SCREENING COMBINATORIAL PEPTIDE LIBRARY FOR
OPTIMAL ENZYME SUBSTRATES AND HIGH
AFFINITY PROTEIN LIGANDS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By

Peng Wang, B. S.

The Ohio State University

2003

Dissertation Committee:

Professor Dehua Pei, Adviser

Professor Ross E. Dalbey

Professor Patrick Hatcher

Professor Michael A. Freitas

Approved by

______________________________

Advisor

Department of Chemistry
ABSTRACT

My Ph.D study at The Ohio State University has mainly focused on the development of new experimental methods to address the general problems faced in combinatorial peptide library screenings and applying these new methods to gain the insight information concerning the specificity of important enzymes or protein domains such as Protein Tyrosine Phosphatases (PTPs), Peptide Deformylase (PDF), and Forkhead Associated (FHA) protein domain.

A method for the rapid identification of high-affinity ligands was used to study the specificity of the interaction between FHA2 domain of Rad53 and phosphotyrosyl peptide. A phosphotyrosyl (pY) peptide library containing completely randomized residues at positions -2 to +3 relative to the pY was synthesized on TentaGel resin, with a unique peptide sequence on each resin bead. The library was screened against the biotinylated FHA2 domains, and the beads that carry high-affinity ligands of the FHA2 domains were identified using an enzyme-linked assay. Peptide ladder sequencing of the hundreds of selected beads using MALDI-TOF revealed consensus sequences for FHA2 domains. Several representative sequences selected from the library were individually synthesized and tested for binding to the FHA2 domains by Fluorescence Resonance Energy Transfer (FRET), which have demonstrated that the selected peptides are capable
of binding to the FHA2 domains with dissociation constants (K_d) in the micromolar range.

To address the problems faced during the peptide sequencing of support-bound peptide library, a new method was developed to rapidly sequence support-bound peptides derived from combinatorial peptide library by adapting the peptide ladder sequencing strategy of Chait and Kent. In this method, support-bound peptides isolated from one-bead-one-compound libraries are subjected to partial Edman degradation in the presence of an N-terminal blocking agent (~5% phenyl isocyanate). Repetition of the degradation reaction for n + 1 cycles (n = No. of residues to be sequenced) results in a series of sequence-specific truncation products (a peptide ladder). The sequence of the full-length peptide is determined by MALDI-TOF analysis of the peptide ladder. This method is highly sensitive, rapid, and inexpensive.

During screening of combinatorial libraries for optimal enzyme substrates, the challenge is to differentiate a reaction product(s) from a complex mixture of substrates. We have overcome this problem by partially labeling the substrates with a heavier isotope (heavy/normal isotope = 1:1), so that each member of the substrate library appears as a doublet in ESI-MS spectrum. Enzymatic reaction removes the functional group that contains the isotopic label and the products appear as singlets in the spectrum, allowing for their unambiguous identification. The strategy has been successfully demonstrated by peptide deformylase screening.

This method was further perfected in the screening of the pY library against the catalytic domain of SHP-1. A 361-member solution-phase peptide library with randomization at the −1 and −2 positions (relative to pY) was synthesized with the
partially $^{18}$O-labeled pY by the split-synthesis method. Each member of the resulting pY peptide library appeared as a doublet peak in the mass spectrum ($m/z$ m and $m + 2.0043$). Limited treatment of the library with a PTP removed phosphoryl group from the most preferred substrates to generate products as singlet peaks, which were readily identified in ESI-FTICR-MS spectrum and sequenced by tandem mass spectrometry. A semi-quantitative measurement of product formation as a function of time allows us to rank the order of the identified optimal substrates in terms of their $k_{cat}/K_m$. Several selected peptides were individually synthesized and assayed against SHP-1 and the kinetic data confirmed the screening results.
Dedicated To My Parents
ACKNOWLEDGEMENT

Looking back at the years I spent at The Ohio state University as a graduate student, I consider myself extremely lucky to have so many intelligent and hard-working people as my teachers and colleagues. I benefit tremendously from this mentally stimulating environment. I am happy to have come here, gratified about what I am now, optimistic about what I am going to be. I ought to express my gratitude to all the people from whom I have learned so much. Let me first thank my advisor for giving me best advices. Not only did he put his great effort and long hours into the work I have been doing, but also tried his best to provide me with best working environment. I also would like to express my deep gratitude to Dr. Michael Freitas for his generous support and mentoring. The majority of my work would not be otherwise possible without his kind help. I am also very grateful to Dr. Ming-Tao Tsai and his former group member Dr. In-Ja L. Byeon and Dr. Hua Liao whom I had great collaboration with on the FHA project. I would also like to acknowledge the people whom I have worked with on various projects: Dr. Kirk Beebe, Dr. Gulnur Arabaci, Dr. Hua Fu and Donald F. Snavley. I sincerely appreciate their help. Finally, I must thank my wife for her continuous support and encouragement.
VITA

1997 ……… B. S Wuhan University

1998-present ……… Graduate Teaching/Research Associate, The Ohio State University

PUBLICATION


FIELD OF STUDY:

CHEMISTRY
TABLE OF CONTENTS

| Abstract | ii |
| Dedication | v |
| Acknowledgments | vi |
| Vita | vii |
| List of Tables | xii |
| List of Figures | xiii |

Chapters:

1. General Introduction.................................................................................1

1.1 Support-Bound Library Screening..........................................................2

1.2 96-Well Solution Phase Library Screening...............................................6

1.3 Analysis Of The Support-Bound Peptide Library....................................6

1.4 Analysis Of Combinatorial Library By Mass Spectrometry.......................8

1.5 Screening Of Combinatorial Library Using Mass Spectrometry...............9

1.5.1 Affinity LC–MS..................................................................................9

1.5.2 Size Exclusion Chromatography (SEC)–Mass Spectrometry.................10

1.5.3 Affinity Capillary Electrophoresis–Mass Spectrometry......................11

1.5.4 Frontal Affinity Chromatography–Mass Spectrometry......................11

1.5.5 Electrospray FT–ICR Mass Spectrometry........................................12
3.1. Introduction……………………………………………………………………...38

3.2. Experimental Procedures……………………………………………………40

3.2.1. Materials…………………………………………………………………40

3.2.2. Peptide Library Synthesis………………………………………………..40

3.2.3. On-Bead Screening of Peptide Library…………………………………..41

3.2.4. Partial Edman Degradation………………………………………………43

3.2.5. Peptide Sequencing by MALDI Mass Spectrometry…………………….44

3.3. Results…………………………………………………………………………...44

3.4. Discussions……………………………………………………………………...48

4. Screening Combinatorial Libraries for Optimal Enzyme Substrates of Peptide Deformylase by Mass Spectrometry……………………………………………...55

4.1. Introduction……………………………………………………………………...55

4.2. Experimental Procedures………………………………………………………..57

4.2.1. Materials…………………………………………………………………57

4.2.2. Synthesis of Peptide Library……………………………………………..57

4.2.3. Library Screening and MS sequencing…………………………………..58

4.3. Results…………………………………………………………………………...60

4.4. Discussions……………………………………………………………………...63

5. Screening Combinatorial Libraries for Optimal Enzyme Substrates of Protein Tyrosine Phosphatase SHP-1 by Mass Spectrometry Screening…………………….71

5.1. Introduction……………………………………………………………………..71
5.2. Experimental Procedure.............................................................................. 74
  5.2.1. Materials................................................................................................. 74
  5.2.2. Synthesis of t-Butyl N\textsuperscript{\alpha} -[(Fluoren-9-yl)methoxycarbonyl]-L-tyrosinate. 74
  5.2.3. Synthesis of \textsuperscript{18}O-Labeled N\textsuperscript{\alpha} -[(Fluoren-9-yl)methoxycarbonyl]-O-
  [bis(allyloxy) phosphoryl]-L-tyrosine......................................................... 74
  5.2.4. Peptide Synthesis...................................................................................... 74
  5.2.5. Library Screening.................................................................................... 75
  5.2.6. Mass Spectrometry.................................................................................. 76
  5.2.7. On-line HPLC-MS Analysis Of Starting Peptide Library...................... 77
  5.2.8. PTP Assay............................................................................................... 78
5.3. Results........................................................................................................ 78
  5.3.1. Library Design and Synthesis................................................................. 78
  5.3.2. MS Analysis of Untreated Library.......................................................... 79
  5.3.3. Identification of Optimal Substrates of SHP-1......................................... 80
  5.3.4. Identification of Poor Substrates of SHP-1.............................................. 82
  5.3.5. Rank Order of Selected SHP-1 Substrates............................................. 82
  5.3.6. Kinetic Analysis of Selected pY Peptides................................................ 84
5.4. Discussion................................................................................................... 85
APPENDIX A..................................................................................................... 101
APPENDIX B..................................................................................................... 103
APPENDIX C..................................................................................................... 107
APPENDIX D..................................................................................................... 122
REFERENCE..................................................................................................... 130
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. Sequences of selected peptides from the library</td>
<td>36</td>
</tr>
<tr>
<td>2.2. Dissociation constants of pTyr peptides to FHA2</td>
<td>37</td>
</tr>
<tr>
<td>3.1. Selected Sequences for <em>E. coli</em> PDF*</td>
<td>54</td>
</tr>
<tr>
<td>4.1. The Most Abundant Products from PDF Treated Library</td>
<td>70</td>
</tr>
<tr>
<td>5.1. Most Abundant Products from the SHP-1-Treated Library</td>
<td>97, 98</td>
</tr>
<tr>
<td>5.2. Poor Substrates of SHP-1</td>
<td>99</td>
</tr>
<tr>
<td>5.3. Kinetic Properties of pY Peptides</td>
<td>100</td>
</tr>
<tr>
<td>C.1. The inhibition test of the peptide inhibitors chosen from the library</td>
<td>121</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Split-Pool Synthesis</td>
</tr>
<tr>
<td>2.1</td>
<td>The design of the pY peptide library for the screening of FHA2</td>
</tr>
<tr>
<td>2.2</td>
<td>Enzyme-linked assay on bead for screening the binding of FHA2 domain</td>
</tr>
<tr>
<td>2.3</td>
<td>Ladder peptide sequencing of one of the selected peptides</td>
</tr>
<tr>
<td>2.4</td>
<td>Picture of the three positive beads (blue colored)</td>
</tr>
<tr>
<td>2.5</td>
<td>pTyr peptide sequence specificity for binding to the FHA2 domain</td>
</tr>
<tr>
<td>2.6</td>
<td>Binding assay of the peptide DNS-GF(pY)-LYFIR</td>
</tr>
<tr>
<td>2.7</td>
<td>The binding interface for FHA2 and the Rad9 pTyr peptide</td>
</tr>
<tr>
<td>3.1</td>
<td>The enzyme-linked assay used for the screening of peptide deformylase</td>
</tr>
<tr>
<td>3.2</td>
<td>(a) Chemistry of Edman degradation (b) Partial Edman degradation on bead-bound peptide</td>
</tr>
<tr>
<td>3.3</td>
<td>Sequencing of support-bound peptides by partial Edman degradation and mass spectrometry</td>
</tr>
<tr>
<td>3.4</td>
<td>Substrate specificity of <em>E. coli</em> Co(II) PDF</td>
</tr>
<tr>
<td>4.1</td>
<td>Scheme showing library screening by mass spectrometry</td>
</tr>
</tbody>
</table>
4.2. ESI FTICR mass spectra (positive ion mode) of the 361-member peptide library……………………………………………………………………………...68

