varying amounts of pervanadate as indicated, with assays initiated by the addition of enzyme. Time courses of inactivation were fit to the first-order rate equation. (B) Replot of k_{obs} for inactivation *vs.* pervanadate concentration as calculated from time courses in (A). (C) Effect of FDP concentration on the rate of pervanadate inactivation. PTP1B was assayed as described above with varying amounts of FDP and 250 nM pervanadate. The time courses of inactivation were fit to the first-order rate equation, and the calculated k_{obs} values were replotted against the FDP concentration and fit to a standard four-parameter fit. (D) Effect of F_2Pmp hexapeptide on the rate of pervanadate inactivation. PTP1B was assayed as described above without any inhibitors (closed circles), with 250 nM pervanadate (open circles), with 83 nM hexapeptide (closed squares), and with both pervanadate and hexapeptide (open squares). Time courses of inactivation were fit to the first-order rate equation as above.

From the double-reciprocal plot of $1/k_{obs}$ *vs.* 1/[pervanadate], the values of K_i and k_{inact} can be estimated as approximately 1 μM and 0.007 s^{-1}, respectively (data not shown). The observed rate of inactivation depends on the concentration of substrate, as increasing the concentration of FDP caused a decrease in the k_{obs} (Fig. 5.9c). The concentration of FDP required to reduce the k_{obs} by 50% was 22 ± 12 μM, which is approximately equal to the K_m for FDP (~10 μM). As well, the onset of inhibition was delayed by the presence of a competitive peptide inhibitor containing the non-hydrolyzable phosphotyrosyl analogue F_2Pmp, DADE(F_2Pmp)L (2) (Fig. 5.9d), reducing the k_{obs} at the IC_{50} concentration of 83 nM by ~50% (0.00057 s^{-1} for 250 nM pervanadate alone *vs.* 0.00027 s^{-1} in the presence of the F_2Pmp hexapeptide). Both these results are consistent with the pervanadate inhibition being active-site directed.

Inhibition by Vanadate in Imidazole Buffer – The fact that DTT reduces pervanadate to vanadate meant that the time-dependent inhibition thought to be due to pervanadate was actually due to vanadate. However, our previous studies with vanadate (1) indicated that the inhibition was completely reversible and not time-dependent. These studies, though, were carried out in HEPES buffer while the pervanadate studies here were performed in imidazole buffer. Thus, the inhibition of vanadate in imidazole buffer was examined, showing the same time-dependent inhibition as observed with pervanadate (Fig. 5.10a).

We next asked how vanadate could be capable of irreversible time-dependent inactivation of PTP1B. One possibility was that hydrogen peroxide was being generated in the assay mixture that would combine with the vanadate to form pervanadate. To test this, the inhibition by vanadate was assayed in the presence of catalase. As shown in Fig.

5.10b. catalase completely eliminated the time-dependent inhibition observed with vanadate, consistent with this proposal. Exactly the same result was observed with pervanadate added to the assay mixture before enzyme (data not shown).

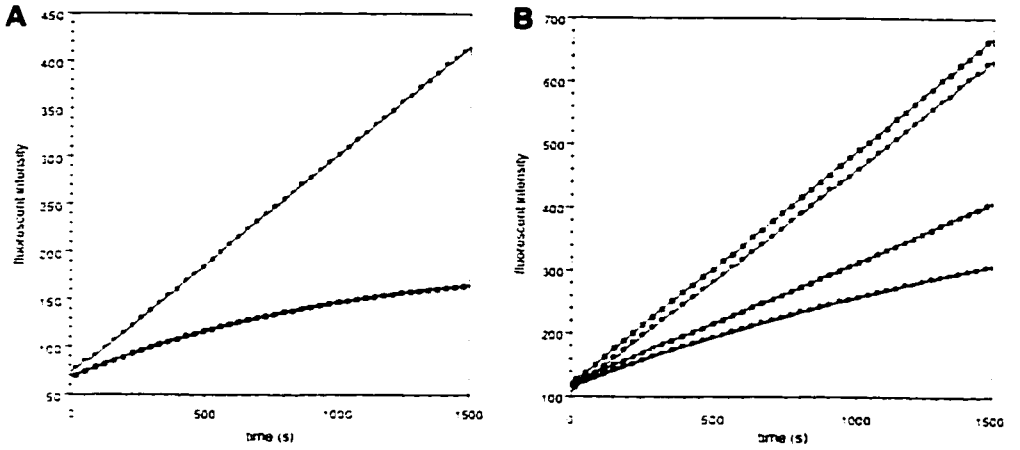


Fig. 5.10. **Inhibition of PTP1B by vanadate in imidazole buffer, and effect of catalase.** (A) Time courses showing activity of PTP1B in the presence of 500 nM vanadate in 100 mM HEPES pH 7.3 (circles) or 50 mM imidazole pH 7.2 (squares). (B) Effect of catalase on vanadate inhibition. PTP1B was assayed in 50 mM imidazole pH 7.2 in the absence (circles) or presence (squares) of 10 µg/ml catalase. Closed symbols, no vanadate; open symbols, 500 nM vanadate.

Hydrogen peroxide can be produced following metal-catalyzed oxidation of DTT (Fig. 5.11) (3-5). Imidazole coordinates transition metals strongly and is therefore contaminated with many trace metal ions, which is the likely source of metals that would catalyze the series of reactions in Fig. 5.11. The addition of EDTA would be predicted to chelate these metal ions and therefore eliminate the generation of hydrogen peroxide. However, EDTA will also chelate vanadate (6-9) and therefore this could not be tested. The reaction series suggest that the addition of superoxide dismutase (SOD) or metal ions such as Cu^{++} should enhance formation of hydrogen peroxide and therefore accelerate the

vanadate time-dependent inhibition. This was in fact observed (data not shown), consistent with the proposed mechanism.

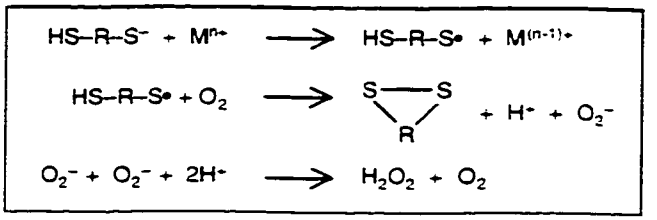


Fig. 5.11. **Mechanism for generation of hydrogen peroxide by thiols and transition metals.** In the first reaction, a transition metal (M^{n+}) initiates the thiol oxidation. O_2^- is subsequently generated from molecular oxygen and disproportionates to hydrogen peroxide. Adapted from Misra (4).

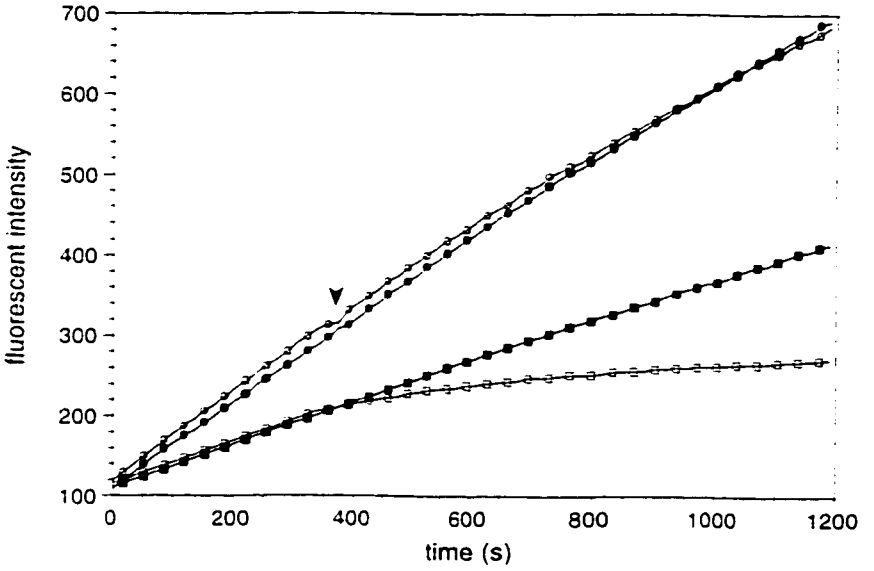


Fig. 5.12. **Potentiation of vanadate inhibition by addition of hydrogen peroxide.** PTP1B was assayed in 25 mM HEPES pH 7.3 with (squares) and without (circles) 500 nM vanadate added at time zero. At the point indicated by the arrow, 100 μM H_2O_2 was added to the time courses represented by the open symbols.

Another prediction from this mechanism is that exogenous hydrogen peroxide should synergize with vanadate to inhibit PTP1B. Hydrogen peroxide alone is a poor inhibitor of PTP1B, requiring at least 200 μM to observe any appreciable inhibition in 25 mM HEPES and 5 mM DTT. However, when hydrogen peroxide is added to a kinetic assay in which a low concentration of vanadate is present, a dramatic inhibition of PTP1B activity is observed (Fig. 5.12), consistent again with vanadate and hydrogen peroxide combining *in situ* to form pervanadate.

DISCUSSION

We have previously described the inhibition of PTPs by vanadate and pervanadate in some detail (1). The purpose of this study was to explore further the effect of buffer on the inhibition of vanadate and to make clear the importance of assessing the inhibition under a variety of conditions. In regards to vanadate and pervanadate, initially their modes of inhibition seemed quite simple: vanadate was a reversible, competitive inhibitor while pervanadate was a slow time-dependent inhibitor that inactivated the enzyme by oxidizing the catalytic Cys (Fig. 5.9). However, it was subsequently discovered that pervanadate is rapidly reduced to vanadate by thiol reducing agents such as DTT in the assay buffer, and the time-dependent inhibition observed was an artifact of the buffer system.

A more detailed examination of the effect of buffer on vanadate inhibition was undertaken in this study. In HEPES, vanadate was a reversible time-independent inhibitor of PTP1B, while in imidazole slow irreversible time-dependent inhibition was observed

(Fig. 5.10a). This difference was due to the generation of hydrogen peroxide in imidazole buffer that could combine with vanadate to form pervanadate *in situ*, as catalase eliminated the time-dependent inhibition (Fig. 5.10b). Addition of hydrogen peroxide potentiated vanadate inhibition (Fig. 5.12), consistent with this mechanism. The relatively slow time-dependent inhibition observed is a combination of the rate of hydrogen peroxide generation (and by extension the rate of pervanadate formation), and the competition between pervanadate oxidation of the catalytic Cys of PTP1B and destruction by DTT (as described in Fig. 5.7).

Vanadate–hydrogen peroxide inhibition may provide a model for *in vivo* regulation of PTP activity. Tyrosine phosphorylation-dependent signaling pathways are known to be regulated by modulation of the cellular redox state (10), and there is evidence to suggest that this is through oxidation of PTPs, presumably at the catalytic Cys (11-14). Many stimuli lead to increased hydrogen peroxide levels in the cell (10-14), which in our *in vitro* assay system can be thought to be mimicked by the reactions described in Fig. 5.12. To prevent non-specific oxidation of cellular components by hydrogen peroxide, it is likely that cells have a mechanism of capturing this oxidizing potential and/or directing it to the appropriate target. In our system, vanadate can be thought as functioning as a hydrogen peroxide delivery system to the catalytic Cys of PTPs. Similarly, cells may use a functional analogue (or perhaps even vanadate).

In summary, this study reinforces the need to examine an enzyme inhibitor under a number of conditions before ascribing an inhibition mechanism to it. In particular, order of inhibitor addition (before or after enzyme), choice of buffer, and effect of other

components of the assay mixture need to be assessed. In regards to pervanadate, the buffer effects (*i.e.*, reduction to vanadate by DTT and generation of hydrogen peroxide in imidazole buffer) initially lead to incorrect conclusions. However, the understanding of the buffer effects may in fact shed some light on the *in vivo* redox control of signaling pathways by providing a model for the kind of redox control of PTP activity that may occur in the cell.

REFERENCES

1. Huyer, G., Liu, S., Kelly, J., Moffat, J., Payette, P., Kennedy, B., Tsaprailis, G., Gresser, M. J., and Ramachandran, C. (1997) *J. Biol. Chem.* **272,** 843-851

2. Burke, T. R., Jr., Kole, H. K., and Roller, P. P. (1994) *Biochem. Biophys. Res. Comm.* **204,** 129-134

3. Trotta, P. P., Pinkus, L. M., and Meister, A. (1974) *J. Biol. Chem.* **249,** 1915-1921

4. Misra, H. P. (1974) *J. Biol. Chem.* **249,** 2151-2155

5. Costa, M., Pecci, L., Pensa, B., and Cannella, C. (1977) *Biochem. Biophys. Res. Comm.* **78,** 596-603

6. Przyborowski, L., Schwarzenbach, G., and Zimmermann, T. (1965) *Helv. Chim. Acta* **48,** 1556-1565

7. Kustin, K., and Toppen, D. L. (1973) *J. Am. Chem. Soc.* **95,** 3564-3563

8. Crans, D. C., Bunch, R. L., and Theisen, L. A. (1989) *J. Am. Chem. Soc.* **111,** 7597-7607

9. Crans, D. C. (1994) *Comments Inorg. Chem.* **16,** 1-33

10. Monteiro, H. P., and Stern, A. (1996) *Free Radical Biol. Med.* **21,** 323-333

11. Fialkow, L., Chan, C. K., Grinstein, S., and Downey, G. P. (1993) *J. Biol. Chem.* **268,** 17131-17137

12. Sundaresan, M., Yu, Z.-X., Ferrans, V. J., Irani, K., and Finkel, T. (1995) *Science* **270,** 296-299

13. Brumell, J. H., Burkhardt, A. L., Bolen, J. B., and Grinstein, S. (1996) *J. Biol. Chem.* **271,** 1455-1461

14. Bae, Y. S., Kang, S. W., Seo, M. S., Baines, I. C., Tekle, E., Chock, P. B., and Rhee, S. G. (1997) *J. Biol. Chem.* **272,** 217-221

CHAPTER 6

Peptide Substrate Specificity Studies

With the Protein Tyrosine Phosphatases CD45, LAR, and PTPß

SUMMARY

The protein tyrosine phosphatases (PTPs) CD45, LAR, and PTPß are all transmembrane PTPs with one (PTPß) or two (CD45 and LAR) intracellular PTP domains. CD45 is known to dephosphorylate the tyrosine kinase $p56^{lck}$ at Tyr_{505} *in vivo*, but little is known about the substrates of the other two enzymes. In order to understand better the substrate specificity of these PTPs, the enzymes were characterized for their K_m, V_{max}, and catalytic efficiency towards a number of peptide substrates. Starting with a sequence surrounding Tyr_{505} of $p56^{lck}$ [TEGQ(pY)QPQP], an alanine scan was performed in which each residue is substituted one at a time with Ala, to determine the importance of the individual positions to enzyme recognition. In general, little variation in the catalytic efficiencies of CD45 or PTPß was observed with all of the peptides tested. Kinetic parameters could not be determined for LAR as the K_m was found to be too high (>1 mM) to be measured in the kinetic assay. The analysis of some of the peptide substrates was complicated by apparent substrate inhibition, which was in fact due to contamination by a potent trace inhibitor.

INTRODUCTION

Protein tyrosine phosphatases (PTPs)[1] are important regulators of many cellular signaling processes (1-4). The first PTP identified was purified 10 years ago from human placenta and termed PTP1B (5, 6). The sequence of this protein showed high homology to CD45 (7), a type I transmembrane protein of previously unknown function identified as a surface antigen on T cells and subsequently shown to be a PTP (8). A large family of PTPs (>75 members) has since been identified. These enzymes fall into two broad classes: transmembrane and cytosolic. PTP1B is representative of the cytosolic enzymes, while CD45 is the archetypal transmembrane PTP.

CD45, also known as the leukocyte common antigen (LCA), is a 180 – 220 kDa transmembrane glycoprotein with a variable extracellular domain. It is predicted to interact with an extracellular ligand, although none has been identified. The cytoplasmic portion consists of two tandem PTP domains, of which only the membrane-proximal (D1) domain appears to be active (9, 10). There is one report of catalytic activity associated with D2 of CD45 (10a), but this result has not been reproduced. The role of D2 is unclear but may be involved in regulation or association with substrates.

CD45 is a highly-abundant protein in T-cells, comprising as much as 10% of the membrane proteins. It plays a critical role in the initiation of TCR signaling, as evident from a T-cell lymphoma cell line, J45.01, which expresses ~8% of the normal levels of CD45 (11). Stimulation of the TCR of these cells by presentation of antigen or by crosslinking with anti-CD3 antibody (a component of the TCR) does not lead to a cellular

[1] Abbreviations: PTP, protein tyrosine phosphatase; TCR, T-cell receptor; GST, glutathione *S*-transferase; HEPES, N-(hydroxyethyl) piperazine-N'-(2-ethane sulfonic acid); DTT, dithiothreitol; BSA, bovine serum albumin; TFA, trifluoroacetic acid.

response, as measured by intracellular Ca^{2+} release, production of IL-2, or mitogenesis. The role of CD45 in the initiation of TCR signaling is now fairly well understood (12). CD45 is peripherally associated with the TCR complex, and in response to TCR stimulation CD45 dephosphorylates a number of intracellular targets, including the Src-family protein tyrosine kinases $p56^{lck}$ and $p59^{fyn}$. Dephosphorylation of these enzymes in turn leads to their activation, resulting in amplification and propagation of the initial signal.

LAR, the leucocyte common antigen related PTP (13), shares a great deal of homology with CD45. It too is a transmembrane PTP with two tandem intracellular PTP domains, of which only D1 (the membrane-proximal domain) appears to be active. The extracellular region resembles that of cellular adhesion molecules in that it is composed of three immunoglobulin (Ig)-like and eight fibronectin type III (FN III)-like domains. Human LAR is expressed as a ~235 kDa intracellular precursor that is processed to 150 kDa and 85 kDa subunits that are non-covalently associated on the cell surface (14). The smaller domain contains the cytoplasmic and transmembrane portions, while the larger subunit contains the extracellular portion. The larger subunit is shed under certain conditions such as high-density cell growth, and shedding may play a role in regulation of LAR activity (14, 15). LAR appears to localize to focal adhesions and may play a role in focal adhesion turnover (16). Although LAR is expressed in many tissues (17), the only apparent phenotype in an LAR knock-out mouse is a defect in mammary gland development (18). Actual substrates for LAR and involvement in particular signaling pathways are unclear, though.