4.3. (a) Expanded spectrum of the box regions of Figure 4.2,(b) A Q-tof MS/MS spectrum of the product peak b (m/z 435.2)……………………………………………………………...69

5.1. Scheme showing library screening by ECLIPSE………………………………………...91

5.2. The design of the pY peptide library for the identification of optimal enzyme substrates of protein tyrosine phosphatases by mass spectrometry-based screening…………………………………………………………………………………...92

5.3. ESI-FTICR mass spectra (positive ion mode) of the 361-member pY peptide library before enzymatic treatment (A) and after treatment with SHP-1 (8.0 µg) for 3.5 (B), 10 (C), and 40 min (D)…………………………………………………………...93

5.4. (A) Expanded spectra for the m/z 878–881 region in Figure 5.3, showing the appearance of product peaks with increasing SHP-1 reaction time. a, peptide RNNDZYA-NH₂; b, peptide RNNDQ(EN)YA-NH₂; c, peptide RNNDEYA-NH₂. Peak a+1 is due to the existence of natural isotopes in a (e.g., ¹³C). (B) A Q-TOF MS/MS spectrum of the product peak a (m/z 878.4). (C) Plot of peak intensities for products a–c in part A as a function of time (reaction progress curves). Z, Leu, Ile, or Nle………………………………………………………………………………………………………………...94

5.5. Reaction progress curves for the 23 most abundant SHP-1 products…………………...95

5.6. On-line HPLC-MS analysis of Starting peptide Library……………………………………...96

B.1. Synthesis of partially ¹⁸O-labeled pY building block……………………………………106

C.1. Structure of the three compounds 1, 2, 3……………………………………………………118

C.2. ESI-MS spectrum of the elution solution…………………………………………………...119
C.3. MALDI-MS spectra of the eluting solution (b) and control 1 (a)………….120

D.1. (a) Structure of the inhibitor, (b) Core structure of the designed
    compounds.................................................................127, 128

D.2 The inhibitor library screening by ESI-MS........................................129
Combinatorial chemistry, contrary to traditional organic chemistry which produces single compound at one time, offers a power means to simultaneously generate a large number of molecules known as combinatorial chemical library. The broad molecular diversity represented in the combinatorial library is often exploited and examined to search for high affinity protein ligands, potent enzyme inhibitors, optimal enzyme substrates and even the catalysts of certain chemical processes [1].

Split-Pool synthesis (Figure 1.1) has been widely used by biochemists to generate one-bead one-compound combinatorial peptide library due to its simplicity [2-4]. The solid support material (resin) is physically separated into equal portions for coupling to each of the individual amino acid by standard peptide chemistry. Uniform coupling of amino acid to the resin takes place since competition between amino acid is eliminated. The resins are combined in a single vessel for washing and deprotection and then divided equally again into individual portions for the next amino acid coupling. The process can be repeated for additional random position coupling. The resulting synthetic peptides cover all possible peptide sequences. Since only one peptide sequence is presented on each
solid support (bead) as a result of Split-Pool synthesis, the bead-bound peptide library was often directly used for screening by identifying the positive beads that carry the bioactive peptides. Alternatively, peptides can be cleaved from their resin support and the peptides assayed in aqueous solution for their biological activity. The active peptides remain unknown before their identities are revealed by proper characterization.

However, it is no easy task to screen sometimes hundreds and thousands of compounds in the combinatorial library to find a handful of lead compounds with desired properties. The key to the success of combinatorial library screening is throughput. The assay method ought to be able to handle hundreds and thousands of compounds in a reasonable time frame and the characterization method has to be fast and compatible to the screening format. Over the past decade, many ingenious experimental approaches have been invented to address these issues so that an ordinary lab can perform successful combinatorial library screenings. Nevertheless, the advancement of the analytical techniques especially in the area of mass spectrometry has provided new opportunities for better and faster combinatorial library screening [5].

1.1 Support-Bound Library Screening

After Split-Pool synthesis, synthetic peptides with various sequences are presented on different solid polymer beads and only one peptide sequence is presented on one bead (one-bead one compound). The peptides are spatially separated from each other so that the positive bead can be manually isolated and the peptide on the bead can be analyzed to determine its sequence. The peptides resulted from split-pool synthesis have equal molar concentration. Under these preset screening conditions, the peptides in the library are
allowed to compete with each other for the binding or the enzymatic catalysis. Many binding assays and enzyme assays were developed to detect the binding events between the protein target and the synthetic peptides or the enzymatic reactions of certain peptide substrates. The peptides with strong binding capacity to the protein target or the optimal enzyme substrates can be identified and isolated for analysis. For example, the peptides on the selected beads can be released from the solid support and sequenced by Edman degradation. Note that the binding events or the enzymatic reaction happens on the peptides bound to the solid support that may interfere with the binding or enzymatic processes. Therefore, rigorous control ought to be done to eliminate false positives produced by the support-bound library screening.

For support-bound library screening, the polymer bead to which the library is attached should be water compatible since the biological assays are conducted in aqueous solution. The polyethyleneglycol grafted polystyrene bead (TentaGel) was initially introduced by Papp et al [6-8] for this purpose. The TentaGel beads are uniform in size to insure the uniformity of the load capacity of the beads.

1.1.1 Binding Assays

The binding of the protein target to the support-bound synthetic ligands (e.g. peptide) can be detected either by direct visualization if the protein target is intrinsically chromogenic or fluorescent or indirectly by a reporter group such as an enzyme, a fluorescent group or radioactive group that is attached to the target. Chen et al. [9] screened a bead-bound peptide library against fluorescently labeled SH3 domain and isolated the positive beads with a fluorescent microscope. Kassarjian et al. [10] used a


[125I]-labeled anti-β-endorphin to screen a peptide library and used autoradiography to locate the positive beads. Using an enzyme-linked colorimetric assay, Lam et al. [2] screened a random pentapeptide library with anti-β-endorphin monoclonal antibody and identified six peptides that bind specifically to this monoclonal antibody (mAb). Smith et al. [11] were the first to describe the use of the bead-bound peptide library approach to determine the anchor residues for MHC-class I molecules. Stricker et al. [12] screened 13 billion distinct peptides against PDZ domain of neuronal NO synthase (nNOS), and found that the PDZ domain binds tightly to peptides ending with sequence Asp-X-Val. Guided by the Asp-X-Val consensus, candidate nNOS interacting proteins have been identified including glutamate and melatonin receptors. Muller et al. [13] used the binding of fluorescein isothiocyanate-labeled antibodies directed against GST to monitor the phosphopeptide interaction of a given SH2 domain fused with GST. High-fluorescence beads were isolated by flow cytometric sorting for sequencing. Dooley et al. [14] identified several selective ligands for the mu, delta, and Kappa opioid receptors from a single mixture based tetrapeptide positional scanning combinatorial library. Beebe et al. [15] have determined the binding specificity of the SH2 domains of protein tyrosine phosphatase SHP-1 through the screening of a combinatorial phosphotyrosyl peptide library.

1.1.2 Enzymatic Assays

Many unique assays have been devised to detect the enzymatic reactions of the optimal substrates. Lam and Wu [16] reported the use of a support-bound peptide library method to identify peptide substrate for posttranslational modifications such as protein
phosphorylation. In this assay, a random peptide library is incubated with $[^\gamma-^{32}\text{P}]$ATP and protein kinase. Autoradiography is then used for localization of the positive beads. Meldal et al. [17] developed a FRET assay for the identification of a proteolytic substrate from a random peptide library. In this assay, a fluorescent molecule was attached to the carboxyl terminus of a random peptide and a fluorescent quenching molecule is attached to the amino terminus. After incubating with the protease, peptides with the proper sequences will be cleaved, and the quencher on the amino terminus of the peptide will be released resulting in an increase in fluorescence of the positive bead. This assay strategy has been applied in the combinatorial peptide library screening of leader peptidase [18], and Capthepsin [19].

A similar fluorescent assay strategy was developed by Harris et al. [20] for rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. Hu et al. [21] has used similar strategy to screen immobilized peptide library for determination of the substrate specificity of Clostridium histolyticum $\beta$-collagenase. Hu et al. [22] have used N-terminally formylated tetrapeptides library to screen against peptide deformylase by using a solid phase enzyme-linked assay.

A related combinatorial library screening method used for systemic identification of optimal enzyme substrate is phage displays. Up to tens of billions of peptides and proteins, including human antibodies and enzymes are biologically displayed on the surface of a small bacterial virus called a phage. By searching through these large peptide libraries, peptide of interest can be selected and identified. This method has been successfully used for studying the enzyme specificity of subtilsin [23], furin [24], matrix metalloproteinases [25], kinases [26] and prostate-specific antigen [27].
1.2 96-Well Solution Phase Library Screening

To screen a one-bead-one-compound library in solution phase, the 96-well two-stage release assays were developed. In these assay, double orthogonally cleavable linkers are incorporated during the preparation of the library [28-31]. Approximately 100-500 beads are added into each well of a 96-well filtration plate. With the cleavage of the first linker, the mixtures of ligands were released and transferred into a 96-well plate placed beneath the filtration plate. The filtrates are then tested for biological activity. Beads from the positive wells are then redistributed into filtration plates with one bead per well. With the second linker cleaved by proper treatment, the filtrates from each well are then tested for biological activity. With this deconvolution strategy, the beads that correspond to the positive wells are then isolated for structure characterization. Alternatively, an on-bead binding assay can first be performed. The positive beads can then be collected for a releasable functional solution-phase assay to identify the true positive beads.

1.3 Analysis Of The Support-Bound Peptide Library

Edman degradation on automatic peptide sequencers can be used to sequence the support-bound peptides with minimal modification of the standard sequencing protocol [2, 18, 32-34]. However, the low throughput of this approach greatly impedes the application of this method. New methods for compound characterization had to be developed if non-α-amino acids and novel building blocks were required in the library building, since Edman sequencing was no longer feasible.

Brenner and Lerner [35] first suggested the use of polynucleotides as tags to encode
peptide libraries. The structure of the compound can be inferred by examining the tag on the same bead that carries the compound. Each individual library member is encoded with a unique nucleotide sequence which can be easily measured. This method has gained wide applications in many cases [36-38].

Additional coding strategies were also developed by using various chemical tags such as peptide tag [39, 40], haloaromatic tag [41], secondary amine tag [42] and fluorescent colloid [43] etc. An ingenious method of incorporating a radio frequency transmitting chip in the library support to record each coupling reaction was developed to encode a bead-bound peptide library [44, 45].

To prevent the undesirable interaction between the encoding structure and the biological target, novel techniques were developed, in which the encoding structure was attached in the internal volume of the bead and the screening structure on the surface [46, 47].

To minimize the interference from the chemical tags, the amount of tags presented on the bead needs to be kept at minimal level. Therefore, these tags are generally amenable to sensitive detection method such as gas chromatography, fluorescence and mass spectrometry [41-43].

To eliminate possible interference from the encoding structure, “isotope ratio encoding” has been developed for the mass spectrometric identification of bioactive leads using stable isotopes incorporated into the library compounds [48, 49]. The mass spectrum of the compound shows a molecular ion with a unique isotope ratio that encodes for a particular library compound.

Youngquist et al. [50] developed an ingenious and popular approach to the sequencing
of the library-derived peptide. Partial (~10%) capping of a peptide chain was conducted in each step of the random residue synthesis. As a result, every bead contains a series of capped peptide and the peptide sequence on the bead can be easily interpreted from mass differences between these capped peptides. This rapid and inexpensive method has been successfully applied to both peptide and peptidomimetic libraries [15, 22, 51].

1.4 Analysis Of Combinatorial Library By Mass Spectrometry

The advancement of mass spectrometry has made itself ideally suited for the mixture analysis of large numbers of compounds prepared using combinatorial library due to its high sensitivity, high selectivity and high throughput. Mass spectrometry can be used to obtain very selective information regarding molecular weight of each of the compounds in the mixture, elemental composition (derived from isotopic distribution) and structural features based on fragmentation patterns in tandem mass spectrometry. Furthermore, the analysis can be conducted using minimal materials as little as femtomoles and completed at a high rate. When coupled to a separation technique such as high performance liquid chromatography (HPLC), mass spectrometry may be applied to the analysis of more complex mixtures.

Molecular ions are routinely detected using virtually any type of mass analyzer equipped with suitable ionization techniques such as electrospray and MALDI (matrix-assisted laser desorption ionization) ionization. However, high resolving power in combination with high mass accuracy is especially critical for characterization of complex combinatorial library mixtures. Fourier transform ion cyclotron resonance (FTICR) mass spectrometry has the highest performance in terms of resolving power and
mass accuracy. FTICR instrument is capable of resolving nominally isobaric components for identification. FTICR mass spectrometry has been routinely used for the characterization of complex peptide library up to hundreds of components [52]. In addition, FTICR mass spectrometers can also be used to analyze other types of combinatorial libraries such as pyrazoles [53, 54] and pseudo-peptide macrocyclic hydrazones [55].

The compound libraries have been purified on HPLC-mass spectrometry system with automated mass-directed fractionation [56-58]. Only those chromatographic peaks displaying the desired molecular ions were collected by automatic control. Recently, supercritical fluid chromatography–mass spectrometry (SFC–MS) has been used for the high-throughput analysis of combinatorial libraries [59, 60], which has faster chromatographic step compared to normal HPLC.

1.5 Screening Of Combinatorial Library Using Mass Spectrometry

During the last few years, many mass spectrometry-based screening approaches have been developed toward faster and more reliable screening. Different from solid phase screening, these methods are performed by screening combinatorial library mixtures dissolved in aqueous solution. Based on the competitive binding of the ligands prepared by combinatorial synthesis, the lead compounds are identified by mass spectrometry techniques.

1.5.1 Affinity LC–MS

In this method, a binding protein or enzyme is immobilized on a solid support in a
chromatography column through which the library mixture is then pumped in a suitable binding buffer so that the ligands with the highest affinity for the protein would be able to bind. After the unbound compounds in the mixture were washed away, the bound ligands are eluted and identified using electrospray mass spectrometry. Kelly et al. [61] has used an affinity column containing immobilized phosphatidylinositol-3-kinase for the screening of a 361-member peptide library. Electrospray mass spectrometry and tandem mass spectrometry were used to identify the ligands released from the affinity column. Similar strategy has also been reported independently by two laboratories to study the substrate specificity of protein tyrosine phosphatase PTP1B [62, 63]. Nevertheless, immobilization of the receptor might change its affinity properties causing false negative or false positive hits. Furthermore, the stationary phase of the column might selectively bind some compounds in the combinatorial mixture, resulting in false positive hits.

1.5.2 Size Exclusion Chromatography (SEC)–Mass Spectrometry

Size Exclusion chromatography (SEC) has been combined with mass spectrometry as a screening system. During the screening, the large protein–ligand complexes are separated from the unbound compounds in the mixture by SEC. Since the binding ligands begin to dissociate from the protein very quickly once the unbound compounds are separated from the complex, the separation step must be carried out quickly. This approach allows the screening to be conducted in solution, which avoids the problems associated with the screening by affinity columns. Subsequently, the protein–ligand complex is denatured to release the ligands, which are identified using electrospray mass spectrometry. Dunayevskiy et al first demonstrated this strategy by screening five known inhibitors of
serum albumin [64]. The method was further perfected by Siegel et al. by using human cytomegalovirus protease and several drug candidates as a model system [65]. It was Kaur et al. who really demonstrated the usefulness of this method by screening a real combinatorial library with about a hundred members [66].

1.5.3 Affinity Capillary Electrophoresis–Mass Spectrometry

Chu et al. [67] developed an approach in which the affinity capillary electrophoresis–mass spectrometry was used for the screening of a 100-member peptide against vancomycin. Without vancomycin present in the electrophoresis buffer, all peptides eluted quickly (within 3 min). With vancomycin present in the electrophoresis buffer, the highest affinity compounds were detected between 4.5 and 5 min. The peptides eluted in the order of affinity to vancomycin. Electrospray mass spectrometry was coupled to affinity capillary electrophoresis to identify the highest affinity ligands.

1.5.4 Frontal Affinity Chromatography–Mass Spectrometry

In frontal affinity chromatography, the compounds with the highest affinity for the immobilized protein elute last, and compounds with no affinity elute in the void volume. Recently, frontal affinity chromatography was coupled to mass spectrometry and used for the screening of combinatorial libraries [68-70]. In these methods, the elution of all compounds in the library are detected and identified by mass spectrometry. The last compounds eluted at their infusion concentrations are the highest affinity compounds. However, the same disadvantages of using immobilized receptors still exist.
1.5.5 Electrospray FTICR Mass Spectrometry

A combinatorial library is pre-incubated with a protein in solution and then analyzed directly using electrospray in order to identify receptor–ligand complexes in the gas phase [71-75]. The mass difference between the protein-ligand complex and the protein alone can be measured with sufficient accuracy to determine the identities of the ligands. Gao et al have used the FTICR mass spectrometric to screen a 289-member peptide library for ligands to carbonic anhydrase [72]. The enzyme-ligand complexes were detected by electrospray ionization and were subsequently isolated and trapped in the FTICR mass spectrometer. Tandem mass spectrometry experiments were used to dissociate and identify the peptide ligands. RNA receptors were incubated with an aminoglycoside library to identify potential ligands for the RNA binding ligands [74, 75]. The advantage of this screening method over other methods is the direct detection of the receptor-ligand complex. However, the receptor–ligand complex might dissociate during electrospray process under certain harsh solvent and ion source conditions or it might have relatively low ionization efficiency. Furthermore, the buffer solution has to contain low level of salt.

1.5.6 Pulsed Ultrafiltration–Mass Spectrometry

Protein-ligand complexes are confined in the ultrafiltration chamber while unbound library components and buffer were washed away during pulsed ultrafiltration. After unbound compounds are removed, the ligand–receptor complex was denatured to release the ligands from the complex. The released ligands were then directed from chamber to on-line electrospray mass spectrometry [76] or collected and analyzed by off-line mass
spectrometry [77]. Library screening using this strategy have been reported for a variety of proteins including dihydrofolate reductase [77], adenosine deaminase [78], cyclooxygenase-2 [79], serum albumin [80, 81] and estrogen receptors [82].

1.6 Challenges In Combinatorial Library Screening

We can summarize the common practice of combinatorial library screening into three steps. In the first step, the mixture of compounds is incubated with the proteins or enzymes to compete for the protein binding or enzyme catalysis. As a result, the protein binds to the high affinity ligands while the optimal enzyme substrates undergo enzymatic reactions (deformylation, dephosphorylation etc.). In the following second step, the compounds of interest are detected and separated from the rest of the compounds in the library. In support-bound screening strategy, the beads that carry the peptide of interest are detected by certain assays and separated manually from the rest of the beads in the library. In the third step, the compounds of interest are analyzed by proper method to reveal their identities. Since a large number of compounds are dealt with in each step, the throughput is always a general challenge.