PTPß is one of the few examples of a mammalian transmembrane PTP that has only one intracellular catalytic domain (19). The PTPß cDNA encodes a protein of 1975 amino acids that contains a large extracellular domain (~1600 amino acids) consisting of 16 FN-III-like repeats. Little if anything, though, is known about the *in vivo* functions of this enzyme.

In order to gain a better understanding of the catalytic activity and substrate specificities of these enzymes, kinetic experiments were undertaken with synthetic phosphotyrosine (pY)-containing peptide substrates. The sequence surrounding the *in vivo* pY target of CD45 in $p56^{lck}$, namely TEGQ(pY)QPQP, was chosen. To assess the importance of each amino acid for recognition by the enzymes, an alanine scan was performed in which each residue was replaced, one at a time, with alanine. The utility of such an alanine scan with PTPs was demonstrated by Zhang *et al.* (20) with PTP1B and the *Yersinia* PTP. The studies reported here demonstrate that although CD45 and PTPß have different levels of activity towards the substrates, overall they display only subtle differences in catalytic efficiency towards the Ala-substituted peptides. Apparent substrate inhibition was observed with some of the peptide substrates that was in fact due to a potent trace inhibitor, demonstrating the need to assess the purity of peptide substrates.

EXPERIMENTAL PROCEDURES

Materials – Peptides were purchased from California Peptide Research (Napa, CA). All peptides had the expected mass and were estimated to be >95% pure by HPLC. All chemicals were of reagent grade from Sigma.

Purification of Proteins – All PTPs used in this study were expressed in and purified from *E. coli*. The 708 amino acid cytosolic portion of CD45 containing both PTP domains (starting with the last residue of the transmembrane domain and continuing through to the natural C-terminal end) was cloned as a GST fusion in the vector pGEX-2T (Pharmacia). Subcloning resulted in codons for an additional six amino acids being added to the 3' end of the sequence. CD45 was expressed and purified as described (21), and the GST portion was removed by thrombin digestion essentially as described (22).

The cytosolic domain of LAR containing both PTP domains (amino acids 1264 – 1881), which included the natural stop codon, was cloned into the expression vector pT7-7 (23). The clone was a generous gift of D. Pot (Purdue University). LAR was expressed and purified by chromatography as described (23). Before use, LAR was desalted using a Pharmacia desalting column in 20 mM imidazole pH 7.2, 0.2% ß-mercaptoethanol.

The cytosolic domain of PTPß (amino acids 1621 – 1975, consisting of the entire cytosolic portion starting with the first residue after the transmembrane domain) was cloned as a GST fusion protein in the vector pGEX-2T, with a ten-residue influenza hemagglutinin (HA) epitope (followed by a Pro-Gly-Ala-Ser spacer) inserted between the GST portion and the PTPß sequence. The construct was kindly provided by F. Jirik (U.B.C.). The protein was expressed, purified, and thrombin-cleaved as described above for CD45.

Assays for PTP Activity – For activity assays, a buffer consisting of 100 mM HEPES pH 7.3, 10 mM DTT, and 0.1 mg/ml BSA was used. Appearance of dephosphorylated peptide product was followed spectroscopically as described (24). For CD45 and LAR,

peptide substrate concentrations ranging from 100 μM to 1.5 mM were used, and product formation was monitored by following the increase of absorbance at 282 nm in 1-ml cuvettes using a Cary 3 spectrophotometer. At this pH, $\Delta\varepsilon_{282} = 830\ M^{-1}$ (24). For PTPß, substrate concentrations ranging from 10 – 125 μM were used, and product formation was monitored by fluorescence (excitation = 280 nm, emission = 305 nm, slit widths = 5 nm) using a Perkin-Elmer luminescence spectrometer (LS50B). Under these conditions, Δfluorescence = 1.97 μM^{-1}. The concentrations of enzyme assayed were approximately 300 ng/ml for CD45, 670 ng/ml for LAR, and 435 ng/ml for PTPß.

HPLC Purification of Peptide – 0.5 mg of LCK Ala4 peptide (250 μl of a 2 mg/ml solution in 5% TFA) was injected onto a C_{18} Aquapore column connected to a Waters HPLC system. A gradient of 0 – 30% CH_3CN containing 0.07% TFA at 0.4 ml/min over 18 min was used to elute the peptide, and a single fraction was collected as the peptide was eluted (as determined by monitoring the absorbance at 218 nm: retention time ~12-13 min). Three injections were performed, and the peptide fractions were pooled, evaporated on a Speed Vac, and resuspended in water for use in the kinetic analysis.

RESULTS

Kinetic Studies with CD45 – The kinetic parameters K_m and V_{max} were determined by assaying CD45 with varying concentrations of peptide substrate and determining the initial rates to generate a plot of rate *vs.* substrate concentration. In general, CD45 displayed simple Michaelis-Menten saturation kinetics, and the data were fit to the equation $v = V_{max}[S]/(K_m + [S])$ (Fig. 6.1a). However, for several peptides, namely the peptides

substituted with Ala at positions 4, 6, or 8, inhibition of enzyme activity was observed at higher concentrations of substrate (Fig. 6.1b), and kinetic parameters could therefore not be determined.

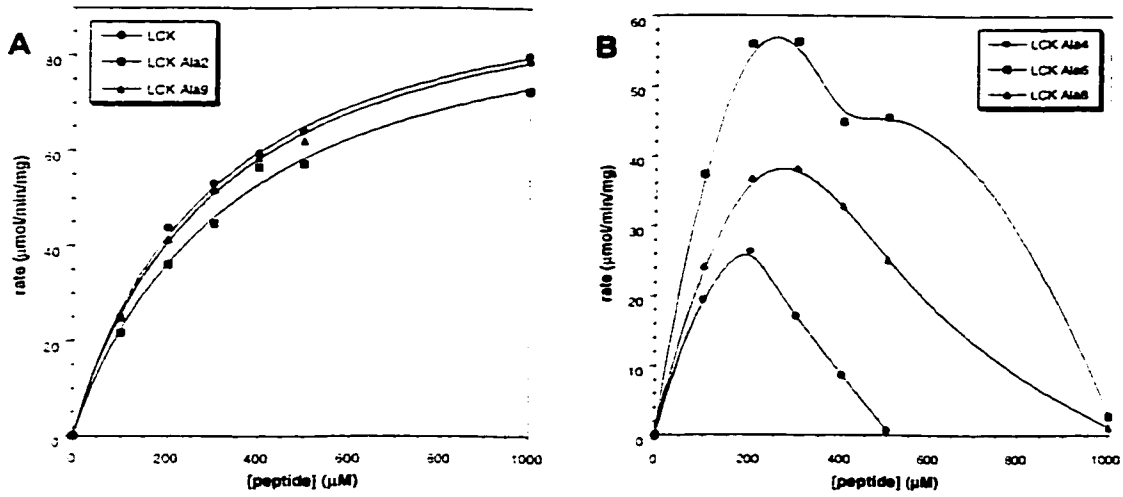


Fig. 6.1. **Velocity *vs.* substrate concentration plots for various LCK peptides.** PTP activity of CD45 (~300 ng/ml) was assayed with varying concentrations of peptide substrate and initial rates were determined as described under "Experimental Procedures." (A) Velocity *vs.* peptide concentration plot for the LCK parent, Ala2, and Ala9 peptides, representative of the peptides that displayed normal Michaelis-Menten kinetics. (B) Velocity *vs.* peptide concentration plot for the LCK Ala4, Ala6, and Ala8 peptides that displayed apparent substrate inhibition.

The nature of this apparent substrate inhibition was further investigated. As these three Ala substitutions all replaced Gln residues, it was initially hypothesized that perhaps these Gln residues were essential and the Gln→Ala substitutions actually created a substrate that could also inhibit the enzyme at high concentrations. To examine this, peptides with double- and triple-Ala substitutions at positions 4, 6, and 8 were synthesized and tested, as well as an Asn4 peptide. No substrate inhibition was observed with any of these peptides, though.

In fact, further studies suggested that the apparent substrate inhibition was simply due to a potent trace inhibitor(s) in the peptide sample that was inactivating the enzyme. For example, it was observed that for the Ala4- and Ala6-substituted peptides, the inhibition at 500 μM or 1 mM, respectively, could be overcome by increasing the enzyme concentration (Fig. 6.2). Plots of rate *vs.* enzyme concentration for these peptides were parallel to the curve for the parent peptide beyond a certain enzyme concentration, consistent with an inhibitor being titrated by the enzyme. The presence of an inhibitor in the Ala4 peptide was confirmed by purifying the peptide on HPLC and using the HPLC-purified peptide in an activity assay. As shown in Fig. 6.3, no activity was observed with 1 mM unpurified Ala4 peptide at a low enzyme concentration, while with 0.89 mM HPLC-purified Ala4 peptide significant activity was observed. Addition of five-fold more CD45 overcame the inhibition with the unpurified peptide and increased the rate with the purified peptide ~5.5-fold (from 5.17 mAU/min to 28.38 mAU/min), consistent with the removal of an inhibitor from the Ala4 peptide.

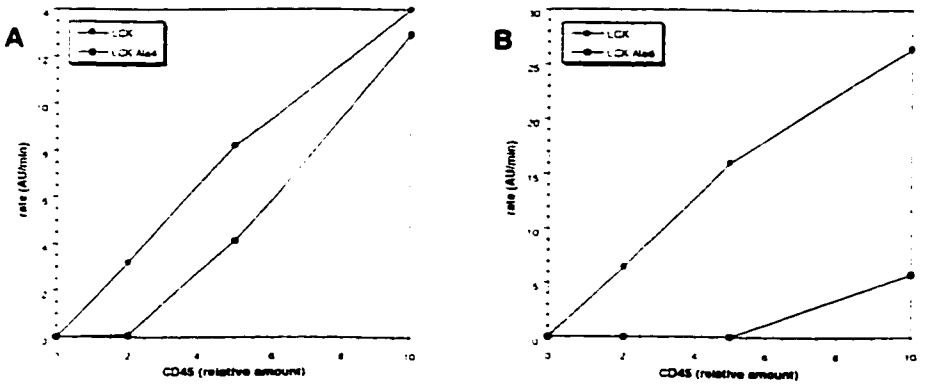


Fig. 6.2. **CD45 titration with various LCK peptide substrates.** CD45 was assayed with 500 μM LCK Ala4 peptide and LCK parent peptide (A), or 1 mM Lck Ala6 peptide and LCK parent peptide (B), and initial rates were determined as described under "Experimental Procedures." "Relative amount" of CD45 is in fact the volume in μl of CD45 stock assayed; however, as different preparations of CD45 were used in (A) and (B), the two experiments cannot be directly compared.

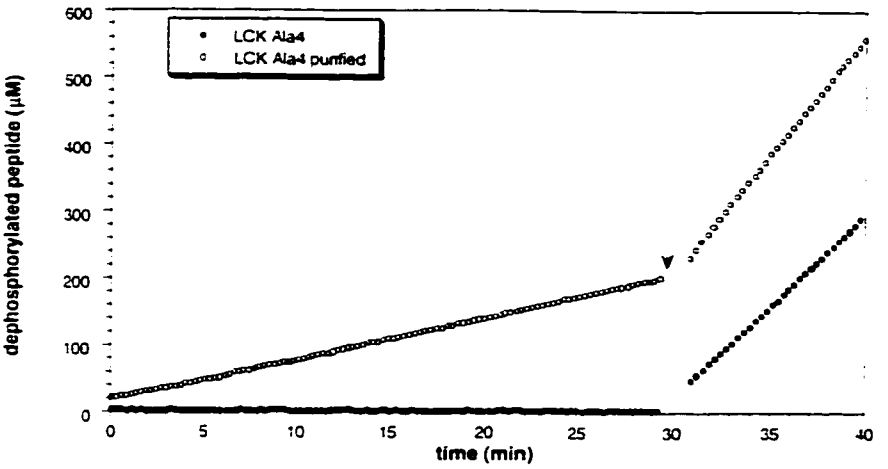


Fig. 6.3. **CD45 activity with LCK Ala4 peptide substrate with and without HPLC purification.** The timecourse shows the activity of CD45 with 1 mM LCK Ala4 peptide or 0.89 mM HPLC-purified LCK Ala4 peptide. To initiate the assay, 0.8 μl of CD45 stock was added to the 400 μl reaction mix. At the point marked by the arrow, an additional 4 μl of CD45 were added.

The kinetic parameters for the peptides with CD45 are summarized in Table 6.1. The K_m's for the LCK peptides varied over an approximately three-fold range (from 110 μM to 363 μM). Little variation was seen in the V_{max}'s towards the peptides assayed with the same preparation of enzyme, consistent with what was reported for PTP1B (19). V_{max} is dependent on the proportion of active enzyme in the preparation (*i.e.*, the specific activity), and clearly the second preparation of CD45 (bottom half of Table 6.1) had a lower specific activity, as the V_{max} with the LCK parent peptide was one-half that of the V_{max} of the first preparation of enzyme. K_m is independent of specific activity of the enzyme preparation, and indeed the K_m for the LCK parent peptide was the same with each preparation (Table 6.1). To permit comparisons between the results with different CD45 preparations, the catalytic efficiencies (k_{cat}/K_m) were normalized to that of each

preparation towards the LCK parent peptide (Table 6.1, last column). The lack of variation in the V_{max}'s is consistent with the second step of the catalytic mechanism, the hydrolysis of the enzyme thiol-phosphate intermediate, being rate-limiting (25, 26): this reaction is substrate-independent and therefore should define a similar V_{max} for all pY peptide substrates.

Table 6.1. **Kinetic parameters for LCK peptides with CD45.** The K_m and V_{max} for each peptide were determined by fitting plots of initial rate *vs.* substrate concentration to the Michaelis-Menten equation. The results are separated into two sections by the dotted line as two preparations of CD45 with different specific activities were used. Thus, the results cannot be compared directly for all peptides; however, in the final column the catalytic efficiencies are expressed relative to the LCK parent peptide for each batch to permit a comparison.

peptide	sequence	$K_m \pm$ S. E. (μM)	$V_{max} \pm$ S. E. (μmol/min/mg)	k_{cat}/K_m ($\times 10^{-5}$ $M^{-1}s^{-1}$)	relative k_{cat}/K_m
parent	TEGQ(pY)QPQP	313 ± 19	102.2 ± 2.5	4.08	1.00
Ala1	AEGQ(pY)QPQP	363 ± 43	108.2 ± 5.9	3.73	0.92
Ala2	TAGQ(pY)QPQP	309 ± 32	97.5 ± 4.3	3.95	0.97
Ala3	TEAQ(pY)QPQP	176 ± 39	99.8 ± 7.4	7.08	1.74
Ala7	TEGQ(pY)QAQP	208 ± 16	85.7 ± 2.6	5.15	1.26
Ala9	TEGQ(pY)QPQA	225 ± 11	102.0 ± 2.1	5.68	1.39
parent	TEGQ(pY)QPQP	307 ± 58	55.8 ± 5.4	2.27	1.00
Ala4,6	TEGA(pY)APQP	111 ± 12	50.8 ± 1.9	5.71	2.51
Ala4,8	TEGA(pY)QPAP	202 ± 34	54.5 ± 3.3	3.38	1.49
Ala6,8	TEGQ(pY)APAP	110 ± 13	48.8 ± 1.5	5.57	2.45
Ala4,6,8	TEGA(pY)APAP	117 ± 7	45.5 ± 0.8	4.85	2.14
Asn4	TEGN(pY)QPQP	229 ± 46	39.8 ± 2.9	2.17	0.96
IRP-5[a]	TRDI(pY)ETDYYRK	90 ± 13	50.7 ± 2.0	7.05	3.11

[a] Sequence derived from residues 1142 – 1153 of the insulin receptor.

Interestingly, none of the LCK Ala substitutions resulted in a substrate that was turned over less efficiently by CD45 than the parent sequence. This is in contrast to the results from the Ala scan performed with PTP1B and the *Yersinia* PTP (20). In fact, several substitutions resulted in better catalytic efficiencies, notably Ala3, Ala4,6, Ala4,8, and Ala4,6,8 (Table 6.1). Unfortunately the kinetic parameters with the individual Ala4, Ala6, and Ala8 peptides could not be determined because of the presence of the trace inhibitor(s); thus, the importance of the individual positions cannot be assessed. Assuming that the contributions of the individual changes are additive, it is likely that the Ala6 substitution accounts for the greatest increase in catalytic efficiency, as CD45 showed similar catalytic efficiencies towards Ala4,6 and Ala6,8 (which both share the Ala6 substitution) while Ala4,8 is only slightly improved. An unrelated peptide derived from the insulin receptor sequence (IRP-5) was the best substrate (Table 6.1). In general, though, the CD45 only displayed a three-fold range of catalytic efficiencies towards the substrates, suggesting that CD45 has low substrate specificity.

Kinetic Studies with LAR – Similar studies were undertaken with the phosphatase LAR to see if it displayed any specificity differences with these peptide substrates. However, the K_m for the LCK parent peptide was >1 mM; an accurate value could not be determined as the highest substrate concentration that could be tested was 1.5 mM because of solubility and substrate supply. Several of the Ala-substituted peptides were also tested, and no saturation was observed up to 1.5 mM, again making a determination of the kinetic parameters impossible.

Interestingly, LAR did not display the same inhibition observed with CD45 at higher concentrations of the Ala4, Ala6, and Ala8 peptides. As LAR was prepared under different conditions, it is possible that something in the LAR preparation was able to titrate or inactivate the inhibitor. Changing the buffer of the LAR enzyme stock to the same buffer as for CD45 would address this issue; however, this experiment was not undertaken. Alternatively, something specific about LAR may have rendered it insensitive to the inhibitory factor. Finally, a higher concentration of LAR was assayed as compared to CD45 that may have been sufficient to titrate the inhibitor, thereby masking the inhibition which may have been revealed at lower enzyme concentrations.

Kinetic Studies with PTPß – As PTPß displayed much lower K_m's than CD45 or LAR towards the LCK peptides, it was necessary to follow the dephosphorylation reaction using the more sensitive method of fluorescence because of the low substrate concentrations (10 – 125 μM). More variation was observed between the V_{max}'s for the various peptide substrates with PTPß than with CD45. With the Ala4, Ala6, and Ala8 peptides, the V_{max} values are likely reduced because of the presence of the trace inhibitor(s). Lower concentrations of these peptides were assayed as compared to the CD45 assays, and therefore the apparent substrate inhibition was not observed; however, the inhibitor was still present (see below) and would be expected to inactivate a portion of the enzyme, reducing the apparent V_{max}. For the parent, Ala7, and Ala9 peptides, the K_m was as high or higher than the largest substrate concentration assayed; thus, the kinetic parameters with these substrates are less certain because of the extrapolations of the data. It is possible that the V_{max} variations arise from an influence of the substrates on the rate-

determining step of PTPß, although a more extensive analysis is required to assess this. In general, the V_{max}'s for PTPß are lower than in other studies where values of 12 to 350 µmol/min/mg were reported for a variety of pY peptide substrates (27-29). In particular, Cho *et al.* (27) obtained a V_{max} of 174 µmol/min/mg for the LCK parent peptide, compared to 6.93 ± 1.43 µmol/min/mg in this study (Table 6.2). The differences may reflect an enzyme preparation of low specific activity or may be due to differences in the assay buffer conditions.

Table 6.2. **Kinetic parameters for LCK peptides with PTPß.** The K_m and V_{max} for each peptide were determined by fitting plots of initial rate *vs.* substrate concentration to the Michaelis-Menten equation.

peptide	sequence	K_m ± S. E. (µM)	V_{max} ± S. E. (µmol/min/mg)	k_{cat}/K_m ($x10^{-4}$ $M^{-1}s^{-1}$)	relative k_{cat}/K_m
parent	TEGQ(pY)QPQP	116 ± 40	6.93 ± 1.43	3.97	1.00
Ala1	AEGQ(pY)QPQP	97 ± 10	5.64 ± 0.34	3.87	0.98
Ala2	TAGQ(pY)QPQP	63 ± 10	3.86 ± 0.30	4.08	1.03
Ala3	TEAQ(pY)QPQP	79 ± 2	6.58 ± 0.09	5.57	1.40
Ala4	TEGA(pY)QPQP	56 ± 30	3.65 ± 0.89	4.34	1.09
Ala6	TEGQ(pY)APQP	32 ± 7	2.97 ± 0.27	6.16	1.55
Ala7	TEGQ(pY)QAQP	118 ± 31	7.37 ± 1.08	4.17	1.05
Ala8	TEGQ(pY)QPAP	43 ± 6	3.40 ± 0.20	5.34	1.35
Ala9	TEGQ(pY)QPQA	144 ± 11	9.28 ± 0.48	4.31	1.09
Ala4.6	TEGA(pY)APQP	68 ± 8	5.64 ± 0.47	5.57	1.40
Ala4.8	TEGA(pY)QPAP	73 ± 9	5.37 ± 0.35	4.89	1.23
Ala6.8	TEGQ(pY)APAP	44 ± 9	5.12 ± 0.46	7.72	1.95
Ala4.6.8	TEGA(pY)APAP	61 ± 16	6.45 ± 1.08	7.09	1.79

As for CD45, an approximately three-fold difference in K_m's was observed between the highest and lowest affinity peptides, but again little variation in the catalytic

efficiencies (k_{cat}/K_m) was observed (Table 6.2). The Ala substitutions resulted in substrates that were turned over as well or better than the LCK parent peptide, with the best increases observed with the Ala3, Ala6, Ala8, Ala4,6, Ala6,8, and Ala4,6,8 peptides. The ability to determine kinetic parameters for the Ala4, Ala6, and Ala8 peptides with PTPß indicated that Ala6 had the biggest positive effect of the three, which was what was predicted indirectly for CD45 (see above). However, the variation in PTPß catalytic efficiencies was only two-fold, again suggesting that, like CD45, PTPß does not discriminate much between these peptide substrates.

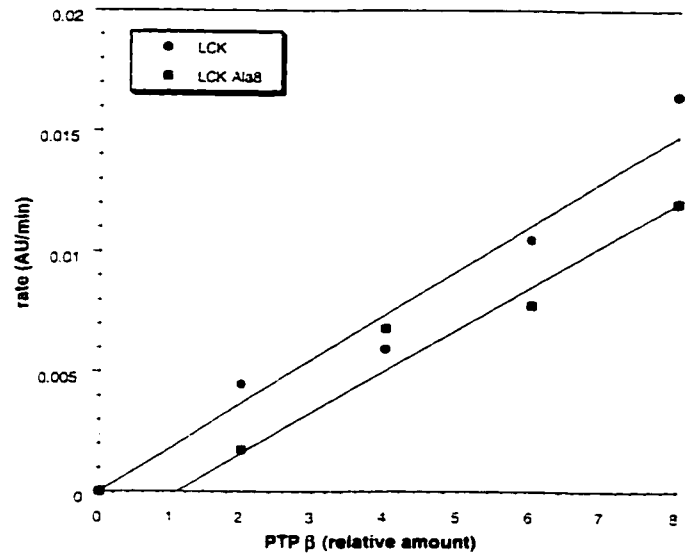


Fig. 6.4. **PTPß titration with LCK Ala8 peptide substrate.** PTP activity of PTPß was assayed with 1 mM LCK parent or Ala8 peptide as described under "Experimental Procedures." Initial rates were determined and plotted *vs.* relative amount of PTPß assayed (1 = 500 ng/ml final concentration).

The reduced V_{max} for PTPß with the Ala4, Ala6, and Ala8 peptides (see above) suggested that the enzyme was being affected by the trace inhibitor responsible for the

apparent substrate inhibition with CD45. To test this hypothesis, an enzyme titration was carried out for PTPß with the parent and Ala8 peptides at 1 mM. At this concentration of substrate, the activity had to be monitored by absorbance and not fluorescence because of the high fluorescent background. While PTPß displayed activity with the Ala8 peptide at all concentrations of PTPß assayed, the enzyme titration plot (Fig. 6.4) showed that the curve for the Ala8 peptide was parallel to that for the parent peptide but shifted to the right, such that it did not intersect the origin. This type of plot is again consistent with a trace inhibitor that is titrated by the enzyme; the x-intersect indicates the amount of enzyme required to completely titrate the inhibitor. Thus, PTPß appears to be similar to CD45 in its sensitivity towards this trace inhibitor.

DISCUSSION

The kinetic data obtained with these enzymes is consistent with published data. For example, CD45 was reported to have a K_m of 130 µM and a k_{cat} of 66 µmol/mg/min towards the LCK peptide (27). The difference in values as compared to those reported here is likely due to the fact that the kinetic assays in (27) were carried out at pH 6.0: repeating the analysis of CD45 with the LCK peptide at pH 6.0 gave very similar results to those published (K_m = 133 ± 16 µM, V_{max} = 33.7 ± 1.2 µmol/min/mg). Similarly, LAR is reported to have K_m values higher than CD45, while PTPß displays lower K_m's as observed here (27).

The inhibition of CD45 observed with the Ala4, Ala6, and Ala8 peptides at high concentrations of substrate revealed a potential danger with using these peptide substrates

in that they may be contaminated with a trace inhibitor(s) that inactivates the enzyme. In fact, this is a concern with any enzyme kinetic study, and certainly an important control if such an apparent substrate inhibition pattern is observed is to see whether the inhibition is eliminated at high enzyme concentrations.

CD45 and PTPß did not appear to discriminate very much between the various pY-peptide substrates with which they were presented, as the catalytic efficiencies did not vary by more than three-fold (Tables 4.1 & 4.2). In the only study of CD45 with pY peptide substrates (27), a nine-fold range in catalytic efficiencies was observed between the best and worst substrates. In the same study, PTPß displayed a 200-fold range in catalytic efficiencies. It is clear, then, that CD45 and PTPß can display a certain amount of sequence selectivity, although it appears to be much less pronounced for CD45. A problem with the use of individual peptide substrates is that it is not possible to screen all possible sequence combinations, and the choice of sequence will greatly influence the results. In this study, substituting the positions surrounding Tyr_{505} of $p56^{lck}$ with Ala residues had little effect on catalytic efficiency for both CD45 and PTPß. However, substitution with different residues may have generated very different results. Another problem with the use of peptide substrates is that substrate recognition *in vivo* may involve the secondary structure around the pY and not simply the primary sequence, which cannot be mimicked by short peptides.

An on-going debate regarding PTPs is how much specificity is intrinsic to the enzyme and how much is due to localization in the cell and access to potential substrates. For example, the transmembrane PTPs are embedded in the cell membrane and would be

expected to have a much more restricted access to substrates than a PTP free in the cytosol. Thus, it may not be necessary for transmembrane PTPs to have as much intrinsic specificity as a cytosolic PTP. For example, CD45 is associated with the TCR and its presumed substrates are recruited to the TCR complex upon T-cell stimulation, which may obviate the need for stringent intrinsic substrate selectivity on the part of CD45. The apparent lack of sequence specificity observed *in vitro* with many PTPs supports such a hypothesis. However, more systematic studies with pY peptide substrates must be carried out before such general conclusions on PTP specificity can be drawn.

ACKNOWLEDGMENTS

We thank D. Pot and F. Jirik for providing the expression clones of LAR and PTPß, respectively, and S. Liu for the purification of LAR.

REFERENCES

1. Sun, H., and Tonks, N. K. (1994) *Trends Biochem. Sci.* **19,** 480-485
2. Tonks, N. K., and Neel, B. G. (1996) *Cell* **87,** 365-368
3. Tonks, N. K. (1996) *Adv. Pharmacol.* **36,** 91-119
4. Neel, B. G., and Tonks, N. K. (1997) *Curr. Opin. Cell Biol.* **9,** 193-204
5. Tonks, N. K., Diltz, C. D., and Fischer, E. H. (1988) *J. Biol. Chem.* **263,** 6722-6730
6. Tonks, N. K., Diltz, C. D., and Fischer, E. H. (1988) *J. Biol. Chem.* **263,** 6731-6737
7. Charbonneau, H., Tonks, N. K., Walsh, K. A., and Fischer, E. H. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 7182-7186
8. Tonks, N. K., Diltz, C. D., and Fischer, E. H. (1990) *J. Biol. Chem.* **265,** 10674-10680

9. Streuli, M., Krueger, N. X., Thai, T., Tang, M.,and Saito, H. (1990) *EMBO J.* **9,** 2399-2407

10. Itoh, M., Streuli, M., Krueger, N. X., and Saito, H. (1992) *J. Biol. Chem.* **267,** 12356-12363

10a. Tan, X., Stover, D. R., and Walsh, K. A. (1993) *J. Biol. Chem.* **268,** 6835-6838

11. Koretzky, G. A., Picus, J., Schultz, T., and Weiss, A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 2037-2041

12. Trowbridge, I. S., and Thomas, M. L. (1994) *Ann. Rev. Immunol.* **12,** 85-116

13. Streuli, M., Krueger, N. X., Hall, L. R., Schlossman, S. F., and Saito, H. (1988) *J. Exp. Med.* **168,** 1523-1530

14. Streuli, M., Krueger, N. X., Ariniello, P. D., Tang, M., Munro, J. M., Blatter, W. A., Adler, D. A., Disteche, C. M., and Saito, H. (1992) *EMBO J.* **11,** 897-907

15. Serra-Pages, C., Saito, H., and Streuli, M. (1994) *J. Biol. Chem.* **269,** 23632-23641

16. Serra-Pages, C., Kedersha, N. L., Fazikas, L., Medley, Q., Debant, A., and Streuli, M. (1995) *EMBO J.* **14,** 2827-2838

17. Pulido, R., Serra-Pagès, C., Tang, M., and Streuli, M. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 11686-11690

18. Schaapveld, R. Q. L., Schepens, J. T. G., Robinson, G. W., Attema, J., Oerlemans, F. T. J. J., Fransen, J. A. M., Streuli, M., Wieringa, B., Hennighausen, L., and Hendriks, W. J. A. J. (1997) *Dev. Biol.* **188,** 134-146

19. Kreuger, N. X., Streuli, M., and Saito, H. (1990) *EMBO J.* **9,** 3241-3252

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

21. Huyer, G., Li, Z. M., Adam, M., Huckle, W. R., and Ramachandran, C. (1995) *Biochemistry* **34,** 1040-1049

22. Dechert, U., Adam, M., Harder, K. W., Clark-Lewis, I., and Jirik, F. (1994) *J. Biol. Chem.* **269,** 5602-5611

23. Pot, D. A., Woodford, T. A., Remboutsika, E., Haun, R. S., and Dixon, J. E. (1991) *J. Biol. Chem.* **266,** 19688-19696

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

25. Zhang, Z.-Y., Malachowski, W. P., Van Etten, R. L., and Dixon, J. E. (1994) *J. Biol. Chem.* **269,** 8140-8145

26. Zhang, Z.-Y. (1995) *J. Biol. Chem.* **270,** 11199-11204

27. Cho, H., Krishnaraj, R., Itoh, M., Kitas, E., Bannwarth, W., Saito, H., and Walsh, C. T. (1993) *Protein Science* **2,** 977-984

28. Wang, Y., and Pallen, C. J. (1992) *J. Biol. Chem.* **267,** 16696-16702

29. Harder, K. W., Owen, P., Wong, L. K. H., Aebersold, R., Clark-Lewis, I., and Jirik, F. (1994) *Biochem. J.* **298,** 395-401

CHAPTER 7

Affinity Selection from Peptide Libraries to Determine Substrate Specificity of Protein Tyrosine Phosphatases

SUMMARY

Affinity selection from peptide libraries is a powerful tool that has been used for determining the sequence specificities of a number of enzymes and protein binding domains, including protein kinases, SH2 domains, and PDZ domains. We have extended this approach to protein tyrosine phosphatases using peptide libraries containing a nonhydrolyzable phosphotyrosine analog, difluorophosphonomethyl phenylalanine (F_2Pmp). A size-exclusion method is used to separate enzyme–peptide complexes from free peptide, providing several advantages over the traditional immobilized protein affinity column approach. In addition, the feasibility of mass spectrometric detection to quantitate peptides rapidly and reproducibly is demonstrated as an alternative to quantitation by peptide sequencing. The validity of this analysis is demonstrated by synthesizing individual peptides and comparing their affinity for enzyme with the predictions from the affinity selection process. As a model for these studies the protein tyrosine phosphatase PTP1B is used, providing additional insights into the sequence specificity of this enzyme. In particular, a selection for aromatic amino acids at the pY-1 position (immediately N-terminal to the phosphotyrosine), as well as a broad pY+1 selectivity, is observed in addition to the general preference for acidic residues N-terminal to the phosphotyrosine.

The approach described here should prove applicable to protein tyrosine phosphatases in general as well as for the study of non-peptidyl combinatorial libraries.

INTRODUCTION

The reversible phosphorylation of proteins on tyrosine residues regulates many cellular processes (1-4). The tyrosyl phosphorylation status of a protein is controlled by the competing activities of the protein tyrosine kinases (PTKs)[1] and the protein tyrosine phosphatases (PTPs). Over 100 PTKs and 50 PTPs have been identified, and it is predicted that the human genome contains up to 500 of each (3, 4). Clearly one of the major challenges in the study of these enzymes is to determine the signaling pathway(s) in which they participate; *i.e.*, to identify their *in vivo* targets and substrates.

One approach to determine potential substrates of PTKs and PTPs has been to study their *in vitro* catalytic efficiency towards purified proteins or synthetic peptides. For PTPs, the use of synthetic pY-containing peptide substrates in particular has been informative, and a number of groups have characterized various PTPs in this way in an attempt to define recognition elements (5-12). However, it is very difficult to carry out anything more than a limited study with individual peptides because the process is costly, time consuming, and laborious.

The use of peptide libraries provides a means to overcome some of these drawbacks. Songyang *et al.* pioneered this technique for the determination of the

[1] Abbreviations: PTK, protein tyrosine kinase; PTP, protein tyrosine phosphatase; EGF, epidermal growth factor; pY, phosphotyrosine; SH2, Src homology 2; F_2Pmp, difluorophosphonomethyl phenylalanine; FDP, fluorescein diphosphate; FMP, fluorescein monophosphate; DTT, dithiothreitol; BSA, bovine serum albumin; MS, mass spectrometry; GST, glutathione *S*-transferase.

recognition requirements of SH2 domains (13). Their approach was to use the SH2 domain as an affinity ligand to 'purify' high-affinity peptides from a degenerate peptide library: the pooled affinity-selected peptides were then sequenced. A similar method was used to probe the carboxy-terminal motifs recognized by PDZ domains (14). A modified approach for studying kinases was also developed in which a peptide library was incubated with the kinase of interest in a phosphorylation reaction mixture, and the phosphorylated peptides were isolated and sequenced (15, 16). As an alternative to sequencing peptides, mass spectrometric detection has also been used qualitatively to analyze SH2 domain–peptide interactions (17). The library approach has the advantage of exposing the protein of interest to many different potential ligands at the same time, resulting in a much more complete analysis in significantly less time than could be accomplished with individual peptides.

We report here the use of peptide libraries to probe the peptide substrate specificity of PTPs, using PTP1B as a model. The sequence surrounding Tyr_{992} of the EGF receptor [DADE(pY)L] previously shown to have high affinity for PTP1B (5, 6) was used as a template for the libraries. Five libraries of 19 peptides each were synthesized in which each position in the hexapeptide (except pY) was substituted with all the naturally-occurring L-amino acids except Cys. Because peptides containing pY would be rapidly dephosphorylated by the enzyme, a nonhydrolyzable pY analog (difluorophosphonomethyl phenylalanine, F_2Pmp) was used (18). PTP1B was incubated with the libraries, and enzyme–peptide complexes were separated from free peptides by gel filtration. The bound peptides were identified by electrospray ionization mass spectrometry to determine amino

acid preferences at each position. To validate this method, individual peptides were analyzed (both F_2Pmp-containing peptides as inhibitors and pY-containing peptides as substrates) and the data showed complete agreement with the affinity selection predictions. The gel filtration separation and the mass spectrometric quantitation of peptides provide several advantages over previous affinity selection techniques. The results of this study provide additional insights into the specificity of PTP1B and demonstrate the applicability of this affinity selection approach to PTPs in general.

EXPERIMENTAL PROCEDURES

Materials – The FDP substrate[2] and the protected F_2Pmp monomer (19) were synthesized as described. The resin and HATU coupling reagent for peptide synthesis were obtained from PerSeptive Biosystems. Phosphotyrosine-containing peptides were obtained from California Peptide Research (Napa, CA). All other chemicals were of reagent grade from Sigma. Amino acid analysis and sequencing of peptide libraries were carried out by Harvard Microchem (Cambridge, MA).