Specifically, since hundreds to thousands library members are mixed together, it presents enormous technical challenges to separate particular ones from the rest of them physically and to characterize them in high throughput. For example, the screening of pY peptide library for optimal enzyme substrates of protein tyrosine phosphatases is awfully difficult. Since the only difference between the pY peptide and its possible dephosphorylated enzyme product is a phosphate, there are really no good approaches to separate the enzyme products from the other pY peptides in the starting peptide library.
On the other hand, although a few of methods have been developed for the sequencing of library derived peptides, they all have drawbacks in certain aspects. For instance, peptide ladder sequencing by MALDI-MS [50] was very popular due to its high throughput in peptide sequencing. However, it may introduce bias to some extent during the screening process (see Chapter 3). We have made some improvements in these areas to address the challenges. At the same time, we applied existing screening technique to new system such as FHA2 domain to reveal the specificity of the protein domain.
Figure 1.1 Split-Pool Synthesis.
CHAPTER 2

SCREENING SUPPORT-BOUND PEPTIDE LIBRARY FOR HIGH AFFINITY LIGAND TO FHA2 DOMAIN OF RAD53

2.1 Introduction

Cellular signaling processes involve protein-protein interactions, which are often mediated by small modular domains by recognizing short stretches of amino acid residues in their partner proteins. A list of such modules, including SH2, SH3, PTB, and 14-3-3 domains, have been identified in signaling proteins of organisms from yeast to human [83]. Recently, a new module, the forkhead-associated (FHA) domain, has been identified by sequence analyses [84]. It was first identified within a subset of forkhead transcriptional factors, located outside of the conserved DNA-binding forkhead domain. It was subsequently found in more than 20 other proteins, mostly nuclear protein kinases and transcriptional factors. New FHA containing proteins of diverse functions continue to be identified. For instance, an FHA domain was recently identified in the N terminus of nuclear RNA-binding protein NIPP1 [85]. It has been suggested that the FHA domain is involved in the regulation of pre-mRNA splicing.

Rad53, a yeast checkpoint protein involved in the DNA damage response [85, 87],
contains two FHA domains and binds to phosphorylated Rad9 [88, 89]. The NMR structure of the FHA2 domain of Rad53 (residues 573-730) indicates that the FHA2 domain of Rad53 requires approximately 160 amino acid residues to fold into a stable three-dimensional (3D) structure [90], which is considerably larger than previously predicted on the basis of sequence homology (55-75 amino acid residues). Sun et al. [88] also reported that deletion of 52 amino acid residues from either end of a functional fragment of Rad53 (residues 549-730) abolished its interaction with Rad9.

FHA has been proposed to be a new phosphoprotein binding domain, which is supported by extensive experimental evidence [84]. However, it remains to be established whether the FHA domain recognizes pSer/pThr or pTyr peptides. The FHA domain of kinase-associated protein phosphatase (KAPP) from Arabidopsis thaliana binds to a phosphorylated receptor-like protein kinase RLK5 [91]. Likewise, the FHA domain of NIPP1 interacts with CDC5L in a phosphorylation-dependent manner [85]. In most of these cases, the reported evidence seems to suggest FHA as a pSer/pThr binding domain. For instance, it has been shown that the N-terminal FHA domain of Rad53 (FHA1) binds to a pThr peptide derived from p53 (Kd ~2 μM), and this pThr peptide is capable of disrupting the association between FHA1 and phosphorylated Rad9 [89]. However, a pTyr peptide derived from Rad9, 826EDI(pY)YLD832, binds to the FHA2 domain of Rad53 with a moderate affinity (Kd ~ 100 mM) [90]. These results taken together made us to believe that FHA domains may have dual specificity, or different FHA domains may recognize different phosphoamino acids. In this report, by screening a combinatorial pTyr peptide library, we have identified specific pTyr peptides which bind to FHA2 of Rad53 with low μM Kd values.
2.2 Experimental Procedures

2.2.1 Materials

TentaGel S NH2 resin, Wang resin, N-9-fluorenylmethoxycarbonyl (Fmoc)-amino acids, 1-hydroxybenzotriazole (HOBT) and 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were purchased from Advanced ChemTech (Louisville, KY). N-Hydroxysuccinimido-biotin, acetylglycine, N-acetyl-D,L alanine, 5-bromo-4-chloro-3-indolyl phosphate, streptavidin-alkaline phosphatase and dansyl chloride were obtained from the Sigma Chemical Co. (St. Louis, MO). All other chemicals were purchased from Aldrich Chemical (Milwaukee, WI). The FHA2 domain of Rad53 (residues 573-730) were cloned into pGEX-4T vector (Pharmacia Biotech) for expression of glutathione S-transferase (GST) fusion proteins in BL21 (DE3) (Novagen). The fusion proteins (GST-FHA2) were purified using glutathione agarose (Sigma). The GST tag was removed by thrombin (Sigma) digestion and gel-filtration chromatography.

2.2.2 NMR Experiment

See Appendix A

2.2.3 Synthesis Of The Phosphotyrosyl Peptide Library

TentaGel S NH2 resin (0.3 mmol/g loading, 2.86 × 10^6 beads/g, 80-100 µm) was used as the solid support for the peptide library. Synthesis was carried out on a 5.0 g scale on a home-made peptide synthesis apparatus using standard Fmoc/HBTU/HOBT chemistry [92]. A common linker of four amino acid residues. BBRM (B=β-alanine), was first...
synthesized on the resin by using a fourfold excess of amino acids [93]. The five random positions were generated by using the Split-Pool method, resulting in a one bead-one-peptide library [2-4] (Figure 2.1). The coupling reaction at each random position was carried out twice with a fivefold excess of the proper amino acid for three hours to ensure complete coupling. To produce a small percentage of truncated peptides for later sequence analysis, 10% (mol/mol) acetylglycine was added to the coupling reactions during the synthesis of the random residues and phosphotyrosine (for all amino acids except for norleucine, glutamine, and isoleucine) [22, 50]. For norleucine and glutamine, 10% acetylalanine was used, whereas a 1:1 mixture of acetylglycine and acetylalanine (total 10%) was used during the coupling of isoleucine. After library synthesis was complete, deprotection was carried out using a cocktail containing 90% (v/v) TFA, 2% anisole, 3% ethanedithiol, and 5% thioanisole for one hour at room temperature. The resin was washed with CH$_2$O$_2$ (10 × 10 mL) and methanol (5 × 10 mL), dried under vacuum, and stored at -20 °C.

2.2.4 Synthesis Of Dansylated Peptides

Peptides were synthesized on Wang resin using standard Fmoc/HBTU/HOBT chemistry on a 0.15 mmol scale. Coupling reactions were carried out with a fourfold excess of amino acids for at least one hour and the completion of the reaction was ensured by ninhydrin tests. Addition of a dansyl group to the N terminal amino group was effected by incubating the protected, resin-bound peptide with 1.5 equivalents of dansyl chloride and three equivalents of triethylamine for 45 minutes at room temperature. Deprotection of the side-chains and cleavage of the peptide off the resin were carried out
using the same cocktail as described above. The resulting peptide solution was drained into a glass vial and the solvent was evaporated under a flow of nitrogen gas. The semi-solid residue was triturated with anhydrous diethyl ether (3 × 10 mL) and stored at -20 °C. The crude product was purified by reversed-phase HPLC before use. The identity of all dansylated peptides was confirmed by matrix-assisted laser desorption ionization mass spectrometry (MALDI MS) analysis.

2.2.5 Preparation Of Biotinylated GST-FHA2

Biotinylated GST-FHA2 was obtained by treating GST-FHA2 (114 µM) with 2.5 molar equivalents of N-hydroxysuccinimidobiotin (NHS-biotin) in PBS buffer (50 mM Na/PO4, 350 mM NaCl (pH 7.4)) for three hours before the reaction was quenched by the addition of 4 µL of 2-aminoethanol (17 M). The biotinylated protein was quickly frozen and stored at -80 °C until use.

2.2.6 On-Bead Screening Of The pTyr Peptide Library Against GST-FHA2

Enzyme-linked on bead assay (Figure 2.2) was used for the library screening, which was conducted in three separate batches. For each batch, 300 mg of resin was incubated with DMF (2 × 10 mL) for 30 minutes and exhaustively washed with water (20 × 10 mL) in a plastic column fitted with a filter/disc (Bio-Rad) before use. After incubation of the resin in 3.5 mL of TBS buffer (25 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1 mg/mL gelatin) for one hour, 70 µL of biotinylated GST-FHA2 (final concentration=0.9 µM) was added to the solution. After overnight incubation at room temperature, the solution was drained, and 2.5 mL of TBS buffer, 70 µl of 1.5 M potassium phosphate (pH 7.4),
and 10 µL of streptavidin-alkaline phosphatase (final concentration =22 nM) were added to the resin. After incubation at room temperature for 15 minutes, the solution was again drained and the resin was washed with PBS buffer (10 mL) and then TBS buffer (10 mL). The resin was resuspended in 12 mL of TBS buffer (25 mM Tris (pH 8.5), 150 mM NaCl) in a Petri dish, and 6 mg of 5-bromo-4-chloro-3-indolyl phosphate was added to the mixture. The Petri dish was placed on a shaker at room temperature. Positive beads stained with intense turquoise color emerged from the dish within one hour. The staining reaction was stopped by treating the resin with 4 mL of 6 M guanidine-HCl for 20 minutes. The resin was washed with double-distilled water (10 × 3 mL) and the positive beads were removed from the dish and placed in individual Eppendorf tubes using a pipette under a low-power microscope. A control screening was conducted under the same conditions except that no GST-FHA2 protein was included in the screening. No colored beads were observed.

2.2.7 Peptide Sequencing by MALDI Mass Spectrometry

Each positive bead selected from the library was treated with 20 µL of a cyanogen bromide (CNBr) solution (20 mg/mL in 70% formic acid) in a microcentrifuge tube in the dark for 20-24 hours at room temperature. After evaporation of the solvent under vacuum, the residue was dissolved in 5 µL of 0.1% TFA solution. For MALDI MS analysis, 1 µL of the peptide solution was mixed with 2 µL of a saturated solution of α-cyano-4-hydroxycinnamic acid in 50% aqueous acetonitrile. A 1 µL volume of the mixture was spotted onto the MALDI sample slide. After crystallization of the mixture, the sample was analyzed on a Kratos Kompact MALDI-III mass spectrometer.
2.2.8 Determination Of Binding Constants

Binding of the dansylated peptides to GST-FHA2 or free FHA2 domain was monitored by fluorescence energy transfer from tryptophan residues in the protein to the dansyl group in the peptides on a Perkin-Elmer LS-5 spectrophotometer. The tryptophan residues were excited at 290 nm while the emission of the dansyl group at 520 nm was monitored. All binding assays were carried out in a phosphate buffer (4.3 mM Na/PO4, 1.5 mM K/PO4, 140 mM NaCl, 2.7 mM KCl (pH 7.3)). Typically, the mixture of the peptide (0.2 µM) and varying concentrations of the protein (1-10 µM) was incubated for 10-20 minutes before the reading was taken. The net fluorescence increase (ΔF) at 520 nm due to energy transfer was obtained for each protein concentration by subtracting the background fluorescence caused by the peptide, the protein, and the buffer solution from the total fluorescence. The experimental data were fitted against the equation:

\[ \Delta F = \Delta F_{\text{max}} \cdot \frac{C}{K_d + C} \]

where \( K_d \) is the dissociation constant, \( \Delta F \) is the fluorescence enhancement at a given protein concentration \( C \), and \( \Delta F_{\text{max}} \) is the maximum fluorescence enhancement. Protein concentrations were determined by Bradford assay. The dansylated peptides were dissolved in 50% DMF and their concentrations were determined by measuring the absorbance of the dansyl group at 350 nm (\( \varepsilon_{350} = 4.57 \times 10^3 \, \text{M}^{-1} \, \text{cm}^{-1} \)).

2.3 Results

2.3.1 Design, Synthesis, and Screening of a pTyr Peptide Library

Although the pTyr829 peptide of Rad9 binds to FHA2 in a specific fashion [90], the
binding affinity is weak, raising the question of whether the pTyr peptide binding by FHA2 is physiologically relevant. To determine whether the FHA2 domain is capable of high-affinity binding to pTyr peptides, we have constructed a pTyr peptide library on TentaGel S resin, in which two residues immediately N-terminal and three residues C-terminal to the pTyr residue were randomized. Each of the random positions had an equal representation by 18 natural amino acids (except for cysteine and methionine) plus norleucine (Nle), which was used as a substitute of methionine. Therefore, the theoretical diversity of the library is $19^5$ or $2.48 \times 10^6$. This library contains all of the possible hexapeptides and each bead carries ~100 pmol of a unique peptide sequence. A peptide linker, IBBRM (B=β-alanine), was synthesized to the C terminus of the random region, whereas glutamate and aspartate were added to the N terminus of the random region to improve aqueous solubility. Therefore, each member of the library bears the sequence: acetyl-EDXX(pY)XXXIBBRM-resin (X = random amino acid). In addition to a full-length peptide, each bead also carries a family of truncated peptides derived from the full-length peptide on that bead, which form a peptide ladder in a mass spectroscopy (MS) spectrum to allow for sequence identification of the full-length peptide via peptide ladder sequencing [50] (Figure 2.3).

A total of 900 mg of the peptide library, which contained $2.57 \times 10^6$ beads, was screened against GST-FHA2 in three batches. Statistically, each possible pTyr hexapeptide is represented once in the library. Approximately 150 positive beads (identified by their turquoise color) were selected from this library (Figure 2.4), whereas control screening in the absence of GST-FHA2 always resulted in completely colorless beads. Out of the 150 beads, 53 produced high-quality MS spectra and their sequences
were unambiguously determined (Table 2.1). A total of 25 beads gave only partial sequences C-terminal to the pTyr, with their N-terminal sequences unassigned. The rest of the beads gave poor mass spectra that could not be reliably analyzed. Analysis of the selected sequences revealed that FHA2 clearly binds to pTyr peptides of correct sequence contexts (Figure 2.5). FHA2 is most selective at the +2 position, where Tyr is by far the most preferred residue, followed by the structurally related Phe and, to a lesser extent, Leu and Met (Nle). Clearly, a large hydrophobic side-chain is required for binding at this position. Strong selectivity is also evident at the +1 and +3 positions, where large hydrophobic residues are again preferred (Figure 2.5). At the +1 position, Leu, Nle, and Phe are almost equally represented among the selected sequences, whereas Ile, Val, and Tyr are significantly less frequent. A similar pattern was observed at the +3 position, where Phe, Leu, and Nle are about equally preferred. There is also some selectivity for hydrophobic residues (e.g. Phe, Val, and Leu) at the -1 position, but this selectivity is much weaker than that of C-terminal residues. Very little, if any, selectivity was observed at the -2 position. In summary, the sequence specificity of the FHA2 domain resembles that of SH2 domains in the overall preference for large hydrophobic residues at three residues immediately C-terminal to pTyr. Residues beyond the -2 and +3 positions may also contribute to the overall binding affinity but were not examined here.

2.3.2 Characterization Of Binding of Selected Peptides By Fluorescence Resonance Energy Transfer

Three of the peptides in Table 2.2, each representing a different subfamily, were selected for further analysis. Peptide GF(pY)LYFIR (entry no. 3) features the most
preferred amino acid residues at all positions that are important for binding. In peptide DV(pY)FYMIR (entry no. 28), Phe and Met at the +1 and +3 positions, respectively, are not the most preferred residues. Peptide IQ(pY)IYHIR (entry no. 14) has still less preferred residues at the +1 (Ile) and +3 (His) positions. These three peptides were individually synthesized on large scales with a dansyl group fused to their N termini, and tested for binding to FHA2 using fluorescence resonance energy transfer (FRET) from protein tryptophan residues to the dansyl group. Binding assays were performed at a fixed concentration of the pTyr peptides and varying concentrations of the GST-FHA2 protein (Figure 2.6). Association was monitored by measuring the fluorescence yield of the dansyl group at 520 nm while exciting the tryptophan residues at 290 nm. Peptide GF(pY)LYFIR (no. 3) showed the highest affinity to GST-FHA2, with a $K_d$ of 1.1 µM. Peptide DV(pY)FYMIR bound GST-FHA2 less tightly, with a $K_d$ of 5.0 µM. The third peptide, IQ(pY)IYHIR, showed little fluorescence enhancement with up to 10 µM GST-FHA2 protein so that the $K_d$ value could not be determined by the FRET method (Table 2.2). Our NMR titrations confirmed that this peptide binds to FHA2, but less tightly, with a $K_d$ of 100 µM (see below). To ensure that the observed binding is to the FHA2 domain but not to the GST portion, control experiments were carried out with GST only; no significant fluorescence enhancement was observed under the same conditions (Figure 2.6). Furthermore, when the untagged FHA2 domain was tested for binding to peptide GF(pY)LYFIR, a $K_d$ value of 2.7 µM was obtained (Figure 2.6). These results demonstrate that the selected pTyr peptides are indeed specific ligands for the FHA2 domain.
2.4 Discussion

2.4.1 Agreement Between The Library Screening And The Complex Structure

Dr. In-Ja L. Byeon, a NMR specialist, has attempted to pursue the possibility of determining the structure of the complex between FHA2 and one of the high-affinity pTyr peptides identified from the combinatorial library. Unfortunately, neither the peptide GF (pY) LYFIR nor its derivatives, DGF (pY) LYFIR and DDGF (pY) LYFIR, were sufficiently soluble for NMR studies. Thus, we have utilized the weaker-binding but more soluble peptide, IQ (pY) IYHIR, for our studies. The addition of two charged residues to the N terminus rendered the resulting peptide (Ac-DEIQ (pY) IYHIR) sufficiently soluble to conduct NMR studies to map out the binding site on FHA2. The NMR titration studies showed that the $K_d$ value of this peptide is similar to that of the Rad9 pTyr peptide, and that similar residues were shifted in the $^{15}$N-HSQC spectra. However, the magnitudes of shifts were smaller, and a smaller number of intermolecular NOEs were observed. Thus, complete determination of the complex structure was not performed. However, the results are sufficient to suggest that the two peptides interact with the same set of residues of FHA2.