Expression and Purification of PTP1B – Two recombinant forms of PTP1B were used: Flag-PTP1B and GST-PTP1B. Studies showed that the two forms are catalytically identical and are therefore interchangeable in both kinetic and binding studies[3]. Flag-PTP1B was expressed and purified as previously described (20). To prepare the GST-PTP1B fusion, the catalytic domain of PTP1B (amino acids 1-320) was amplified by PCR from a full-length PTP1B cDNA clone, incorporating *Bam*HI and *Eco*RI sites at the 5'

[2] Scheigetz, J., Gilbert, M., and Zamboni, R., manuscript submitted.
[3] G. H., unpublished observations.

and 3' ends, respectively. The amplified fragment was digested with *Bam*HI and *Eco*RI, isolated from a 1% agarose gel, and cloned into the pGEX-2T GST fusion vector (Pharmacia) that had been digested with the same restriction enzymes. Sequencing ensured that no errors had been introduced by PCR.

Escherichia coli cells containing the GST-PTP1B plasmid were grown in Luria Broth with 100 μg/ml ampicillin at 37°C to $A_{600} \equiv 0.8$. The cultures were then induced with the addition of IPTG to 0.5 mM and grown overnight at 27°C. Cells were harvested and resuspended in ice-cold lysis buffer [phosphate-buffered saline containing 2.5% Triton X-100, 5 mM DTT, and Complete protease inhibitor cocktail (Boehringer Mannheim)] and lysed using a Bead Beater (Biospec Products) with 0.1-mm glass beads. The lysate was centrifuged for 15 min at 43,000 x *g* and then incubated with glutathione-Sepharose (Pharmacia) for 1 h at 4°C with end-over-end mixing. The beads were washed extensively with lysis buffer and packed into a column for elution of the GST-PTP1B fusion protein by competition with 10 mM reduced glutathione in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM DTT. Approximately 8 mg of GST-PTP1B (>95% pure as estimated by SDS-polyacrylamide gel electrophoresis) were recovered from a 1-litre culture.

Peptide Synthesis – F_2Pmp-containing peptides were synthesized on a 9050 Plus PepSynthesizer (PerSeptive Biosystems). The resin used for synthesis was Fmoc-PAL-PEG-PS. Peptides were synthesized in the "fast cycles" mode using the coupling reagent HATU. All amino acids were Fmoc-protected. The peptides were synthesized with a free amino terminal and a carboxyamide carboxy terminal. The quality of the peptide syntheses was confirmed by amino acid analysis, mass spectrometry, and HPLC.

The peptide libraries were synthesized in the same manner as the individual peptides except that an isokinetic mixture of 19 amino acids (all naturally-occurring L-amino acids except cysteine) was added at the degenerate position. The libraries were analyzed by amino acid analysis, sequencing, and mass spectrometry. Overall, peptides with Val, Leu/Ile, and Thr at the degenerate position tended to be over-represented, while those with Gly, Ser, and Pro tended to be under-represented. However, the concentration range between the most and least abundant peptides was estimated to be less than an order of magnitude. All peptide libraries were reversible competitive inhibitors of PTP1B.

Activity Assays with PTP1B – Kinetic parameters for F_2Pmp-containing peptides were determined by assaying Flag-PTP1B (1 or 5 nM final concentration) with FDP as substrate[4] in assay buffer (50 mM HEPES pH 7.3, 5 mM DTT, 10 μg/ml BSA) in a total volume of 200 μl in 96-well plates, containing varying amounts of peptide inhibitor. Appearance of FMP product was followed by fluorescence using a CytoFluor II plate reader (PerSeptive Biosystems) with excitation at 440 nm (slit width 20 nm) and emission at 530 nm (slit width 25 nm). Initial rates were determined at six substrate and five inhibitor concentrations, and the 30 data points were fit simultaneously to the Michaelis-Menten equation for competitive inhibition, *i.e.*, $v = (V_{max}[S]) / (K_m(1 + [I]/K_i) + [S])$, using the program GraFit (Erithacus Software), with the K_m for FDP fixed at 10 μM.

Kinetic parameters for the pY-containing peptides were determined by assaying Flag-PTP1B (10 nM final concentration) in assay buffer in a total volume of 1 ml in cuvettes. Appearance of dephosphorylated peptide product was followed using a Hewlett

[4] Huang, Z., Wang, Q., Ly, H. D., Govindarajan, A., Scheigetz, J., Zamboni, R., and Ramachandran, C., manuscript submitted.

Packard 8452A diode-array spectrophotometer at 282 nm (21). As the K_m's of the peptides were too low (<50 μM) to determine the values directly by this method, varying amounts of DADE(F_2Pmp)L competitive inhibitor peptide were added to increase the apparent K_m to ~50 – 200 μM ($K_m^{app} = K_m(1 + [I]/K_i)$). Initial rates were determined at six substrate and three inhibitor concentrations, and the 18 data points were fit simultaneously to the Michaelis-Menten equation for competitive inhibition using the program GraFit, with the K_i for the F_2Pmp peptide fixed at 26 nM.

Affinity Selection of Peptides – A buffer consisting of 50 mM Tris-HCl pH 7.2, 150 mM NaCl, 5 mM DTT, and 1 mM EDTA was used for binding reactions. GST-PTP1B (12.5 μM) was incubated with ~250 μM total peptide library for 5 min at room temperature in a 55-μl reaction volume. The sample (50 μl) was loaded onto a Centri-Sep gel filtration spin column (Princeton Separations) to separate PTP1B with bound peptides from free peptides, using the protocol described by the manufacturer. The protein in the eluate was denatured by the addition of TFA and acetonitrile (final concentrations of 0.5% and 25%, respectively) to release bound peptide, and 360 ng of an internal standard peptide [ADE(F_2Pmp)L] was added. Denatured PTP1B was removed by ultrafiltering the sample through a Millipore Ultrafree-MC 30,000 molecular weight cut-off membrane. The ultrafiltrate was evaporated to dryness and resuspended in 50 μl of 0.2% (v/v) formic acid for mass spectrometric analysis. For analysis of the library stocks, 48 μl of 24 μM total peptide library was subjected to the same post-column work-up described above. All library stock and affinity-selected samples were prepared in triplicate for analysis.

Mass Spectrometric Analysis of Peptides – The samples were analyzed by electrospray ionization mass spectrometry in negative ion mode in a similar manner to that described by Kelly *et al.* (17). A 20-µl aliquot of each sample was concentrated on-line using a 1-cm C_{18} small molecule cartridge (Michrom BioResources) in place of the normal injection loop of a HPLC injector. The cartridge was washed successively with 0.2% (v/v) formic acid (100 µl) and 2% (v/v) aqueous acetonitrile (100 µl). The peptides were eluted off the trap directly into the electrospray ionization source of a Finnigan TSQ 7000 triple quadrupole mass spectrometer with 80% acetonitrile at a flow rate of 5 µl/min using a syringe pump (Harvard Apparatus). The electrospray ionization voltage was -3.8 kV, the nebulization gas pressure 40 p.s.i., and the inlet capillary temperature 210°C. The mass spectrometer was tuned and calibrated using the automated procedures supplied by the manufacturer. Instrument resolution was set to approximately unit mass resolution (0.7 mass units full width at half height). Mass spectra (m/z 600-1000) were acquired in profile mode using the third quadrupole (Q3) as the mass analyzer. All spectra acquired as the peptides eluted were summed, and the peak heights of the monoisotopic, singly-deprotonated peptides were determined. The average background was determined in spectral regions devoid of peptide ions and subtracted from the peak intensities.

RESULTS

Development of Affinity Selection Protocol – For the initial development of the technique, the peptide DADE(F_2Pmp)L was used. Our kinetic characterization showed it had a K_i of 26 nM (Table 7.2), in good agreement with published results (18, 22). To establish the

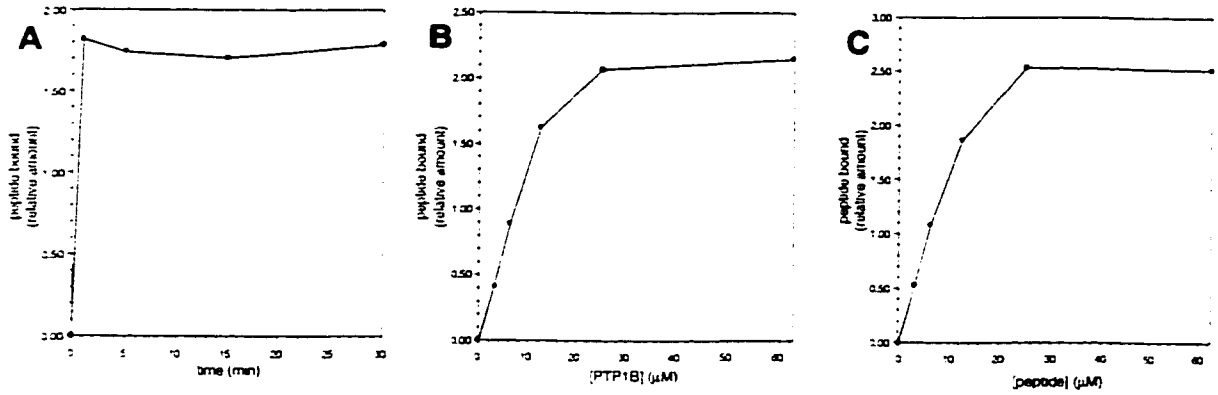


Fig. 7.1. **Development of affinity selection protocol.** (A) Time course of binding. GST-PTP1B (12.5 μM) and DADE(F_2Pmp)L peptide (12.5 μM) were incubated for the times indicated before being loaded onto the gel filtration column as described under "Experimental Procedures." In all cases, the relative amount of PTP1B-bound peptide is expressed as the ratio of the MS peak height of the peptide after affinity selection to the peak height of the internal standard. (B) Binding titration with varying enzyme concentration. GST-PTP1B at varying concentrations was incubated with 12.5 μM DADE(F_2Pmp)L peptide for 5 min before being loaded onto the gel filtration column. (C) Binding titration with varying peptide concentration. GST-PTP1B (12.5 μM) was incubated with varying concentrations of DADE(F_2Pmp)L peptide for 5 min before being loaded onto the gel filtration column.

time course of binding, PTP1B was incubated with the peptide (both at 12.5 μM) for varying times before applying the sample to the spin column. As shown in Fig. 7.1a, maximal binding had occurred by 1 min and remained maximal for at least 30 min. In order to confirm the stoichiometry of binding, the enzyme concentration was first varied at a constant peptide concentration (12.5 μM) in the incubation reaction; and second, the peptide concentration was varied at a constant enzyme concentration (12.5 μM). In both cases, saturation was observed between 12.5 μM and 25 μM (Fig. 7.1b,c), consistent with 1:1 binding as expected from the crystal structure of PTP1B complexed with peptide (23). In control experiments, only 70-80% of the PTP1B activity was recovered from the gel filtration column, indicating that a portion of the enzyme was inactivated or lost. In

experiments without enzyme, with denatured or inactivated PTP1B, or with CD45 (an enzyme very poorly inhibited by the DADE(F_2Pmp)L peptide), a background of ~2% of the peptide load through the gel filtration column was typically observed. Also, no loss of peptide due to non-specific binding to the Ultrafree-MC membrane was detected.

Mass Spectrometric Library Analysis – The mass spectrometric analysis required that all peptides in the library have unique masses in order to be quantitated. As such, only one position in the peptide sequence could be degenerate. Using the high-affinity sequence DADE(F_2Pmp)L from Tyr_{992} of the EGF receptor (5, 6) as a template, five libraries were made in which each position (except the F_2Pmp) was substituted. In referring to the libraries, the residues of the peptide are numbered 1 to 6 and the degeneracy is indicated by an "X". For example, in the X2 library the second position (Ala in the parent sequence) is degenerate.

A small molecule cartridge was used to trap and wash the peptides before elution into the mass spectrometer, which provided an efficient means to concentrate the sample and remove interfering buffer salts. Interestingly, the peptide cartridge made by the same manufacturer did not efficiently trap the peptides, and hence the small molecule cartridge was used. Negative ion mode was chosen for the peptide analysis because of the highly acidic nature of the peptides.

To verify that MS peak height could be used to quantitate the library peptides, different dilutions of the X4 peptide library solution [DAD<u>X</u>(F_2Pmp)L] were analyzed by MS. For all peptides substituted with neutral and acidic residues, the relationship between peptide concentration and peak height was linear over a wide concentration range (40 –

800 ng total stock injected) that bracketed the concentrations observed in a typical affinity-selected sample. The peptides containing Arg or His at the degenerate position displayed attenuated signals at higher concentrations. This effect only became significant above 200 ng stock, and as these peptides were generally selected against in the affinity selection process (see below), this non-linear relationship had no practical effect on the data and interpretation. The peak height *vs.* concentration relation for the Lys-substituted peptide could not be determined as Lys has the same mass as Gln; thus, these peptides were not resolved by the mass spectrometer. Similarly, Leu and Ile are isobaric and were not distinguished in this analysis.

Affinity Selection from Peptide Libraries – The size-exclusion method employed here for the affinity selection analysis requires that excess peptide library be used so that the highest affinity peptides saturate the enzyme and compete out lower affinity peptides. Preliminary experiments were done with a 5:1 molar ratio of total peptide library to enzyme; however, increasing the ratio gave clearer results, and a 20:1 ratio (250 μM library, 12.5 μM PTP1B) was ultimately employed. PTP1B with bound library peptides was separated from free peptides by gel filtration through the spin column, and after denaturation to release bound peptides and removal of PTP1B by ultrafiltration, the peptides were analyzed by mass spectrometry. Representative mass spectra for the X3 library [DAXE(F_2Pmp)L] are shown in Fig. 7.2. To quantitate peptides, the MS peak heights of each peptide were measured in both the library stock and the affinity-selected sample, and isotopic contributions from peptides one or two mass units lower than the peptide of interest were subtracted. For example, in the X3 library, the monoisotopic ion

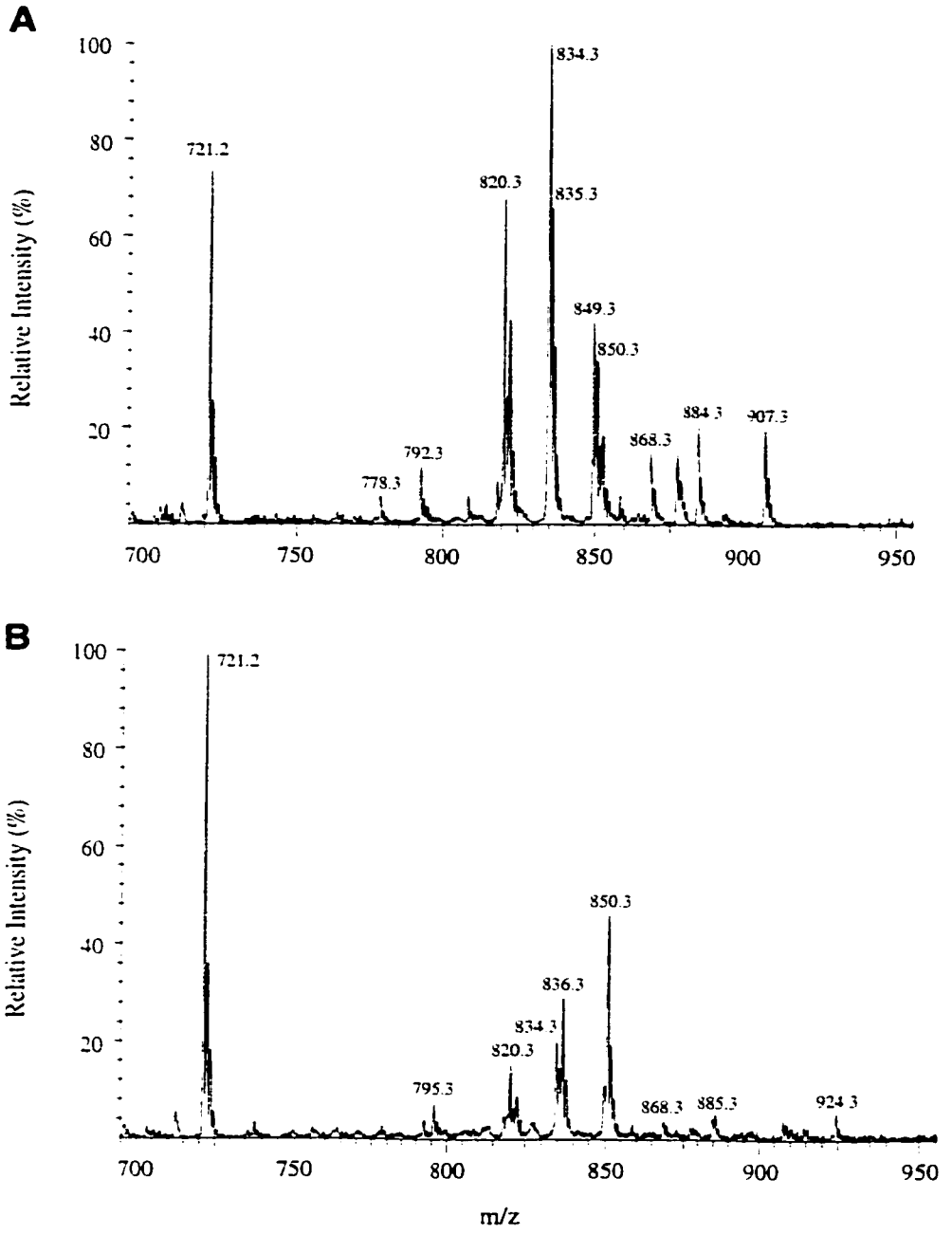


Fig. 7.2. **Representative mass spectra of the X3 peptide library.** Electrospray mass spectra of the X3 library stock (A) and the X3 library after affinity selection (B). Samples were prepared and analyzed as described under "Experimental Procedures." Each mass spectrum represents the summation of all the spectra acquired as the peptides eluted from the small molecule trap. The ion at m/z 721.2 corresponds to the mono-deprotonated ion ($[M-H]^-$) of the internal standard peptide, which has a similar intensity in both spectra. The masses of the mono-deprotonated ions of the X3 library peptides are listed in Table 7.1.

of the Val peptide is observed at m/z 820.3. However, the M+2 isotope of this peptide also contributes to the signal at m/z 822.3, and this must be taken into account when calculating the peak height of the monoisotopic Thr peptide (m/z 822.3).