Both library screening and NMR structure study of the complex between FHA2 and Rad9 pTyr peptide suggest that the residues C-terminal to the pTyr residue of the peptide are involved in the interaction with FHA2. In particular, the result of library screening indicates that FHA2 shows the strongest selectivity at the +2 position. Consistent with this mechanism, the structure shows that a key interaction is mediated by the side-chains of the +2 residue and a hydrophobic surface on the FHA2 domain (Figure 2.7).
2.4.2 Possible Biological Relevance of Our Results

Our results raise some interesting questions. First, why is the Rad9 pTyr829 sequence not selected from the library screening? Is Rad9 pTyr829 the physiological binding site of FHA2? The Rad9 sequence was not selected, most likely because it lacks a hydrophobic side-chain at the +3 position and therefore is weak binding. Among the 78 sequences in Table 2.1, none of them contained an Asp or Glu at the +3 position. Nevertheless, a number of the peptides selected from the library have very similar sequences to the Rad9 peptide (pYYLD): pYYLF (no. 67), pYFLF (no. 40 and 41), pYFLL (no. 71), and pYFLV (no. 72), where large hydrophobic or aromatic residues are selected for the three residues C-terminal to the pTyr. The second question is more difficult to answer, since the actual Rad53 binding site(s) of Rad9 is still unknown. However, among the 13 possible pTyr sites, pTyr829-containing peptide binds to FHA2 with the highest affinity and therefore is the most likely candidate [90] if the biological binding site of FHA2 is a pTyr site. It is less likely that the residues outside the -3 to +3 region are involved in binding, since a Rad9 peptide extended to 13 residues (825DEDI(pY)YLDIRIGD837) did not bind FHA2 any tighter than the shorter peptide (unpublished data). Rather, it may require other Rad9 segments, which are distal in primary sequences but spatially close to pTyr829, to complete the binding to FHA2 of Rad53. It is also possible that FHA2 binds to a pThr site of Rad9, since our preliminary data indicate that FHA2 also binds some pThr peptides (unpublished results).
2.4.3 FHA Domains Have Dual Phosphopeptide Specificity

Liao et al. proposed that FHA domains could have pTyr and pSer/pThr dual specificity, or FHA domains from different proteins could have different specificity [90]. Since the pSer/pThr specificity is widely accepted for FHA domains [89, 91, 94], our demonstration of the pTyr specificity for the FHA2 domain here provides conclusive experimental evidence for our dual specificity proposal. Thus, FHA domains represent the first phosphoprotein-binding module that has pTyr and pSer/pThr dual specificity.
Figure 2.1 The design of the pY peptide library for the screening of FHA2

\[ \text{Ac-EDXXpYXXXIBBRM-} \]

- **B** = β-alanine
- **X** = 19 natural amino acids
  - Excluded: Cys, Met
  - Included: Nle
- Size of Library = \( 19^5 = 2,476,099 \)

Key Feature:
- Each bead carries a unique sequence

TentaGel (100 µM)
- Allows cleavage by CNBr
- Increases aqueous solubility and ionization in MALDI-TOF

Flexible linker
Figure 2.2 Enzyme-linked assay on bead for the screening of FHA2
The sequence: Ac-ENFpYFYLYBBRM*
Figure 2.4 Picture of the three positive beads (blue colored).
Figure 2.5 pTyr peptide sequence specificity for binding to the FHA2 domain. Displayed are the amino acid residues selected at the -2 to +3 positions relative to the pTyr residue. The x axis indicates the identity of the 19 amino acid residues in single letter codes, whereas abundance on the y axis represents the number of occurrence of an amino acid at a certain position (maximum 78). M, norleucine.
Figure 2.6 Binding assay of the peptide DNS-GF(pY)-LYFIR. The concentration of peptide for assay with GST-FHA2 and GST was 96 nM; for assay with FHA2 it was 432 nM.
Figure 2.7 The binding interface for FHA2 and the Rad9 pTyr peptide (a) Surface representation of the FHA2 domain in complex with the peptide; (b) side-chain interactions of the peptide with the FHA2 domain. The backbone ribbons of FHA2 and the peptide are shown in purple and green, respectively. The side-chains of FHA2 and the peptide are shown in red and yellow, respectively. MOLMOL and Insight II were used for (a) and (b), respectively.
<table>
<thead>
<tr>
<th>Entry</th>
<th>Peptide sequence</th>
<th>Entry</th>
<th>Peptide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RLpYLYF</td>
<td>40</td>
<td>RTpYFLF</td>
</tr>
<tr>
<td>2</td>
<td>MLpYLYF</td>
<td>41</td>
<td>SHpYFLF</td>
</tr>
<tr>
<td>3</td>
<td>GFpYLYF</td>
<td>42</td>
<td>PHpYMFF</td>
</tr>
<tr>
<td>4</td>
<td>XXpYLYF</td>
<td>43</td>
<td>LApYMFF</td>
</tr>
<tr>
<td>5</td>
<td>QVpYLYF</td>
<td>44</td>
<td>PPpYKFF</td>
</tr>
<tr>
<td>6</td>
<td>GFpYLYY</td>
<td>45</td>
<td>VLpYLFM</td>
</tr>
<tr>
<td>7</td>
<td>DYpYLYY</td>
<td>46</td>
<td>VFpYMFL</td>
</tr>
<tr>
<td>8</td>
<td>VYpYLFF</td>
<td>47</td>
<td>RVpYLFL</td>
</tr>
<tr>
<td>9</td>
<td>RQpYIYF</td>
<td>48</td>
<td>XXpYIIF</td>
</tr>
<tr>
<td>10</td>
<td>XXpYIYF</td>
<td>49</td>
<td>XXpYMFL</td>
</tr>
<tr>
<td>11</td>
<td>NLpYIYM</td>
<td>50</td>
<td>XXpYFW</td>
</tr>
<tr>
<td>12</td>
<td>XVpYIYM</td>
<td>51</td>
<td>VFpYMLL</td>
</tr>
<tr>
<td>13</td>
<td>VVpYIYH</td>
<td>52</td>
<td>KYpYMML</td>
</tr>
<tr>
<td>14</td>
<td>IQpYIYH</td>
<td>53</td>
<td>GPpYMML</td>
</tr>
<tr>
<td>15</td>
<td>RFpYIYL</td>
<td>54</td>
<td>LYpYLLM</td>
</tr>
<tr>
<td>16</td>
<td>XXpYLYM</td>
<td>55</td>
<td>LPpYLLM</td>
</tr>
<tr>
<td>17</td>
<td>KMpYLYL</td>
<td>56</td>
<td>LDpYVLM</td>
</tr>
<tr>
<td>18</td>
<td>EIpYLYI</td>
<td>57</td>
<td>RPpYMFL</td>
</tr>
<tr>
<td>19</td>
<td>EVpYLYH</td>
<td>58</td>
<td>QFPpYLLY</td>
</tr>
<tr>
<td>20</td>
<td>INpYYVH</td>
<td>59</td>
<td>PFpYMFM</td>
</tr>
<tr>
<td>21</td>
<td>PFpYVYH</td>
<td>60</td>
<td>XPpYMM</td>
</tr>
<tr>
<td>22</td>
<td>VXpYYVL</td>
<td>61</td>
<td>XXpYMLF</td>
</tr>
<tr>
<td>23</td>
<td>IMpYYYM</td>
<td>62</td>
<td>XLpYMML</td>
</tr>
<tr>
<td>24</td>
<td>YHpyYMYL</td>
<td>63</td>
<td>EVpYMF</td>
</tr>
<tr>
<td>25</td>
<td>XIPpYMYM</td>
<td>64</td>
<td>SRpYHY</td>
</tr>
<tr>
<td>26</td>
<td>NFpYFYL</td>
<td>65</td>
<td>NWpYHY</td>
</tr>
<tr>
<td>27</td>
<td>XXpYFYL</td>
<td>66</td>
<td>XXpYFR</td>
</tr>
<tr>
<td>28</td>
<td>DVpYFYM</td>
<td>67</td>
<td>XXpYYLF</td>
</tr>
<tr>
<td>29</td>
<td>LVpYFYM</td>
<td>68</td>
<td>XLpYMIF</td>
</tr>
<tr>
<td>30</td>
<td>VMPpYYYM</td>
<td>69</td>
<td>XXpYLRF</td>
</tr>
<tr>
<td>31</td>
<td>KLPpYEYY</td>
<td>70</td>
<td>LWpYLLY</td>
</tr>
<tr>
<td>32</td>
<td>HWpYFFL</td>
<td>71</td>
<td>LMpYFLL</td>
</tr>
<tr>
<td>33</td>
<td>FPpYFFM</td>
<td>72</td>
<td>SFpYFLV</td>
</tr>
<tr>
<td>34</td>
<td>XMpYFFF</td>
<td>73</td>
<td>GYpYGAL</td>
</tr>
<tr>
<td>35</td>
<td>XXpYFFF</td>
<td>74</td>
<td>GVpYVSI</td>
</tr>
<tr>
<td>36</td>
<td>XIpYFFM</td>
<td>75</td>
<td>XXpYYKI</td>
</tr>
<tr>
<td>37</td>
<td>XXpYFFI</td>
<td>76</td>
<td>XXpYMLI</td>
</tr>
<tr>
<td>38</td>
<td>SVpYYFF</td>
<td>77</td>
<td>SApYIQG</td>
</tr>
<tr>
<td>39</td>
<td>XEpYFFL</td>
<td>78</td>
<td>XXpYPAI</td>
</tr>
</tbody>
</table>

M, norleucine; X, identity unknown.

Table 2.1 Sequences of selected peptides from the library
ND, not determined

<table>
<thead>
<tr>
<th>Peptide</th>
<th>GST-FHA2 (µM)</th>
<th>FHA2 (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GF(pY)LYFIR</td>
<td>1.1 ± 0.3</td>
<td>2.7 ±0.6</td>
</tr>
<tr>
<td>DV(pY)FYMIR</td>
<td>5.0 ±0.6</td>
<td>ND</td>
</tr>
<tr>
<td>IQ(pY)IYHIR</td>
<td>~100</td>
<td>ND</td>
</tr>
</tbody>
</table>

Table 2.2. Dissociation constants of pTyr peptides to FHA2
CHAPTER 3

RAPID SEQUENCING OF LIBRARY DERIVED PEPTIDES BY PARTIAL EDMAN DEGRADATION AND MASS SPECTROMETRY

3.1 Introduction

The well-established peptide chemistry on the solid phase permits the ready access to large combinatorial peptide libraries via Split-Pool synthesis [2-4]. Consequently, screening of peptide libraries have been widely practiced to define the substrate specificity of enzymes [17, 18, 22-27, 32, 33], to develop enzyme inhibitors [72, 95-98], to search for ligands of protein modules or receptors [9, 12-15, 99-102] and to identify new catalysts [103-107]. For resin-bound libraries, high-throughput screening often generates a large number of positive beads that need to be individually characterized. This has created the demand for a rapid and inexpensive method for high-throughput sequencing of library derived peptides. Several methods have been used for this purpose. The first method is to directly sequence the resin-bound peptides by Edman degradation [2, 18, 32-34]. However, Edman sequencing is expensive and time-consuming, making it impractical for the routine sequencing of large numbers of peptides (>100). A variation of the Edman method involves pooling all of the selected peptides
and sequencing them as a mixture [13, 100]. While this method is simple, it loses valuable information about the individual sequences and only gives an overall consensus, which may not necessarily represent the most active sequence. A second method utilizes various types of tags to encode the library and the structure of an active compound is deduced by examining the tags present on the same bead that carries the active compound [35-49]. The drawbacks of the coding strategy include requirement of additional, compatible chemical steps in the library synthesis, potential interference with screening by the coding structure, and limited amount information that can be coded by the tags. Finally, Youngquist et al. have developed a peptide ladder sequencing strategy, which encodes the peptide on a bead by generating a series of sequence-specific, partially terminated products during the synthesis of the library [50]. The sequence of the full-length peptide is determined by analyzing the peptide ladder formed by these termination products in a mass spectrum. A drawback of this method is that, due to the different reactivities of 20 amino acids, the amount of chain termination varies at each position in a non-predictable fashion, resulting in heterogeneous amounts of full-length peptides on all library beads. This biases the screening against those peptides that have higher percentages of termination products (thus less full-length peptides). The use of the same amino acid as capping agent (e.g. use of Boc-Ala as capping agent during the addition of Fmoc-Ala) has partially but not completely solved this problem [108]. Another potential problem is that the termination products may interfere with library screening.