Relative selectivities were determined from the observed MS peak heights by applying the following calculations (Table 7.1). (i) The peak heights were normalized to the peak height of the internal standard ($[M-H]^-$ = m/z 721.2) to permit averaging of replicate runs. (ii) The ratio of the normalized affinity-selected peak height over the stock peak height was determined, which compensates for the fact that the libraries are not equimolar mixtures of peptides. (iii) To adjust for differences in peptide quantity loaded on the mass spectrometer, the values from (ii) were multiplied by the ratio of the sum of the stock peak heights over the affinity-selected peak heights. In this way, a value of 1.0 indicates no selection or deselection of a peptide.

The selectivity data (Fig. 7.3) indicate a number of clear preferences and dislikes at each position in the sequence. Acidic residues are clearly preferred at all positions N-terminal to the pY residue, as indicated by a relative selection >1.0. Interestingly, the fourth position also selects for aromatic amino acids (Phe and Tyr). The selection at the sixth position is broader, selecting for a number of residues. The basic amino acid Arg is selected against at all positions; the selection for Lys cannot be determined as it has the same mass as Gln and therefore its signal cannot be resolved.

Table 7.1. **Mass spectral data for X3 library [DAXE(F_2Pmp)L] affinity selection analysis.**

X[a]	mass[b] $(M-H)^-$	normalized stock peak height[c] (average ± S.E.)	normalized affinity-selected peak height[c] (average ± S.E.)	relative selection[d] (average ± S.E.)
G	778.3	0.075 ± 0.004	0.021 ± 0.002	1.11 ± 0.11
A	792.3	0.166 ± 0.012	0.034 ± 0.003	0.82 ± 0.07
S	808.3	0.080 ± 0.006	0.017 ± 0.002	0.86 ± 0.09
P	818.3	0.129 ± 0.009	0.041 ± 0.005	1.29 ± 0.15
V	820.3	0.891 ± 0.060	0.123 ± 0.018	0.55 ± 0.08
T	822.3	0.451 ± 0.022	0.066 ± 0.008	0.59 ± 0.07
L/I	834.3	1.335 ± 0.078	0.185 ± 0.016	0.56 ± 0.05
N	835.3	0.305 ± 0.020	0.058 ± 0.007	0.76 ± 0.09
D	836.3	0.223 ± 0.012	0.224 ± 0.026	4.04 ± 0.47
Q/K	849.3	0.564 ± 0.010	0.094 ± 0.019	0.67 ± 0.13
E	850.3	0.231 ± 0.008	0.379 ± 0.049	6.61 ± 0.85
M	852.3	0.226 ± 0.016	0.034 ± 0.006	0.61 ± 0.10
H	858.3	0.095 ± 0.006	0.025 ± 0.004	1.03 ± 0.17
F	868.3	0.202 ± 0.014	0.032 ± 0.001	0.64 ± 0.01
R	877.3	0.194 ± 0.013	0.020 ± 0.001	0.42 ± 0.01
Y	884.3	0.273 ± 0.012	0.038 ± 0.004	0.56 ± 0.06
W	907.3	0.274 ± 0.020	0.030 ± 0.006	0.44 ± 0.09

[a] Each peptide has the sequence DAXE(F_2Pmp)L, where X is the residue indicated.

[b] The mass is for the singly-deprotonated ion, as observed by mass spectrometry.

[c] The peak heights for each of the peptide ions were normalized to the peak height of the internal standard peptide, and the results of three determinations were averaged.

[d] Relative selection was determined from the ratio of column 4 to column 3. The ratios were multiplied by the ratio of the sum of column 3 over the sum of column 4 to adjust for differences in sample load on MS, such that a value of 1.0 indicates no apparent selection or deselection.

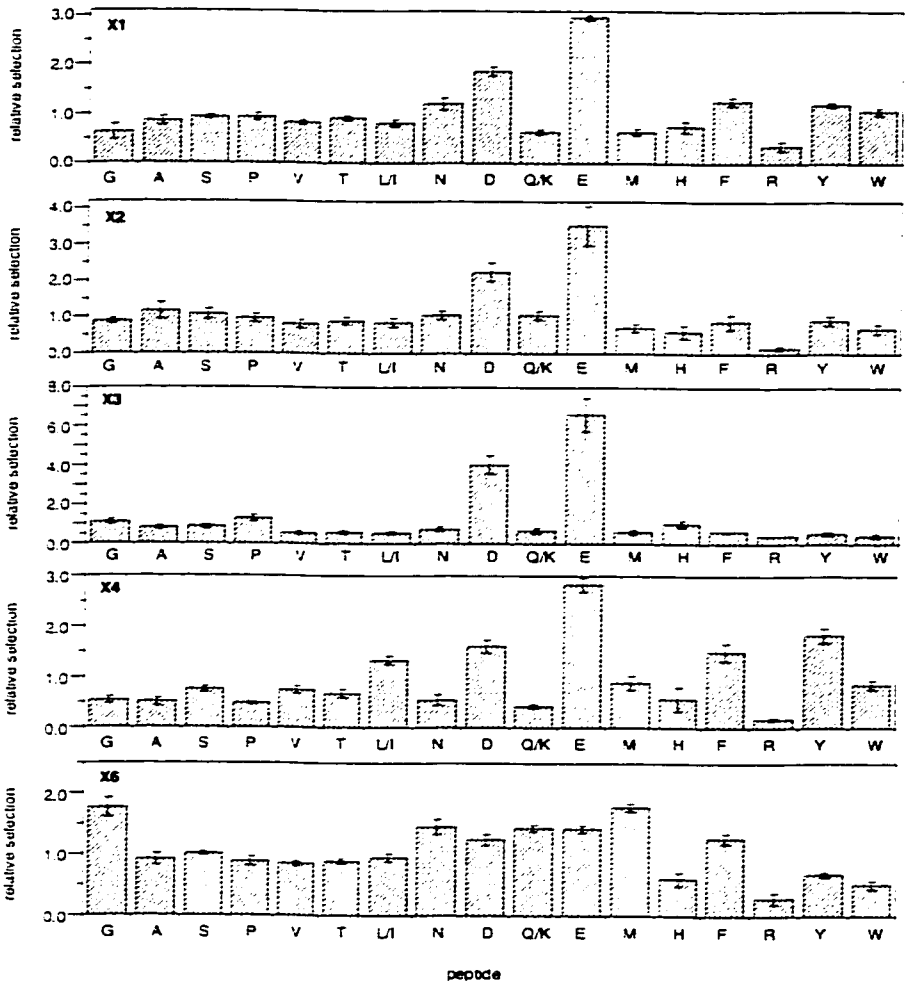


Fig. 7.3. **Summary of library affinity selection results.** Histograms summarizing the results of the affinity selection are shown for each library. The one-letter amino acid code is used to indicate the residue incorporated at the degenerate position. The amino acids leucine/isoleucine (L/I) and glutamine/lysine (Q/K) are isobaric and therefore the peptides containing these pairs of residues cannot be resolved using the current mass spectrometric scanning technique. The data processing and manipulations are described in the "Results" and in Table 7.1. A relative selection value of 1.0 is considered to indicate no apparent selection or deselection by the enzyme.

Kinetic Analysis of PTP1B Inhibition by Individual F_2Pmp-containing Peptides – Mass spectrometry is not frequently used in a quantitative manner as employed here. The fact that we observed a linear relationship between MS peak height and peptide concentration (see above) suggested that MS was quantitative under these conditions. However, in order to confirm the selection results, a number of individual F_2Pmp-containing peptides were synthesized and their K_i's determined (Table 7.2). Overall, there was a very good correlation between K_i and the relative selection observed in the affinity selection. A 'super peptide' incorporating the best changes at each position [*i.e.*, EEDE(F_2Pmp)M] was synthesized and tested. Its K_i was found to be the lowest at 7.2 nM, suggesting that the binding energy contributions from the individual positions are additive. All peptide inhibitors were found to be reversible and competitive with substrate.

Kinetic Analysis of PTP1B Catalysis Using pY-containing Peptide Substrates – It was not possible to use pY-containing peptides for affinity selection as they would be rapidly dephosphorylated by PTP1B. However, it was important to establish whether the selection predictions with the F_2Pmp-containing peptide inhibitors would be truly reflected in catalytic efficiency towards pY-containing peptide substrates. To this end, a number of pY-containing peptides were synthesized and their kinetic parameters (K_m and k_{cat}) determined (Table 7.3). The catalytic efficiencies (k_{cat}/K_m) of PTP1B with these peptides very closely parallel the inhibition results with the F_2Pmp-containing peptides (Table 7.2), indicating that the results with the nonhydrolyzable peptide inhibitors do reflect the true substrate specificity of the enzyme.

Table 7.2. **Inhibition constants of individual F_2Pmp-containing peptides.**

peptide	sequence	relative selection[a]	$K_i \pm$ S.E. (nM)
R1	RADE(F_2Pmp)L	–	205 ± 25
E1	EADE(F_2Pmp)L	+	28.7 ± 1.7
R2	DRDE(F_2Pmp)L	–	79.3 ± 6.6
E2	DEDE(F_2Pmp)L	+	9.7 ± 1.0
E3	DAEE(F_2Pmp)L	+	36.5 ± 5.3
G3	DAGE(F_2Pmp)L	–	177 ± 23
R4	DADR(F_2Pmp)L	–	374 ± 53
Y4	DADY(F_2Pmp)L	+	21.4 ± 2.1
D4	DADD(F_2Pmp)L	+	24.3 ± 1.7
M6	DADE(F_2Pmp)M	+	14.3 ± 2.6
G6	DADE(F_2Pmp)G	+	17.2 ± 2.0
R6	DADE(F_2Pmp)R	–	82.2 ± 6.3
super	EEDE(F_2Pmp)M	(+)	7.2 ± 1.1
P	DADE(F_2Pmp)L		26.1 ± 2.2

[a] Relative selection is taken from the affinity selection data in Fig. 7.3: "+", >1.0; "–", ≤1.0. The relative selection for the 'super peptide' is predicted to be >1.0; however, as it was not determined by affinity selection, it is indicated in parentheses.

Table 7.3. **Kinetic characterization of phosphotyrosine-containing peptides.**

peptide	sequence	$K_m \pm$ S.E. (μM)	$k_{cat} \pm$ S.E. (s^{-1})	x 10^{-7} k_{cat}/K_m ($M^{-1}s^{-1}$)
P	DADE(pY)L	0.89 ± 0.06	14.0 ± 0.3	1.57
R1	RADE(pY)L	6.64 ± 0.20	14.3 ± 0.2	0.22
E2	DEDE(pY)L	0.68 ± 0.05	13.7 ± 0.3	2.03
E3	DAEE(pY)L	1.09 ± 0.10	14.8 ± 0.5	1.36
R4	DADR(pY)L	9.77 ± 0.58	13.4 ± 0.3	0.14
R6	DADE(pY)R	3.73 ± 0.25	12.6 ± 0.4	0.34
M6	DADE(pY)M	0.71 ± 0.06	13.4 ± 0.4	1.87
super	EEDE(pY)M	0.54 ± 0.03	15.0 ± 0.3	2.78

Further Tests of Affinity Selection Protocol – With the variety of F_2Pmp-containing peptides available and characterized as described above, the affinity selection protocol could be tested by using libraries reconstituted from various combinations and proportions of these peptides. In one series of experiments, the parent (P) peptide [DADE(F_2Pmp)L, K_i = 26 nM] and the R1 peptide [RADE(F_2Pmp)L, K_i = 205 nM] were mixed with 12.5 μM PTP1B at the following concentrations: (1) 10.2 μM P and 11.8 μM R1; (2) 20.4 μM P and 5.8 μM R1. The equilibrium concentrations of each peptide-enzyme complex were calculated assuming reversible, competitive binding for both peptides and compared to the actual values determined from the affinity selection. In both cases there was very good agreement between the actual proportions and the theoretical amounts (Fig. 7.4).

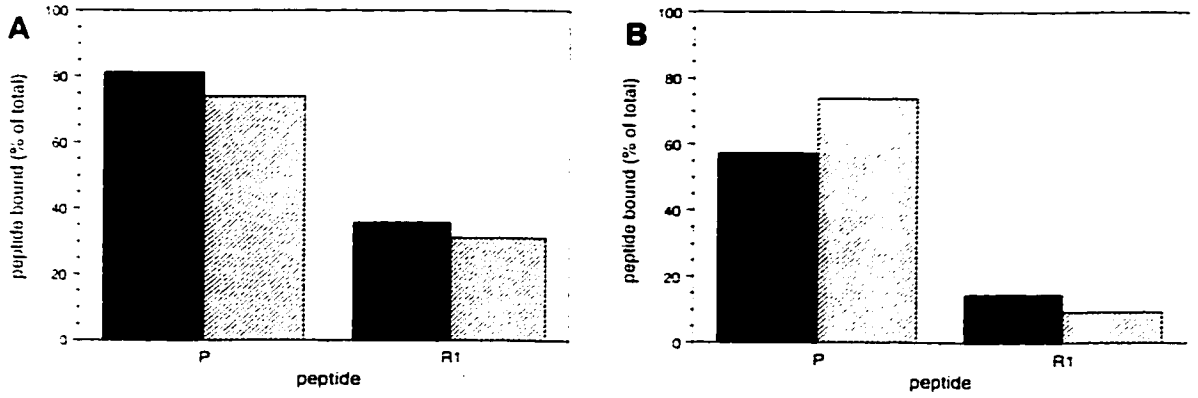


Fig. 7.4. **Affinity selection from two-peptide mixtures.** GST-PTP1B (12.5 μM) was incubated with the hexapeptides DADE(F_2Pmp)L ("P") and RADE(F_2Pmp)L ("R1") at the following concentrations: (A) 10.2 μM P and 11.8 μM R1; (B) 20.4 μM P and 5.8 μM R1. Bound peptides were separated and analyzed as described under "Experimental Procedures." The solid bars show the theoretical amounts of the peptides expected to bind, and the shaded bars show the actual amounts of peptide bound. In both cases, data are expressed as the percent of total peptide in the affinity selection mixture.

In a second experiment, 12 of the peptides synthesized (E1 was excluded because it has the same mass as E3) were mixed together to form an approximately equimolar mixture. This "mix of 12" library was then used for the affinity selection, with 12.5 μM PTP1B and ~180 μM total library (*i.e.*, ~15 μM of each peptide). The results of the affinity selection (Fig. 7.5) exactly paralleled the order of the K_i's of the peptides, further validating the technique and MS analysis.

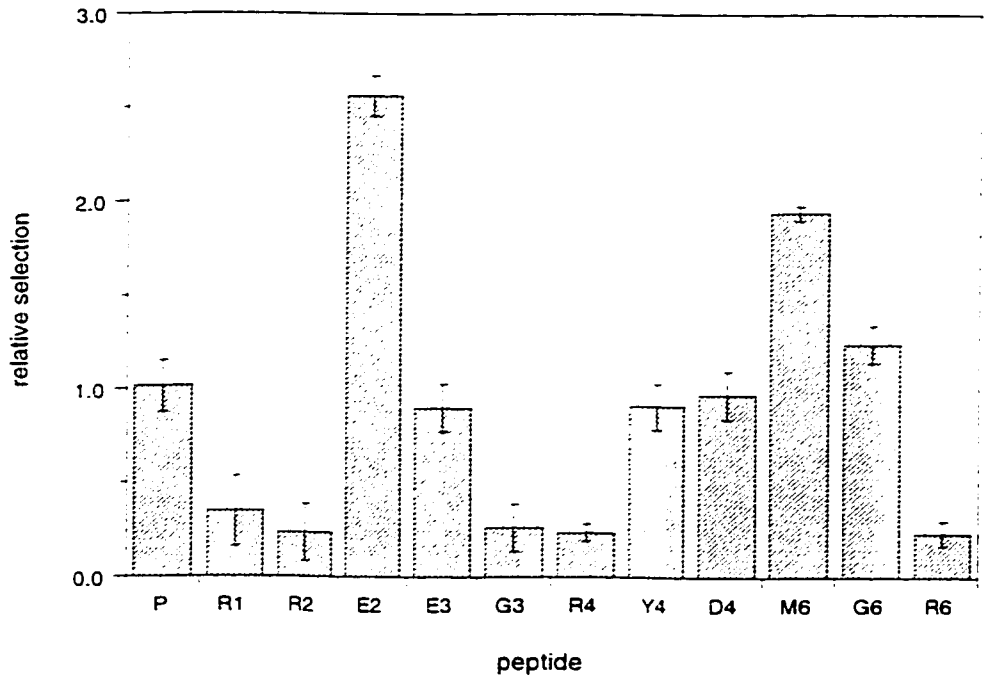


Fig. 7.5. **Affinity selection from the "mix of 12" library.** The indicated peptides were mixed together in approximately equal proportions and used for affinity selection, as described under "Results." In the peptide nomenclature, "P" refers to the parent sequence [DADE(F$_2$Pmp)L], and the other peptides are named according the residue changed from the parent sequence (*e.g.*, "R1" has an arginine substituted at position 1). Data were processed and expressed in the same way as for Fig. 7.3.

DISCUSSION

The published peptide library affinity selection protocols with SH2 and PDZ domains involved immobilizing the protein on a column and 'purifying' high affinity peptides by affinity chromatography (13, 14, 17, 24). An alternative approach is to have both the target protein and peptide library free in solution and to use size exclusion chromatography to separate protein–peptide complexes from free peptide (25). There are clear advantages and disadvantages to each approach. For example, the immobilized protein method can provide a means of concentrating a high affinity peptide present at very low concentration in a large volume of sample. As well, differences in relative affinities of peptides can be assessed by their elution profiles. However, incomplete washing or non-specific binding can result in a high background of unbound peptides; alternatively, bound peptides can dissociate during washing, reducing the recovery. Furthermore, immobilizing a protein may affect its ability to bind peptides efficiently. The size exclusion method has the advantage that both protein and ligand are free in solution and rapidly come to equilibrium; using a gel filtration spin column limits the separation time, minimizing dissociation of bound peptides. As well, the PTP does not need to be in a form that can be immobilized on a solid support, making it easy to adapt this protocol to any available PTP. However, this method does not permit concentration of ligands, and low-affinity peptides present at high concentrations can swamp out a scarce high-affinity peptide. With defined libraries these concerns are minimized, and thus we chose to use the size exclusion method for our analyses.

Combining mass spectrometric detection with the affinity selection protocol provides a rapid and inexpensive means to quantitate the various peptides. The high sensitivity of the mass spectrometer means that less sample is required as compared to direct sequencing of peptides, reducing the scale of the affinity binding reaction. One limitation of MS analysis is that the amino acids Leu/Ile and Gln/Lys are isobaric, so peptides containing these pairs of residues cannot be distinguished based on mass. Initial experiments using capillary-zone electrophoresis (CZE) to separate the peptides before infusion into the mass spectrometer suggest that Gln and Lys can be separated, but not Leu and Ile[5]. A second limitation is that this MS detection is restricted to peptide libraries with only one degenerate position, as the peptides in a library with more degenerate positions would no longer have unique masses. As such, the libraries are inherently biased by the residues at the fixed positions, and the combined effect of changes at multiple positions cannot be assessed. Preliminary experiments have been undertaken with libraries with two degenerate positions, using CZE to separate peptides before MS analysis. While the analysis is more complex, similar results were observed as for the libraries with one degenerate position[5], thus reinforcing the validity of using singly-degenerate libraries.

The validity of the affinity-selection technique employed here is apparent from comparing the results in Table 7.2 with the data from Fig. 7.5. In general, the order of the relative selections of the 12 peptides (Fig. 7.5) paralleled the order of their affinities (K_i's) for PTP1B (Table 7.2). The correlation is less strong with the libraries used in Fig. 7.3, although qualitatively there is a very good agreement in that peptides with a relative selection >1.0 have an affinity as strong or better than the parent peptide and those with a

[5] J. K., unpublished observations.

relative selection ≤1.0 have lower affinity. The weaker correlation in Fig. 7.3 may simply be a reflection of the quality and/or complexity of the libraries, since these libraries are not equimolar mixtures and are not as clean synthetically as the "mix of 12" used in Fig. 7.5.

The limited published peptide substrate specificity studies of PTP1B serve to validate further the results obtained here. Zhang *et al.* screened a panel of ten pY-peptides, identifying the sequence surrounding Tyr_{992} of the EGF receptor as a high-affinity substrate for PTP1B (5). Varying the peptide length and an alanine scan indicated the importance of acidic residues N-terminal to the pY residue in the DADE(pY)L peptide (5, 6). Our results confirm this and also demonstrate that replacement of Ala at the second position with an acidic residue significantly improves the affinity of the peptide for PTP1B. Furthermore, our results reveal a selection for aromatic residues (Phe and Tyr) at the position immediately N-terminal to pY, as well as some selectivity immediately C-terminal to pY, which were not suggested from the above studies.

An examination of the crystal structure of the Cys_{215}→Ser mutant complexed with the DADE(pY)L peptide (23) and additional modeling help to rationalize the results obtained here. The side chains of the Asp and Glu immediately N-terminal to pY form salt bridges with Arg_{47} of PTP1B, providing a structural reason for the preference for acidic residues at these positions. As well, modeling shows that the side chain of Phe or Tyr at the first position N-terminal to pY could be "sandwiched" by the Arg_{47} and Asp_{48} side chains without causing any severe steric hindrance; the resulting extensive van der Waals interactions may account for the selection observed for these residues. Modeling studies also reveal that acidic residues at the second position (Ala in the parent sequence) may be

able to interact with Arg_{45}, consistent with the strong selection for Asp and Glu from the X2 library, and the high affinity of the E2 peptide. The binding pocket for the residue at position 6 appears to be of an amphipathic nature, large enough to accommodate most neutral and acidic residues. Arg_{24} is sufficiently close to this binding pocket that it may interact unfavorably with the basic residues Arg and Lys, consistent with the negative selection observed for Arg at position 6.

While there are clearly structural reasons for sequence specificity, a number of observations suggest that the amino acid sequence surrounding the pY may not be the only determinant of PTP substrate specificity. This suggestion arises in part because of the apparent lack of specificity that many PTPs display towards *in vitro* substrates, although few detailed analyses have been done. It is likely that PTP localization and substrate proximity also play a role in substrate specificity, particularly for transmembrane PTPs and those containing protein association domains. However, other evidence suggests that at least some PTPs have their substrate specificity determined by the substrate sequence. For example, PTP-PEST has a very restricted substrate specificity *in vitro* and *in vivo* (26), and several MAP kinase phosphatases have been identified [for example, MKP-1 (27)] that are specific for MAP kinases. As well, studies with chimeras of SHP-1 and SHP-2 have suggested that there are intrinsic catalytic specificity differences between the two enzymes (28). In fact, substrate specificity of PTPs is probably a combination of localization that restricts the potential substrates and catalytic specificity that directs the PTP to its particular substrate(s). Thus, the use of a peptide library affinity

selection approach to identify the sequence specificities of PTPs should assist in the identification of their *in vivo* substrates.

Beyond defining sequence specificities, the tight binding F_2Pmp-containing peptide inhibitors identified in the affinity selection could themselves be useful tools in the study of signal transduction pathways. Preliminary results suggest that these peptides can show strong selectivity for a particular PTP, and a number of techniques have been reported to introduce such peptides into cells (for examples, see 29-31). Thus, F_2Pmp-containing peptides could be used to inhibit a PTP specifically *in vivo* and potentially uncover its involvement in a particular signaling pathway, similar to what has been achieved with protein kinase inhibitors (for examples, see 32-35). Furthermore, the present affinity selection technology can be adapted to other mixtures such as cell extracts and non-peptidyl combinatorial libraries, providing a powerful means to uncover novel ligands and inhibitors of PTPs.

ACKNOWLEDGMENTS

We thank R. L. Erikson for providing the full-length PTP1B cDNA clone, and B. Kennedy and P. Payette for the subcloning of the GST-PTP1B fusion. We also thank Z. Jia for helpful discussions on structural interpretations and for molecular modeling, A. English for critical reading of the manuscript, and the members of the Merck Frosst phosphatase group for their support and insights. G. H. was supported by a Medical Research Council/Pharmaceutical Manufacturers Association of Canada (MRC/PMAC) Health Program Studentship and a Stanley G. French Graduate Fellowship from Concordia

University, and J. K. was supported by a Natural Sciences and Engineering Research Council (NSERC) of Canada Industrial Research Fellowship.

REFERENCES

1. Sun, H., and Tonks, N. K. (1994) *Trends Biochem. Sci.* **19,** 480-485

2. Tonks, N. K. (1996) *Adv. Pharmacol.* **36,** 91-119

3. Hunter, T. (1995) *Cell* **80,** 225-236

4. Tonks, N. K., and Neel, B. G. (1996) *Cell* **87,** 365-368

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

6. Zhang, Z.-Y., Maclean, D., McNamara, D. J., Sawyer, T. K., and Dixon, J. E. (1994) *Biochemistry* **33,** 2285-2290

7. Cho, H., Krishnaraj, R., Itoh, M., Kitas, E., Bannwarth, W., Saito, H., and Walsh, C. T. (1993) *Protein Science* **2,** 977-984

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

9. Dechert, U., Adam, M., Harder, K. W., Clark-Lewis, I., and Jirik, F. (1994) *J. Biol. Chem.* **269,** 5602-5611

10. Dechert, U., Affolter, M., Harder, K. W., Matthews, J., Owen, P., Clark-Lewis, I., Thomas, M. L., Aebersold, R., and Jirik, F. R. (1995) *Eur. J. Biochem.* **231,** 673-681

11. 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

12. Kwon, M., Oh, M., Han, J., and Cho, H. (1996) *Biochem. Mol. Biol.* **29,** 386-392

13. 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., and Cantley, L. C. (1993) *Cell* **72,** 767-778

14. Songyang, Z., Fanning, A. S., Fu, C., Xu, J., Marfatia, S. M., Chishti, A. H., Crompton, A., Chan, A. C., Anderson, J. M., and Cantley, L. C. (1997) *Science* **275,** 73-77

15. Songyang, Z., Blechner, S., Hoagland, N., Hoekstra, M. F., Piwnica-Worms, H., and Cantley, L. C. (1994) *Curr. Biol.* **4,** 973-982

16. Songyang, Z., Carraway, K. L., Eck, M. J., Harrison, S. C., Feldman, R. A., Mohammad, M., Schlessinger, J., Hubbard, S. R., Smith, D. P., Eng, C., Lorenzo, M. J., Ponder, B. A. J., Mayer, B. J., and Cantley, L. C. (1995) *Nature* **373,** 536-539

17. Kelly, M. A., Liang, H., Sytwu, I.-I., Vlattas, I., Lyons, N. L., Bowen, B. R., and Wennogle, L. P. (1996) *Biochemistry* **35,** 11747-11755

18. Burke, T. R., Jr., Kole, H. K., and Roller, P. P. (1994) *Biochem. Biophys. Res. Comm.* **204,** 129-134

19. Solas, D., Hale, R. L., and Patel, D. V. (1996) *J. Org. Chem.* **61,** 1537-1539

20. Huyer, G., Liu, S., Kelly, J., Moffat, J., Payette, P., Kennedy, B., Tsaprailis, G., Gresser, M. J., and Ramachandran, C. (1997) *J. Biol. Chem.* **272,** 843-851

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

22. Chen, L., Wu, L., Otaka, A., Smyth, M. S., Roller, P. P., Burke, T. R., Jr., den Hertog, J., and Zhang, Z.-Y. (1995) *Biochem. Biophys. Res. Comm.* **216,** 976-984

23. Jia, Z., Barford, D., Flint, A. J., and Tonks, N. K. (1995) *Science* **268,** 1754-1758

24. Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R. A., and Cantley, L. C. (1994) *Mol. Cell. Biol.* **14,** 2777-2785

25. Williams, K. P., Evans, D. M., Rosenberg, S., and Jindal, S. (1996) in *Techniques in Protein Chemistry VII* (Marshak, D. R., ed.) pp. 57-64, Academic Press, Inc., San Diego

26. Garton, A. J., Flint, A. J., and Tonks, N. K. (1996) *Mol. Cell. Biol.* **16,** 6408-6418

27. Sun, H., Charles, C. H., Lau, L. F., and Tonks, N. K. (1993) *Cell* **75,** 487-493

28. Tenev, T., Keilhack, H., Tomic, S., Stoyanov, B., Stein-Gerlach, M., Lammers, R., Krivtsov, A. V., Ullrich, A., and Böhmer, F.-D. (1997) *J. Biol. Chem.* **272,** 5966-5973

29. Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L. L., Pepinsky, B., and Barsoum, J. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 664-668

30. Kole, H. K., Garant, M. J., and Bernier, M. (1996) *J. Biol. Chem.* **271,** 14302-14307

31. Giorgetti-Peraldi, S., Ottinger, E., Wolf, G., Ye, B., Burke, T. R., Jr., and Shoelson, S. E. (1997) *Mol. Cell. Biol.* **17,** 1180-1188

32. Birchall, A. M., Bishop, J., Bradshaw, D., Cline, A., Coffey, J., Elliott, L. H., Gibson, V. M., Greenham, A., Hallam, T. J., Harris, W., Hill, C. H., Hutchings, A., Lamont, A. G., Lawton, G., Lewis, E. J., Maw, A., Nixon, J. S., Pole, D., Wadsworth, J., and Wilkinson, S. E. (1994) *J. Pharmacol. Exp. Ther.* **268,** 922-929

33. Lee, J. C., Laydon, J. T, McDonnell, P. C., Gallagher, T. F., Kumar, S., Green, D., McNulty, D., Blumenthal, M. J., Heys, J. R., Landvatter, S. W., Strickler, J. E., McLaughlin, M. M., Siemens, I. R., Fisher, S. M., Livi, G. P., White, J. R., Adams, J. L., and Young, P. R. (1994) *Nature* **372,** 739-746

34. Alessi, D. R., Cuenda, A., Cohen, P., Dudley, D. T., and Saltiel, A. R. (1995) *J. Biol. Chem.* **270,** 27489-27494

35. Meydan, N., Grunberger, T., Dadi, H., Shahar, M., Arpaia, E., Lapidot, Z., Leeder, J. S., Freedman, M., Cohen, A., Gazit, A., Levitzki, A., and Roifman, C. M. (1996) *Nature* **379,** 645-648

CHAPTER 8

Discussion and Perspectives

Cellular signal transduction pathways appear quite complex at first glance: a single stimulus can lead to the involvement of dozens or even hundreds of proteins and the modulation of multiple signal cascades, all leading to a specific regulated response by the cell. Upon further analysis, though, it becomes apparent that certain fundamental themes in signaling are repeated. For example, signaling pathways that involve tyrosine phosphorylation all use the same basic machinery to elicit a cellular response. All pathways start with a protein tyrosine kinase (PTK) activity, either intrinsic to a transmembrane receptor or associated with its cytosolic portion, that becomes activated upon ligand binding. Tyrosine phosphorylation of substrates can directly modulate an enzyme's activity to continue the signal propagation. Alternatively, tyrosine phosphorylation can create binding sites for SH2 domains to direct formation of specific protein complexes. Protein tyrosine phosphatases (PTPs) also play an important role, both in opposing PTK activity to downregulate the signal and in activating other proteins to propagate the signal. Conservation of signaling machinery is more efficient and economical than evolving a distinct mechanism to transmit each of the signals with which the cell is presented. However, with so much related signaling machinery, how does a cell maintain specificity in the transmission of various signals? The results presented in this thesis help to shed some light on the determinants of SH2 domain and PTP specificity, two of the major players in tyrosine phosphorylation signaling pathways.

SH2 Specificity Studies

The human PTP SHP-2 was cloned at the same time in a number of laboratories, including ours (1-6), and had been demonstrated to bind through its SH2 domains to the PDGF receptor *in vivo* (7). The wide tissue distribution of SHP-2 implied that it might bind to other targets and be involved in additional signaling pathways. In order to help identify other potential *in vivo* targets, studies were undertaken to characterize the binding of the SH2 domains of SHP-2 and to uncover unique specificity determinants that allow them to maintain specific signaling interactions.

In Chapter 2, two approaches using competitive binding assays were developed to measure the relative affinities of pY-peptides for the SH2 domains compared to the affinity for a peptide modeled on the *in vivo* binding target surrounding Tyr_{1009} of the PDGF receptor: DTSSVL(pY)TAVQPN. The first technique used surface plasmon resonance (BIAcore), in which the peptide was immobilized on a sensor chip and binding of the SH2 domains was measured in real time from changes in refractive index on the chip's surface when the SH2 domains bind. The other technique immobilized the SH2 domains as a GST fusion on GT-Sepharose beads. Binding was measured as the amount of a ^{32}P-labeled peptide (based on the same PDGF receptor sequence used above) that remained associated with the beads after washing. In both systems, relative affinities of peptides could be determined by their ability to compete with the PDGF receptor peptide for binding to the SH2 domain.

There are a number of advantages and disadvantages associated with each of the techniques used to determine peptide affinities. The BIAcore provides a direct means for

following the binding in real time and measuring the kinetics of association and dissociation. Since a physical property (*i.e.*, change in refractive index) is measured, labeling of the protein or ligand for detection is not required. However, because the SH2 domain is binding to a peptide immobilized on a surface, the dynamics of binding are different than when protein and ligand are both free in solution. As well, because the SH2 domain is in a buffer that flows across the immobilized peptide, a true equilibrium binding situation does not exist. Because of these factors, only dissociation constants in the range of 10^{-5} to 10^{-11} M can be determined with confidence. Compared to the BIAcore, the radioactive competitive binding assay only gives relative affinities of peptides in terms of their IC_{50} values and not true K_d values. Although the assay is an equilibrium binding assay, bound peptide is lost during the washes and will affect the results. It is clearly desirable to use both methods because of these technical concerns: if similar data are obtained with both systems, there is obviously more confidence in the results.

Binding competitions with truncated versions of the PDGF receptor peptide were performed by both methods to define the minimum sequence required for high-affinity binding to the SH2 domains of SHP-2, which was found to be VL(pY)TAV. An Ala scan, in which each of the residues (except pY) was replaced individually with Ala, indicated the importance of the pY+1 and pY+3 residues (*i.e.*, the first and third residues C-terminal to the pY) as well as the pY-2 residue. The requirement for N-terminal residues was unexpected: for all other SH2 domains studied, the sequence specificity was directed exclusively by residues C-terminal to pY, and in particular the pY+1 and pY+3 residues (8, 9).

The study in Chapter 2 identified the minimum sequence required for high affinity binding and suggested some important residues in that sequence. To obtain additional information about the optimal sequence and the range of sequence variations tolerated, an affinity selection technique from peptide libraries was developed in Chapter 3. The methods used were modeled on those published by Songyang *et al.* (10) to determine SH2 domain binding specificity at the pY+1 to pY+3 positions. Because we were particularly interested in assessing the specificity at the pY-2 position, a peptide library with only this position degenerate was required. As such, it was possible to explore the use of mass spectrometry (MS) for analysis of the affinity selection instead of peptide sequencing as previously used. In a peptide library with a single degenerate position, all peptides have a unique mass (except for Leu/Ile and Gln/Lys) and therefore can be identified (and quantitated) based on their mass. MS has a number of advantages over sequencing, including the speed of MS analyses and the requirement for less sample because of the greater sensitivity of MS. However, the continuous flow–liquid secondary ion MS (CF-LSIMS) protocols needed to be developed, optimized, and validated as CF-LSIMS is not typically used in a quantitative fashion as required in this study. As a control for the MS analysis, samples were also analyzed by peptide sequencing for comparison, since the sequencing methodology has been validated (10).

Before affinity selection from peptide libraries was attempted, pilot experiments with a single peptide [TSSVL(pY)TAVQP or VL(pY)TAV] were performed to assess column binding capacity and efficiency of phenyl phosphate elution. These experiments used HPLC to quantitate the peptide; however, numerous difficulties were encountered

with the HPLC (in particular an intermittent autosampler problem) that hampered the analysis of results. As such, the extent of the development experiments was limited. From the column binding titration (Fig. 3.1), the capacity appeared to be only about 15% of that expected based on 1:1 peptide:SH2 domain binding. A number of factors may account for this apparent low capacity. First, only a fraction of the SH2 domains may have been competent for binding peptide, perhaps because of denaturation or accessibility problems associated with the immobilization. As well, bound peptide may have dissociated during the column washes, reducing the apparent capacity of the column. It is also possible that more time was required for the binding to reach equilibrium. Finally, the phenyl phosphate elution may not have quantitatively eluted peptide. Because of the HPLC technical problems, these factors could not be properly assessed. However, as the system appeared to be reproducible, it was decided to proceed with the library affinity selection.