We have adapted a protein ladder sequencing method pioneered by Chait et al [109] to rapidly sequence support-bound peptides derived from combinatorial libraries. In this method, a peptide library is synthesized on solid support without partial chain termination
(except for Ile and Nle) and, therefore, all of the library beads presumably carry an equal amount of full-length peptides. After screening against a molecular target (e.g., a protein), the peptide on a selected bead is subjected to multiple cycles of partial Edman degradation to generate a peptide ladder and its sequence determined by matrix assisted laser desorption ionization (MALDI) mass spectrometry.

3.2 Experimental Procedures

3.2.1 Materials.

TentaGel S NH₂ resin (80–100 µm, 0.26 mmol/g loading, 2.86 × 10⁶ beads/g), N-9-Fluorenlymethoxycarbonyl (Fmoc)-amino acids, 1-hydroxybenzotriazole (HOBr), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were purchased from Advanced ChemTech (Louisville, KY). N-Hydroxysuccinimidobiotin, Acetylglycine, N-acetyl-D, L-alanine, streptavidin-alkaline phosphatase conjugate were obtained from Sigma Chemical Company (St. Louis, MO). Peptide deformylase (PDF) was purified as described [110]. All other chemicals were purchased from Aldrich Chemical (Milwaukee, WI).

3.2.2 Peptide Library Synthesis.

TentaGel S NH₂ resin was used as the solid support for the peptide library, f-XXXXBBRM-resin (X = norleucine or any of the 16 natural amino acids excluding cysteine, lysine, arginine, and methionine; B = β-alanine). Library synthesis was carried out on 1.0-g scale using standard Fmoc/HBTU/HOBt chemistry [92] on a homemade peptide synthesis apparatus. A common four-amino-acid linker BBRM was first
synthesized on the resin. The randomized positions were generated using a split-pool synthesis method to construct the one-bead-one-sequence peptide library. Specifically, after deblocking the N-Fmoc group with 20% (v/v) piperidine in DMF twice (5 + 15 min) and washing five times with DMF, the resin was evenly divided into 17 aliquots and placed into 17 separate reaction vessels. A different amino acid was coupled to the resin in each of the reaction vessels using a fivefold excess of amino acid (2 h). The coupling reaction was repeated once to ensure complete coupling at each randomized position. After coupling is complete, the resin from all vessels was combined, thoroughly mixed, washed with 5 mL of DMF, deprotected twice with 20% piperidine (5 min and then 15 min), and redistributed into the reaction vessels. This process was repeated to produce the tetrapeptide library. During the coupling of norleucine and isoleucine, 10% capping reagents (N-acetylglycine for Ile and N-acetylalanine for Nle) were added into the coupling reactions. After coupling at all four randomized positions and deblocking with piperidine, a formyl group was added to the N-termini of the peptides using 96% formic acid, dicyclohexylcarbodiimide, and dimethylaminopyridine in CH₂Cl₂ (2 × 1 h). Deprotection of side chains was effected with a cocktail containing 4.75 mL of trifluoroacetic acid, 0.1 mL of anisole, 0.1 mL of ethanedithiol, and 0.25 mL of thioanisole for 1 h at room temperature. The resin was washed with CH₂Cl₂ (10 × 4 mL) and stored at 4 °C.

3.2.3 On-Bead Screening of Peptide Library.

Library screening was carried out at room temperature in a 1mL plastic
chromatography column fitted with a filter disc at the bottom (Bio-Rad Laboratories). Approximately 35 mg of resin (~100,000 beads) was placed into the column and washed five times each with 1 mL of DMF and deformylase assay buffer (50 mM sodium phosphate, pH 7.0, 10 mM NaCl). Deformylation was initiated by the addition of 1–5 µg of PDF to the resin in 1.0 mL of 1 × PDF buffer. After gentle shaking for 10 min, the solution was drained with vacuum suction and the resin was washed five times each with 1 mL of doubly-distilled H₂O and DMF. The resin was then incubated in 1 mL of 0.2 mg/mL N-hydroxysuccinimidobiotin in 1:1 (v/v) DMF/sodium phosphate buffer (pH 8.0) for 30 min to derivatize the exposed N-terminal amine. The excess biotin reagent was removed by washing the resin with DMF (6 × 1 mL), deionized water (6 × 1 mL), and TBS buffer (25 mM Tris•HCl, pH 8.0, 150 mM NaCl, and 0.1% Tween 20) (2 × 1 mL). Any nonspecific protein binding sites were blocked by incubating the beads in 1 mL of TBS buffer containing 4 mg/mL bovine serum albumin for 1 h. The buffer was then removed and replaced with 1 mL of TBS buffer containing 0.5 µg/mL streptavidinalkaline phosphatase conjugate. After incubation for 10 min at room temperature, the solution was removed and the beads were washed three times with PBST buffer (40 mM sodium phosphate, pH 7.4, 280 mM NaCl, and 0.2% Tween-20), three times with a high-salt buffer (20 mM Tris•HCl, pH 8.5, 1 M NaCl), and three times with a low-salt buffer (20 mM Tris•HCl, pH 8.5, 10 mM NaCl). The resulting beads were stained by the addition of 1 mL of the low-salt buffer containing 0.5 mg/mL 5-bromo-4-chloro-3-indolyl phosphate (BCIP) and subsequent incubation (with shaking) for 2–4 h. Staining was terminated by washing the resin four times with TBS buffer. Positive beads were readily identified by their intense green color and were manually removed with a
micropipette under a low-power microscope. A control screening was carried out under the same conditions except for the exclusion of PDF treatment. This screening resulted in no colored beads.

### 3.2.4 Partial Edman Degradation.

All of the selected positive beads (total 70) were placed in a 0.5-mL plastic column fitted with an 8-µm glass filter at one end [prepared by inserting a glass filter (ACE Glass, Catalog No. A7421) into a 1.5 mL microcentrifuge tube and cutting off the bottom]. The resin-bound peptides were deformylated by incubating in 1 mL of 7% HCl in methanol for 60 h at room temperature. The beads were resuspended in 250 µL of 1:1 (v/v) pyridine/water and 250 µL of a 2:20:78 (v/v) mixture of phenyl isocyanate (PIC), phenyl isothiocyanate (PITC), and pyridine was added to the reaction vessel. This mixture was incubated at room temperature for 10 min and the solution was drained. The beads were washed three times each with 1 mL of CH₂Cl₂ and dried under the vacuum. The dried beads were treated with 500 µL of anhydrous trifluoroacetic acid (TFA) at room temperature for 10 min, followed by drying under vacuum. This partial Edman degradation procedure was repeated three times. Finally, the beads were suspended in 250 µL of 1:1 pyridine/water and treated with 250 µL of a 2:76:22 (v/v) mixture of PIC, pyridine and deionized water for 10 min at room temperature. This converted any remaining N-terminally unblocked peptides to the corresponding phenylcarbamyl derivatives. After draining the solution, the beads were washed with CH₂Cl₂ and dried under vacuum. The treated beads were picked under a microscope and placed in separate microcentrifuge tubes.
3.2.5 Peptide Sequencing by MALDI Mass Spectrometry.

Individual beads placed in separate microcentrifuge tubes were each treated with 20 µL of 50 mg/mL CNBr in 70% formic acid for 16–24 h in the dark. The released peptide in each tube was lyophilized to remove the excess CNBr and dissolved in 4 µL of 0.1% TFA in H₂O. For mass analysis, a 1 µL aliquot of the TFA solution was mixed with 2 µL of saturated solution of 2, 5-dihydroxybenzoic acid in water and 1 µL of the resulting mixture was applied to the spectrometer plate. MALDI mass analysis was performed on a Bruker Reflex III MALDI-TOF mass spectrometer in the positive ion mode.

3.3 Results

The validity of this method was demonstrated by the synthesis and screening of a peptide library to identify the optimal substrates of *E. coli* peptide deformylase (PDF) [111]. An N-formylated tetrapeptide library, f-XXXXBBRM–resin (X = Ala, Asn, Asp, Gln, Glu, Gly, His, Ile, Leu, Nle, Phe, Pro, Ser, Thr, Trp, Tyr, or Val; B = β-alanine), was synthesized on TentaGel S resin via the split-synthesis method [2-4]. A common linker (BBRM) was added to facilitate enzymatic reaction and mass spectrometric analysis [93], whereas Lys, Arg, Cys, and Met (replaced with Nle) were excluded from the randomized region to simplify the screening process [22]. The resulting one-bead-one-compound library (theoretical diversity = 83,521) was screened against Co(II)-substituted *E. coli* PDF as previously described [22]. Briefly, enzymatic removal of the N-terminal formyl group by PDF exposes a free NH₂ group at the N-terminus of a peptide. The exposed amino group was selectively derivatized with a biotin by incubating the library with N-
hydroxysuccinimidobiotin. The resin was then incubated with a streptavidin-alkaline phosphatase conjugate, which is subsequently recruited to the reacted beads by the bound biotin. Hydrolysis of 5-bromo-4-chloro-3-indolyl phosphate (BCIP) by the bound phosphatase resulted in intense turquoise color at the surface of a positive bead (Figure 3.1). The colored beads were readily identified under a dissecting microscope and manually removed from the library.