For the affinity selection, two peptide libraries were made based on the sequence TSSVL(pY)TAVQP: one in which the pY-2 position was replaced with all naturally-occurring L-amino acids (except Cys and Trp); and another in which the pY+3 position was similarly substituted. The pY+3 library was used as a control because the selectivity at this position had already been examined, the preferred residues being Val, Ile, Leu, and Pro, all with similar selectivities (10). A number of difficulties were initially encountered with the quality of the peptide libraries in terms of widely-varying yields of the peptides (including some peptides that were completely absent). Peptide libraries were finally acquired from a different manufacturer that were of much better quality and could be used with confidence for the affinity selection.

The CF-LSIMS analysis worked well in terms of identifying each of the library peptides by its mass, and peak heights could be readily obtained from the spectra. In order to determine if peak height was a quantitative measure of the amount of each peptide, experiments with peptide standards needed to be carried out. Unfortunately, limited MS access restricted the number of experiments that could be analyzed, and as such these important controls were not performed. Differences in peak shapes suggested that peak area might be better correlated with peptide quantity compared to peak height; however, the JEOL MS analysis software only permitted peak height calculations. Fragmentation of peptide ions in the relatively high-energy LSIMS ionization source was also a problem. The peak heights of the singly-protonated peptide molecular ions were used to quantitate peptides, but if the extent of fragmentation varied between the peptides, then MS peak height would not be a quantitative measure of the amount of each peptide. The capillary-LC separation conditions may also have affected the quantitation and reproducibility. For example, the TFA in the mobile phase gave a high spectral background, and the high-aqueous eluent resulted in unstable MS signals. Including an internal standard peptide in the MS samples would permit the data to be normalized and compared more easily between runs, allowing reproducibility to be better assessed and averaging of multiple runs. One preliminary experiment was performed with an internal standard, but further work with the standard could not be carried out because of a lack of MS accessibility.

In spite of all these problems, some interesting preliminary results were obtained with the libraries. At the pY+3 position, there appeared to be a very strong selection for Val, the residue at this position in the PDGF receptor. MS also suggested a weak

selection for Leu/Ile that was not observed by sequencing. In contrast, the published study with the N-terminal SH2 domain of SHP-2 (10) indicated similar preferences at this position for Val, Ile, Leu, and Pro. These differences might reflect the limitations of the MS quantitation method used here, and the fact that our sequencing data were only from one experiment. As well, a fusion protein with both SH2 domains of SHP-2 was used in the affinity selection in Chapter 3, as opposed to only the N-terminal SH2 domain in the published study (10), which may contribute to a different peptide selection. Finally, the peptide libraries used in Chapter 3 have all positions fixed except one, which may influence the selectivity at the degenerate position.

The affinity selection results at the pY-2 position suggested clear preferences but a broad specificity, supporting the original observation that the residue at this position was important in defining binding specificity. Six individual pY-containing peptides were synthesized with substitutions at the pY-2 position to test the predictions from the affinity selection process. There was reasonable but not complete agreement between the affinity selection results and the relative affinities of the six peptides (Table 3.2), suggesting that the method was valid but needed to be optimized as described above. Nonetheless, the analysis of individual peptides did confirm the importance of the pY-2 position for binding as certain substitutions had a marked effect on peptide affinity.

The structural basis for the pY-2 selectivity of the N-terminal SH2 domain of SHP-2 was demonstrated in Chapter 4. Replacement of the αA2 Gly residue of the SH2 domain with Arg, the consensus residue at this position in most SH2 domains, eliminated the requirement for residues N-terminal to the target pY in mediating high-affinity binding.

A similar specificity for N-terminal residues has recently been reported for the SH2 domains of SHP-1 (11) which also have a Gly residue at the αA2 position. It is likely that the structural basis for this requirement is the same for SHP-1 as for SHP-2.

These observations on the role of the pY-2 residue have a number of important implications. The general consensus is that SH2 specificity is directed solely by residues C-terminal to the target pY (8, 9). In fact, this belief is so entrenched that most experiments to study SH2 specificity are biased in this direction. For example, a number of structural studies of SH2 domains with bound peptides have been carried out with pY-peptides having no residues N-terminal to pY (12-14). Even though the structure of the N-terminal SH2 domain of SHP-2 bound to the PDGFR 1009 peptide (15) shows pY-2 residue involvement, the contact was ignored in most of the discussion which focused instead on the C-terminal residue contacts. Perhaps the clearest case of the bias towards C-terminal residue interactions is in the peptide library affinity selection studies performed by Songyang *et al.* (10), who used peptide libraries with only the pY+1, pY+2, and pY+3 positions degenerate. Consequently, the effects of substitutions at other positions in the peptide sequence were not assessed. While for the majority of SH2 domains the exclusive involvement of C-terminal residues in defining selectivity appears to be correct, by assuming that this is a hard-and-fast rule for all SH2 domains the involvement of N-terminal residues in directing specificity for certain SH2 domains, like those of SHP-2 and SHP-1, will be missed. For example, the SH2 domain of Ctk also has an αA2 Gly and therefore may have a similar specificity for residues N-terminal to its target pY residues.

Defining a consensus sequence for high-affinity SH2 binding has been useful in predicting potential *in vivo* binding sites by searching a cDNA or protein database. With this additional information about SHP-2 and SHP-1 selectivity, such search strategies need to be modified to include the consensus N-terminal residue(s). For example, most of the potential binding sites predicted by Songyang *et al.* for SHP-2 (10) and SHP-1 (16) can be eliminated based on the pY-2 residue. It is perhaps not surprising that in order to maintain specificity with at least 100 different SH2 domains, additional modes of SH2 domain binding exist such as that defined by SHP-2 and SHP-1. Recognition of residues in addition to those at the pY+1 and pY+3 positions should provide more scope for fine-tuning SH2 domain–protein interactions.

An important caveat in these studies is that *in vitro* specificity may not always reflect the true cellular situation. SH2 domain-directed associations will depend on many factors, in particular the relative concentrations and affinities of SH2 domain-containing proteins that compete for the same or mutually-exclusive binding sites. As well, identification of a high-affinity consensus binding sequence does not mean that all *in vivo* targets have this sequence. Many sequences that deviate from the high-affinity consensus still have reasonable affinity for the SH2 domain. There are circumstances where it may be advantageous for an SH2 domain to bind to a sequence with moderate affinity such that, for example, the signaling interaction does not persist too long. As in any binding interaction, increasing the concentration of one of the binding partners will drive binding. Thus, in a case where a high amount of a particular SH2 domain is present, modulating the affinity for the binding site could regulate the extent of the interaction. The *in vitro*

studies are certainly informative in describing differences between SH2 domains, but the highly complex situation in the cell must be kept in mind when considering potential SH2 domain-directed associations.

PTP Specificity Studies

Much less is known about the sequence specificity determinants of PTPs compared to SH2 domains as fewer studies have been undertaken. In general, the same principles apply as for SH2 domains, since PTPs, like SH2 domains, bind to pY in the context of the surrounding sequence. PTPs are distinct from SH2 domains in that they hydrolyze the pY, and therefore to measure binding either a mutant enzyme that binds but does not cleave the substrate, or a nonhydrolyzable pY analogue, must be used. Catalytic efficiency of PTPs towards different pY-peptide substrates can also be used to evaluate their substrate selectivity. As well, the catalytic activity can be exploited to study specificity with inhibitors, and in particular, inhibitors that show selectivity for one enzyme over another. By determining how an enzyme interacts with an inhibitor, and the structural features of the inhibitor that are responsible for its selectivity, mechanistic insight can be gained and related to substrate interactions.

For PTPs, no specific inhibitors are known. In fact, few inhibitors at all are known, and the most commonly used are the non-specific inhibitors vanadate and pervanadate. Previous studies in our laboratory suggested that these inhibitors have different inhibitory properties. Thus, an understanding of the mechanisms of inhibition by these two compounds may reveal details of how they are recognized by PTPs and by

extension how substrates are recognized. As such, the inhibition of PTPs by vanadate and pervanadate was explored in Chapter 5.

The complexities of vanadate and pervanadate chemistry quickly became clear in this study, which may have contributed to the fact that no such study had been reported in the literature in spite of the widespread use of these compounds as PTP inhibitors and insulin mimetics. Vanadate, for example, interacts with many organic compounds due to its propensity for forming reversible esters with hydroxyl and sulfhydryl groups (17). As such, the choice of buffer is very important for a careful kinetic study as several buffers have been shown to interact strongly with vanadate, including Tricine and triethanolamine (18, 19). For our studies, HEPES was used as it has been reported not to interact with vanadate (18, 19). A careful kinetic analysis with PTP1B showed that vanadate is a competitive inhibitor as had been expected but never actually shown. Structurally, vanadate is a phosphate analogue and therefore is predicted to interact at the active site where the phosphate of the substrate pY normally binds. In fact, the recently-described crystal structures of the *Yersinia* PTP (20) and a bovine low molecular weight PTP (21) complexed with vanadate show vanadate bound to the catalytic Cys residue in a five-coordinate structure that mimics the transition state.

Another important interaction of vanadate is its chelation by EDTA. This interaction had been reported over 30 years ago in the literature (22, 23), yet many researchers persist in using vanadate as a PTP inhibitor in buffers containing EDTA, presumably out of ignorance of this interaction. In Chapter 5, the effect of EDTA on vanadate inhibition of PTPs was clearly demonstrated: the presence of 1 mM EDTA in the

assay buffer completely blocked inhibition by 10 μM vanadate; similarly, addition of EDTA to vanadate-inhibited PTP1B completely reversed the inhibition. Because of the vanadate chelation by EDTA, many erroneous conclusions have been made in the literature. For example, widely varying inhibitory potencies of vanadate towards PTPs have been reported, which in fact simply reflects the presence of EDTA in the assay buffer that chelates the vanadate and reduces the concentration of free vanadate for inhibition. Perhaps more importantly is the effect in studying pY-protein levels *in vivo*. In such studies, cells are lysed and pY-proteins are either immunoprecipitated or identified by an anti-pY antibody Western blot. Vanadate is generally included in the cell lysis buffer to inhibit endogenous PTPs and prevent non-specific dephosphorylation of pY-proteins. However, EDTA is often also included, and as a result the vanadate will be chelated and endogenous PTPs will not be inhibited, potentially resulting in misleading analyses of pY-proteins.

The chemistry of pervanadate is even more complex than vanadate. The term "pervanadate" refers to the variety of complexes formed between vanadate and hydrogen peroxide. These compounds form in rapid equilibrium and once formed only slowly dissociate with a half-life of several hours (24). In the course of the study, it became apparent that pervanadate is not stable in the presence of reducing agents: DTT included in the assay buffer, for example, very quickly reduced the pervanadate to vanadate. It is necessary to include a thiol reductant in a PTP assay buffer to prevent inactivation of the enzyme by air oxidation of the catalytic Cys. Thus, in order to establish kinetic parameters for pervanadate inhibition, assays would have to be carried out anaerobically which was

beyond the scope of this study. Furthermore, because pervanadate is always a mixture of species, including uncomplexed vanadate and hydrogen peroxide, it would be difficult to assign inhibitory properties to the individual complexes.

When pervanadate is added to a PTP assay, it is able to inhibit a portion of the enzyme irreversibly before it is converted to vanadate by DTT. As shown by mass spectrometry, the irreversible inhibition results from pervanadate-mediated oxidation of the catalytic Cys. This mode of inhibition is clearly different than vanadate, which inhibits PTP1B reversibly, and has important implications for the use of pervanadate as a PTP inhibitor. For example, if irreversible inactivation of a PTP is not desirable, then pervanadate should not be used. If pervanadate is used, it should be in the absence of thiol reductants to prevent its destruction.

Elucidating the different modes of vanadate and pervanadate inhibition did not, unfortunately, provide much insight into how PTPs recognize these compounds. Since vanadate is a phosphate analogue, it is not surprising that it is a competitive inhibitor of PTP1B. However, it is not clear how pervanadate interacts with the active site of PTPs. From an examination of the structures of the various pervanadate complexes, it is not obvious how the affinity for PTPs is defined. The experiment in Fig. 5.8 suggested that diperoxovanadate is a more potent inhibitor of PTP1B than monoperoxovanadate. However, whether this reflects a difference in affinities of the two complexes for the enzyme or simply the greater oxidizing potential of the diperoxo species (complexed with two peroxide molecules instead of one) is unclear.

Peroxovanadium compounds have also been shown to be potent PTP inhibitors (25). These compounds are stable pervanadate derivatives that contain an oxo ligand, one or two peroxo ligands, and an ancillary ligand in the inner coordination sphere of the vanadium. The nature of the ancillary coordinating ligand has been shown to influence significantly the inhibitory potency of the compounds and is suggested to play a role in directing specificity towards a particular PTP. The mechanism of inhibition by peroxovanadium compounds has not been studied but it presumably involves oxidative inactivation similar to pervanadate as described above. If the ancillary ligand is involved in directing PTP specificity, a structure–activity relationship of a series of peroxovanadium compounds has the potential to uncover important determinants of PTP recognition.

The oxidative mechanism of pervanadate inhibition may have parallels in *in vivo* PTP regulation. Intracellular redox modulation has been suggested to control many tyrosine phosphorylation signaling pathways (26-30), presumably through oxidation of the catalytic Cys of PTPs. Many extracellular signals lead to an increase in intracellular reactive oxygen species like hydrogen peroxide (26-30). *In vitro*, pervanadate acts as a delivery system to direct its complexed hydrogen peroxide to the active site of PTPs to oxidize the catalytic Cys. It is intriguing to speculate that cells may have a delivery system to target oxidizing agents to the catalytic Cys of PTPs in a controlled and regulated fashion. The mechanism may be analogous to that described in the addendum to Chapter 5, in which hydrogen peroxide (generated in the imidazole–DTT assay system) complexed with vanadate to inactivate PTP1B. Mild Cys oxidation (*i.e.*, to the sulfenic acid, –SOH) would result in a species that is more easily reduced, providing a reversible means of

regulating PTP activity. Furthermore, a putative oxidant delivery system with intrinsic specificity for a particular PTP would lead to specific inactivation by a non-specific oxidant such as hydrogen peroxide. Further investigations of the roles of reactive oxygen species in regulating PTPs and the mechanisms involved are clearly required and should be very informative.

PTP specificity can also be studied by determining catalytic efficiency towards a variety of pY-peptide substrates to assess the effect of sequence differences. The K_m of pY-peptide substrate reflects its binding affinity (K_d) for the enzyme. As shown in Fig. 1.3 and diagrammed in the scheme below, the PTP catalytic mechanism involves two steps: transfer of the phosphate from the substrate pY to the catalytic Cys of the PTP, and hydrolysis of the thiol-phosphate intermediate.

$$E + S \underset{}{\overset{K_d}{\rightleftharpoons}} ES \xrightarrow{k_2} E\text{–}OPO_3^{2-} \xrightarrow{k_3} E + P_i$$

product

The relation between K_m and K_d is defined by the following expression:

$$K_m = K_d \frac{k_3}{k_2 + k_3}$$

Thus, the first order rate constants k_2 and k_3 must be known to calculate a true K_d from the K_m. However, since k_3 is independent of the substrate's structure, and assuming k_2 is minimally affected by substrate, then the K_m of a range of substrates will always be related to the K_d by the same factor. Thus, a substrate's K_m is a reasonable approximation of its K_d.

In Chapter 6, three transmembrane PTPs were examined (CD45, LAR, and PTPß) with a number of pY-peptide substrates in an attempt to define determinants of substrate

specificity. For these studies, the cytosolic portions of the enzymes that included the catalytic domains were expressed in *E. coli* and purified. The approach taken was modeled on that of Zhang *et al.* (31) in which a high-affinity pY-peptide substrate [DADE(pY)LIPQQG] was identified for the *Yersinia* PTP and PTP1B, and then an Ala scan was performed in which each residue was substituted, one at a time, with Ala to assess the importance of the individual residues in determining the enzymes' catalytic efficiencies towards the parent peptide. Ala was chosen because its small, neutral side chain (simply a methyl group) would not be expected to make significant interactions with the PTP while still maintaining any peptide backbone interactions. These studies demonstrated that replacement of the acidic residues N-terminal to pY with Ala reduced the enzymes' catalytic efficiencies, suggesting that the enzymes required these residues for substrate recognition and binding.

The primary goal was to study the specificity of CD45, a key enzyme in T-cell signaling that activates $p56^{lck}$ by dephosphorylating it at Tyr_{505}. Thus, a peptide based on the sequence surrounding this Tyr [TEGQ(pY)QPQP] was used. CD45 displayed standard Michaelis-Menten saturation kinetics with this peptide substrate, from which kinetic parameters (K_m, V_{max}, k_{cat}/K_m) could be calculated. The results of the Ala scan indicated that none of the Ala substitutions made an appreciable difference in the catalytic efficiency of CD45 towards the peptides. In fact, none of the Ala substitutions resulted in a peptide with lower affinity for the enzyme. The biggest change was observed with two double-Ala substitutions, TEGA(pY)APQP and TEGQ(pY)APAP, which were both ~2.5-fold better substrates than the parent sequence.

These results were somewhat unexpected and suggested that the sequence chosen was not the optimal high-affinity sequence for CD45. For an Ala scan to be successful, a high-affinity peptide substrate must be used as a starting point to assess the negative effect of changes in the sequence that would identify critical residues in the parent sequence. However, none of the Ala substitutions had a negative effect on catalytic efficiency towards the peptide, and in fact some of the substitutions resulted in better substrates. If Tyr_{505} of $p56^{lck}$ is an *in vivo* target of CD45, then why would this be a suboptimal sequence for the enzyme? The simplest explanation is that *in vivo* it is not necessary for this sequence to be the best substrate for CD45. CD45 is a transmembrane PTP peripherally associated with the T cell receptor (TCR) complex. Upon TCR activation, enzymes including $p56^{lck}$ become associated with the TCR complex, bringing them in proximity to CD45. Thus, CD45 may not require strict intrinsic specificity as it does not need to seek out its substrates. Rather, specificity could be a result of localization and access to substrates instead of being intrinsic to the enzyme.

Another possibility relates to $p56^{lck}$ regulation. If the Tyr_{505} sequence were a high affinity substrate for CD45, then Tyr_{505} would be very efficiently dephosphorylated (and $p56^{lck}$ therefore activated) by CD45. As the phosphotyrosine balance at Tyr_{505} is controlled by the competing activities of a PTK and CD45, the extent of phosphorylation will be adjusted by the relative activities of the two enzymes towards this site which in turn controls the proportion of activated $p56^{lck}$. The sequence surrounding Tyr_{505} may affect the enzymes' relative activities, providing a means of fine-tuning this balance. Finally, it must be kept in mind that the kinetic studies in Chapter 6 were performed with a

recombinant form of CD45 which consisted of the cytosolic portion only (containing the PTP domains). It is possible that the extracellular domain plays a role in regulating or modulating the intracellular PTP activity and that the full-length enzyme may display different catalytic efficiencies towards the pY-peptides studied.

To determine if other PTPs would show the same specificity behaviour as CD45, the PTPs LAR and PTPß were examined. Unfortunately, kinetic parameters could not be determined for LAR because the K_m's were too high to be determined in the kinetic assays due to substrate solubility and limited substrate supply. However, for PTPß, similar results were observed as for CD45. Again, none of the Ala substitutions resulted in a peptide with lower affinity for PTPß. In fact, almost all peptides were within 1.5-fold of the parent sequence in terms of the catalytic efficiencies of PTPß; one peptide showed a two-fold increase. Again, similar conclusions can be drawn as for CD45. It is likely that this is a suboptimal substrate for PTPß and therefore an Ala scan will not be overly informative. Like CD45, PTPß is a transmembrane PTP, and it may not require a great deal of intrinsic sequence specificity as its restricted membrane localization may limit the substrates to which it has access. Finally, the kinetic studies were carried out with a recombinant form of PTPß containing only the soluble cytosolic domain; the extracellular domain may play a role in regulating the enzyme's specificity and activity which would help to define its specificity.

The difficulties in assessing determinants of substrate specificity of PTPs using individual pY-peptide substrates are apparent from the studies in Chapter 6. The individual peptides are expensive, and as observed with a few of the peptides the

interpretations can be complicated by the presence of trace inhibitors, even with purified commercial peptides. Furthermore, it is labourious to determine kinetic parameters for many separate peptides. Finally, only a limited number of sequences can practically be analyzed: therefore, important determinants may be missed. Clearly a less restricted approach is required to probe sequence specificity more completely. Affinity selection from peptide libraries is an ideal way to screen many different sequence variations simultaneously. As described above, this approach has been successfully applied to SH2 domains. Extending this method to PTPs required the synthesis of peptide libraries in which the pY residue is replaced by a nonhydrolyzable pY analogue. Difluorophosphono-methyl phenylalanine (F_2Pmp) has been shown to be a very effective pY analogue (32); in fact, F_2Pmp-peptides generally have increased affinity for PTPs as compared to pY-peptides because of additional van der Waals interactions and hydrogen bonding due to the methylene fluorines of F_2Pmp (33). F_2Pmp-containing peptide libraries were synthesized at Merck Frosst after successfully developing synthetic routes based on published procedures (34).

Rather than simply repeating the SH2 affinity selection method with PTPs, we sought to introduce several improvements to the technique. The published methods with SH2 domains used immobilized SH2 domains on a column to affinity-purify peptides that were eluted from the column by competition at the pY binding site or by a non-specific eluent (10, 16, 35). This approach was followed in Chapter 3, revealing a number of limitations: the protein must be in a form that can be immobilized (such as a GST fusion); non-specific binding of peptides can be a problem; bound peptide can be lost during

column washing; and immobilization can affect the efficiency and kinetics of binding. As such, we explored the use of gel filtration to separate enzyme–peptide complexes from free peptide in solution. Secondly, quantitative analysis of SH2 domain affinity-selected peptides in the published protocols was accomplished by sequencing, which is time consuming and expensive. We had previously attempted in Chapter 3 to apply liquid secondary ion–mass spectrometry (LSIMS) to peptide analysis: the experience gained in those experiments led to the different approaches in this study that proved more successful.

The success of the affinity selection protocol and the MS analysis was clearly demonstrated from the analysis of individual peptides. The inhibitory potencies of a number of F_2Pmp-peptides were determined and the rank order correlated completely with their relative affinities for PTP1B. The affinity selection predictions also paralleled the catalytic efficiency of PTP1B towards individual pY-peptide substrates. Furthermore, molecular modeling using the crystal structure of PTP1B complexed with the EGF receptor hexapeptide (36) provided structural rationalizations for the sequence specificities observed at all positions.

The success of the study outlined in Chapter 7 as compared to the SH2 domain study in Chapter 3 is likely a reflection of the gel filtration separation and the improved MS analysis. In Chapter 7, the peptide libraries were mixed with PTP1B in solution and gel filtration was used to separate enzyme–peptide complexes from free peptide, while in Chapter 3 the SH2 domains were immobilized on a column and used to fish out high-affinity peptides from the libraries. As such, the problems in Chapter 3 associated with the

column binding, washing, and elution were avoided in Chapter 7. Furthermore, in Chapter 7 electrospray ionization (ESI) MS was used which is a gentler ionization technique than LSIMS used in Chapter 3, minimizing fragmentation of the peptide ions. Under the ESI conditions used (in negative ion mode), the peptide ions were almost exclusively singly-deprotonated, so no loss of signal resulted due to the formation of multiply-charged ions. Finally, the use of the small molecule trap prior to MS in Chapter 7 eliminated the problems associated with the capillary-LC separation conditions in Chapter 3 that may have negatively affected the LSIMS analysis.

The validation of the peptide affinity selection technique with PTP1B opens the door to apply this system to other PTPs. For best results, a sequence with a reasonable (sub-micromolar) affinity for the enzyme is required upon which the libraries are based. With PTP1B, the EGF receptor sequence surrounding Tyr_{992} had previously been identified as a high-affinity substrate (31, 37) and therefore provided a starting point for the affinity selection libraries. If a reasonable peptide substrate is not known for a PTP of interest, the affinity selection protocol can be used in different ways to identify such a peptide. One approach would be to use a more degenerate library based on a 'neutral' sequence such as a stretch of Ala residues in which multiple positions are degenerate. While such a library is not compatible with the quantitative MS analysis employed here because of the non-unique peptide masses, potential high-affinity peptides may be selected and identified by sequencing through tandem MS. Another approach would be to create a library by mixing a variety of unrelated F_2Pmp-peptides with different masses representing a range of sequences. From the affinity selection results with such a library, a high-affinity

peptide may be identified as a starting point for making individual libraries, or a consensus sequence might be suggested from comparing the selected peptides.

Defining the sequence specificities of PTPs may provide information about potential *in vivo* substrates. However, there are clearly limitations to this approach. For example, the peptide substrate studies with CD45 in Chapter 6 suggested that the chosen sequence around Tyr_{505} of $p56^{lck}$ was not optimal for CD45, even though $p56^{lck}$ is an *in vivo* substrate of the enzyme. This result underscores the fact that *in vivo*, PTP localization may influence substrate specificity which is not revealed by *in vitro* specificity studies.

Knowledge of sequence specificities can be useful in many other ways, particularly in the development of specific PTP inhibitors. Similarities and differences in the high affinity sequences recognized by individual PTPs can be exploited in the rational design of specific inhibitors. As well, knowing which PTPs have a very restricted specificity and which have a broad specificity will help in addressing the on-going debate about intrinsic catalytic specificity *vs.* localization as the primary determinant of PTP specificity. The F_2Pmp-peptides themselves may be useful tools for studying signaling pathways, since peptides of the appropriate sequence are potent PTP inhibitors [for example, the best peptide from the study in Chapter 7, EEDE(F_2Pmp)M, has a K_i of 7 nM]. Thus, specificity can be dialed in by exploiting differences in the preferred residues recognized by PTPs surrounding the F_2Pmp. By using techniques for introducing peptides into the cell (for example, 38-40), the effects of specific peptide inhibitors on signaling pathways could be determined. Such studies may reveal the specific involvement of particular PTPs in

different signaling pathways, similar to what has been achieved with inhibitors specific for different protein kinases (for example, 41-44). Finally, the affinity selection technology outlined in Chapter 7 is adaptable to non-peptidyl mixtures, providing a rapid way to screen mixtures of small molecules such as combinatorial chemical libraries for potential ligands or inhibitors.

Future Directions

The work described in this thesis contributes a great deal to the understanding of SH2 domain and PTP specificity and suggests a number of experimental paths to pursue. The studies with the SH2 domains of SHP-2 revealed another mode of binding specificity that involves residues N-terminal to the target pY and in particular a role for the pY-2 residue. This unique requirement appears to involve the αA2 Gly residue for the N-terminal SH2 domain of SHP-2. At least four other SH2 domains have this αA2 Gly residue: the C-terminal SH2 domain of SHP-2, the two SH2 domains of SHP-1, and the SH2 domain of Ctk. Similar specificity experiments with these SH2 domains would be of interest to determine if this binding mode is shared by all these SH2 domains and to assess the role of the αA2 Gly residue. There appear to be some specificity differences between the N- and C-terminal SH2 domains of SHP-2 (45), in particular regarding the pY-2 residue, and this needs to be further explored. Crystallographic or NMR studies of the αA2 Gly→Arg mutant complexed with peptide may reveal the structural basis of the selectivity and the effects of the mutation. Finally, using the additional pY-2 specificity constraint in a database search for potential binding targets should greatly restrict potential

high-affinity *in vivo* targets and provide a basis for a directed *in vivo* search for novel binding partners.

The techniques developed for studying PTP specificity have a number of applications as well. As the peptide library affinity selection procedure in Chapter 7 has been validated with PTP1B, the next step is to apply the method to other PTPs to determine their specificities. Specificity differences may help to predict potential *in vivo* targets of the various PTPs as has been done with SH2 domains. Furthermore, important structural aspects of the peptide substrate that contribute to specificity could be exploited in the design of specific PTP inhibitors. The specific high affinity F_2Pmp-peptide inhibitors identified in the affinity selection can themselves be useful tools *in vivo* to inhibit a particular PTP. Such experiments are already in progress in our laboratory, although the techniques for introducing these peptides into the cell need to be optimized.

Studies with more complex peptide libraries or non-peptidyl libraries should also be undertaken. Increasing the complexity of the peptide libraries and adapting the MS analysis to deconvolute the libraries are experiments that have been started in our laboratory. The MS approaches outlined in Chapters 3 and 7 require peptide libraries in which all peptides have unique masses and cannot be directly applied to more complex libraries in which the peptide masses are not unique. Preliminary results using capillary zone electrophoresis (CZE) to separate peptides before MS analysis suggest that the separation will help to identify peptides unambiguously, but more work is required. The ability to analyze more complex libraries will make this technology more readily applicable to all PTPs because more generic unbiased libraries can be used. Beyond studies with

peptide libraries, the affinity selection should in theory be adaptable to any mixture of potential PTP ligands. For example, such an approach would provide a rapid means to identify PTP inhibitors by screening small molecule collections, combinatorial chemical libraries, and even natural products such as fermentation extracts or plant extracts. In this way, lead compounds for drug development may be rapidly identified, greatly simplifying the process of screening compounds.

The *in vitro* approaches described above in the studies of SH2 domains and PTPs are powerful methods for uncovering specificity determinants. However, the specificity must also be demonstrated *in vivo* to be relevant to cell signaling pathways. When studying signaling pathways it is easy to forget the networking and crosstalk since pathways are generally studied as isolated, independent units. The exquisite complexities of signaling networks are daunting and it will be a major task to dissect the intricacies of the connections in the cell, but ultimately such an understanding is necessary. Knowledge of signaling pathway specificity should lead to an understanding of many fundamental aspects of life, from regulation of cell growth and adaptation to disease states and cell death. Hence, investigations into the mechanisms by which signaling specificity is achieved are and will continue to be a very important and fascinating problem to unravel.

REFERENCES

1. Bastien, L., Ramachandran, C., Liu, S., and Adam, M. (1993) *Biochem. Biophys. Res. Comm.* **196,** 124-133

2. Freeman, R. M., Jr., Plutzky, J., & Neel, B. G. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 11239-11243

3. Ahmad, S., Banville, D., Zhao, Z., Fischer, E. H., and Shen, S. H. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 2197-2201

4. Adachi, M., Sekiya, M., Miyachi, T., Matsuno, K., Hinoda, Y., Imai, K., and Yachi, A. (1992) *FEBS Lett.* **314,** 335-339

5. Feng, G-S., Hui, C.-C., and Pawson, T. (1993) *Science* **259,** 1607-1611

6. Vogel, W., Lammers, R., Huang, J., and Ullrich, A. (1993) *Science* **259,** 1611-1614

7. Kazlauskas, A., Feng, G.-S., Pawson, T., and Valius, M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 6939-6942

8. Songyang, Z., and Cantley, L. (1995) *Trends Biochem. Sci.* **20,** 470-475

9. Schaffhausen, B. (1995) *Biochem. Biophys. Acta* **1242,** 61-75

10. 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., and Cantley, L. C. (1993) *Cell* **72,** 767-778

11. Burshtyn, D. N., Yang, W., Yi, T., and Long, E. O. (1997) *J. Biol. Chem.* **272,** 13066-13072

12. Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. D., Silverman, L., and Kuriyan, J. (1992) *Nature* **358,** 646-653

13. Tong, L., Warren, T. C., King, J., Betageri, R., Rose, J., and Jakes, S. (1996) *J. Mol. Biol.* **256,** 601-610

14. Breeze, A. L., Kara, B. V., Barratt, D. G., Anderson, M., Smith, J. C., Luke, R. W., Best, J. R., and Cartlidge, S. A. (1996) *EMBO J.* **15,** 3579-3589

15. Lee, C.-H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E., and Kuriyan, J. (1994) *Structure* **2,** 423-438

16. Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R. A., and Cantley, L. C. (1994) *Mol. Cell. Biol.* **14,** 2777-2785

17. Tracey, A. S., and Gresser, M. J. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 609-613

18. Crans, D. C., Bunch, R. L., and Theisen, L. A. (1989) *J. Am. Chem. Soc.* **111,** 7597-7607

19. Crans, D. C. (1994) *Comments Inorg. Chem.* **16,** 1-33

20. Denu, J. M., Lohse, D. L., Vijayalakshmi, J., Saper, M. A., and Dixon, J. E. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 2493-2498

21. Zhang, M., Zhou, M., Van Etten, R. L., and Stauffacher, C. V. (1997) *Biochemistry* **36,** 15-23

22. Przyborowski, L., Schwarzenbach, G., and Zimmermann, T. (1965) *Helv. Chim. Acta* **48,** 1556-1565

23. Kustin, K., and Toppen, D. L. (1973) *J. Am. Chem. Soc.* **95,** 3564-3563

24. Jaswal, J. S., and Tracey, A. S. (1991) *Inorg. Chem.* **30,** 3718-3722

25. Posner, B. I., Faure, R., Burgess, J. W., Bevan, A. P., Lachance, D., Zhang-Sun, G., Fantus, I. G., Ng, J. B., Hall, D. A., Soo Lum, B., and Shaver, A. (1994) *J. Biol. Chem.* **269,** 4596-4604

26. Monteiro, H. P., and Stern, A. (1996) *Free Radical Biol. Med.* **21,** 323-333

27. Fialkow, L., Chan, C. K., Grinstein, S., and Downey, G. P. (1993) *J. Biol. Chem.* **268,** 17131-17137

28. Sundaresan, M., Yu, Z.-X., Ferrans, V. J., Irani, K., and Finkel, T. (1995) *Science* **270,** 296-299

29. Brumell, J. H., Burkhardt, A. L., Bolen, J. B., and Grinstein, S. (1996) *J. Biol. Chem.* **271,** 1455-1461

30. Bae, Y. S., Kang, S. W., Seo, M. S., Baines, I. C., Tekle, E., Chock, P. B., and Rhee, S. G. (1997) *J. Biol. Chem.* **272,** 217-221

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

32. Burke, T. R., Jr., Smyth, M. S., Otaka, A., Nomizu, M., Roller, P. P., Wolf, G., Case, R., and Shoelson, S. E. (1994) *Biochemistry* **33,** 6490-6494

33. Burke, T. R., Jr., Ye, B., Yan, X., Wang, S., Jia, Z., Chen, L., Zhang, Z.-Y., and Barford, D. (1996) *Biochemistry* **35,** 15989-15996

34. Solas, D., Hale, R. L., and Patel, D. V. (1996) *J. Org. Chem.* **61,** 1537-1539

35. Kelly, M. A., Liang, H., Sytwu, I.-I., Vlattas, I., Lyons, N. L., Bowen, B. R., and Wennogle, L. P. (1996) *Biochemistry* **35,** 11747-11755

36. Jia, Z., Barford, D., Flint, A. J., and Tonks, N. K. (1995) *Science* **268,** 1754-1758

37. Zhang, Z.-Y., Maclean, D., McNamara, D. J., Sawyer, T. K., and Dixon, J. E. (1994) *Biochemistry* **33,** 2285-2290

38. Fawell, S., Seery, J., Daikh, Y., Moore, C., Chen, L. L., Pepinsky, B., and Barsoum, J. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 664-668

39. Kole, H. K., Garant, M. J., and Bernier, M. (1996) *J. Biol. Chem. 271,* 14302-14307

40. Giorgetti-Peraldi, S., Ottinger, E., Wolf, G., Ye, B., Burke, T. R., Jr., and Shoelson, S. E. (1997) *Mol. Cell. Biol.* **17,** 1180-1188

41. Birchall, A. M., Bishop, J., Bradshaw, D., Cline, A., Coffey, J., Elliott, L. H., Gibson, V. M., Greenham, A., Hallam, T. J., Harris, W., Hill, C. H., Hutchings, A., Lamont, A. G., Lawton, G., Lewis, E. J., Maw, A., Nixon, J. S., Pole, D., Wadsworth, J., and Wilkinson, S. E. (1994) *J. Pharmacol. Exp. Ther.* **268,** 922-929

42. Lee, J. C., Laydon, J. T, McDonnell, P. C., Gallagher, T. F., Kumar, S., Green, D., McNulty, D., Blumenthal, M. J., Heys, J. R., Landvatter, S. W., Strickler, J. E., McLaughlin, M. M., Siemens, I. R., Fisher, S. M., Livi, G. P., White, J. R., Adams, J. L., and Young, P. R. (1994) *Nature* **372,** 739-746

43. Alessi, D. R., Cuenda, A., Cohen, P., Dudley, D. T., and Saltiel, A. R. (1995) *J. Biol. Chem.* **270,** 27489-27494

44. Meydan, N., Grunberger, T., Dadi, H., Shahar, M., Arpaia, E., Lapidot, Z., Leeder, J. S., Freedman, M., Cohen, A., Gazit, A., Levitzki, A., and Roifman, C. M. (1996) *Nature* **379,** 645-648

45. Sugimoto, S., Wandless, T. J., Shoelson, S. E., Neel, B. G., and Walsh, C. T. (1994) *J. Biol. Chem.* **269,** 13614-13622