Screening of 70 mg of resin (~200,000 beads) resulted in ~100 colored beads and the 70 most colored beads were removed from the library for sequence analysis. All 70 beads were placed in a 0.5-mL column fitted with an 8-µm glass filter at one end. The beads were incubated in a MeOH solution containing 7% hydrochloric acid for 60 h to remove the N-terminal formyl group [112]. The resulting beads were subjected to partial Edman degradation to generate a peptide ladder according to the method originally developed by Chait et al [109] (Figure 3.2). Specifically, the beads were treated with a 10:1 (v/v) mixture of phenyl isothiocyanate (PITC) and phenyl isocyanate (PIC) in pyridine (250 µL), both of which reacted with the N-terminal amine of the support-bound peptides (Figure 3.2b). After washing with CH₂Cl₂ to remove the excess reagents, the beads were treated with trifluoroacetic acid (TFA). For ~90% of the peptides that had reacted with PITC, TFA treatment resulted in the cleavage of the N-terminal amino acid (Edman degradation). For the rest ~10% peptides that had reacted with PIC, no cleavage was possible under the conditions used and these peptides became permanently N-blocked by a phenylcarbamoyl (PC) group. The resulting beads, still placed in the same column, were washed with CH₂Cl₂ to remove the phenylthiohydantoin (PTH)-amino acid derivatives and subjected to the next cycle of partial Edman degradation. Repetition of
the partial Edman degradation reaction for \( n + 1 \) cycles (\( n \) is the number of random positions in a library and equaled to four in this work) resulted in a series of sequence-specific truncation products. Next, the treated beads were individually picked, placed in 70 separate microcentrifuge tubes, and treated with CNBr. This cleaved the peptides off the support at their C-terminal methionine, generating the corresponding peptide homoserine lactones. After lyophilization to remove the excess reagents, the peptides released from each bead were analyzed by MALDI-TOF mass spectrometry as previously described [22] (Figure 3.3). Since each bead carried a unique full-length peptide and four truncation products, MALDI analysis of the cleavage mixture generated a peptide ladder containing five individual peaks. Figure 3.3b shows a typical MALDI spectrum, obtained with 20% of the peptide mixture isolated from a single resin bead. The protonated full-length peptide gave a peak at \( m/z \) 1014.93. The truncation products produced four peaks at \( m/z \) 901.94, 739.01, 576.07, and 519.10. The mass difference between the full-length peptide and the first truncation product (\( m/z \) 901.94) is 113.0, indicating that the N-terminal residue of the peptide is Nle, Leu, or Ile. Likewise, the rest of the residues were determined as Tyr, Tyr, and Gly, respectively, based on the mass differences between adjacent peaks in the ladder. To differentiate norleucine (used as a surrogate for methionine) from the isobaric leucine and isoleucine, a small amount of chain-termination products were generated by adding 10% N-acetylationanine into the coupling reaction of Nle and 10% N-acetylglucoseine into the coupling reaction of Ile during library synthesis [50]. Thus, if a peptide contains a Nle at a given position, its mass spectrum would show at that position a doublet separated by 6 mass units, due to the presence of both the chain-termination product derived from library synthesis and the truncation
product derived from post-screening partial Edman degradation. Similarly, peptides containing Ile should produce doublets separated by 20 mass units at the corresponding positions. The presence of a doublet at m/z 895.93 and 901.94 in Figure 3.3b therefore indicates that the N-terminal residue is a norleucine and the unknown peptide has the sequence of NH2-Nle-Tyr-Tyr-Gly.

Out of the 70 positive beads analyzed by this method, 65 (90%) produced high-quality spectra, which allowed unambiguous sequence assignment at all random positions (Table 3.1). The other 7 beads gave spectra that missed one or a few peaks and, therefore, complete sequence assignment was not possible. It should be noted that no special effort has been made to optimize the degradation reaction or the MALDI MS conditions at this point, and we believe that higher success rate is likely after optimizing these conditions. Analysis of the 65 selected sequences revealed that norleucine is the most preferred amino acid at the P1’ position (36 sequences have norleucine at this position) (Table 3.1 and Figure 3.4). This is consistent with earlier studies [22, 113] and the fact that methionine is the N-terminal residue of all nascent polypeptides, the physiological substrates of PDF. Consistent with these earlier studies, PDF has little selectivity at the P2’ position. At the P3’ and P4’ positions, hydrophobic residues, and tyrosine in particular, are preferred (Figure 3.4). This is again similar to our previous screening results, although the extent of preference for tyrosine is less strong in this work. A number of factors may have contributed to the slight specificity difference observed at the P3’ and P4’ positions including difference in screening stringency, the use of Co-PDF (this work) vs Fe-PDF (used in the earlier work), and bias against peptides containing slow-coupling amino acids such as Val and Thr. We have also recently applied this
method to sequence over a hundred phosphothreonyl peptides selected from a combinatorial library for binding to a forkhead-associated (FHA) domain [114].

3.4 Discussion

The current method offers several advantages over the existing methods. First, this method generates a peptide ladder only after the screening step, so that libraries can be prepared with few chain termination events (except for the homopolymers of Nle and Ile) and therefore, the variation in the quantity of full-length peptides on different beads is minimal. It thus avoids the sequence bias problem associated with the Youngquist method. One additional problem associated with the Youngquist method is that because of the different percentage of capping on the growing peptides for different amino acid coupling, it is very difficult if not impossible to find a capping ratio during the synthesis of the random residues to have all the capped peptides present in uniform quantity. Some of the capped peptides are doomed to be too little to be detected by MALDI-MS, which frequently leads to incomplete sequence information. The current method generates the capped peptide ladders by capping the degrading peptides with PIC capping reagent at each degradation cycle. The percentage of the capping for different amino acid is controlled by the ratio of the capping reagent PIC to the degrading reagent PITC and remains virtually constant if the ratio is kept constant. This leads to more uniform height of the peptide ladders detected in MALDI-MS, which improves the success rate of peptide sequencing. In current study, a 90% success rate of peptide sequencing was achieved without special effort in optimizing the degradation conditions, which forms sharp contrast to the typical <50% success rate we have gotten by using Youngquist
method for peptide sequencing. The new method also simplifies library synthesis and largely eliminates the potential interference with screening by a chain-termination product. Second, since all peptides are support-bound, the operation is very simple; no peptide isolation is necessary after each cycle of Edman degradation and excess reagents and reaction by-products are readily removed by simply washing the beads with solvents. Third, the spatial separation of different peptides on individual beads permits simultaneous Edman degradation of all selected beads (anywhere from one to a few million) in one pot. This tremendously reduces the amount of labor required for sequencing a large number of beads. Finally, this method is highly sensitive, rapid, and yet inexpensive. On a Bruker Reflex III MALDI instrument, we have been able to routinely sequence 20–30 beads in an hour and at a cost of a few US dollars per peptide. Only <20% of the peptides isolated from a single 100-μm bead is needed for each MS experiment. The only limitation of this method is that a peptide to be sequenced must contain only α-amino acids and a free N-terminus or an N-terminal structure that can be converted into a free N-terminus (e.g., N-formyl or N-Fmoc protected) prior to Edman degradation.

In conclusion, a rapid, sensitive, and inexpensive method has been developed to sequence peptides derived from combinatorial libraries. Although the basic concept of peptide ladder sequencing has previously been demonstrated in the solution phase [109], to our knowledge, this work represents the first application of the method to the sequence determination of library derived peptides on the solid phase. We believe that this sequencing strategy will further expand the utility of combinatorial peptide libraries in biomedical research.
Figure 3.1 The enzyme-linked assay used for the screening of peptide deformylase.
Figure 3.2 (a) Chemistry of Edman degradation (b) Partial Edman degradation on bead-bound peptide.
Figure 3.3 Sequencing of support-bound peptides by partial Edman degradation and mass spectrometry. (a) Reactions involved in partial Edman degradation. PITC, phenyl isothiocyanate; PIC, phenyl isocyanate. (b) MALDI mass spectrum of a peptide and its truncation products, obtained with 20% of the material from a single bead. The sequence of the unknown peptide is f-(Nle)YYGBBRM. M, methionine prior to CNBr cleavage and homoserine lactone after CNBr cleavage.
Figure 3.4 Substrate specificity of E. coli Co(II) PDF. Displayed are the amino acids identified at each of the four N-terminal positions (P1' is the N-terminal residue). Number of appearance on the y axis represents the number of selected sequences that contain a particular amino acid at a certain position (out of a total of xx sequences). M, norleucine
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Sequence</th>
<th>Sequence</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFYY</td>
<td>MVVL</td>
<td>MLGP</td>
<td>EMYI</td>
</tr>
<tr>
<td>MGYY</td>
<td>MFVI</td>
<td>MYSH</td>
<td>EEYV</td>
</tr>
<tr>
<td>MGYP</td>
<td>MLVP</td>
<td>MYSE</td>
<td>EFLP</td>
</tr>
<tr>
<td>MPYV</td>
<td>MLVY</td>
<td>MEHT</td>
<td>EYMV</td>
</tr>
<tr>
<td>MHYP</td>
<td>MYVF</td>
<td>MTHL</td>
<td>HMHY</td>
</tr>
<tr>
<td>MMYV</td>
<td>MEAV</td>
<td>MHVH</td>
<td>HATL</td>
</tr>
<tr>
<td>MYYV</td>
<td>MFAT</td>
<td>MMFE</td>
<td>FSPY</td>
</tr>
<tr>
<td>MMYE</td>
<td>MVAE</td>
<td>MSFG</td>
<td>FPHY</td>
</tr>
<tr>
<td>MEYL</td>
<td>MAAA</td>
<td>MMEE</td>
<td>AVLL</td>
</tr>
<tr>
<td>MYPF</td>
<td>MPAH</td>
<td>MMPL</td>
<td>AVFF</td>
</tr>
<tr>
<td>MYYG</td>
<td>MLLY</td>
<td>MMPV</td>
<td>SHYH</td>
</tr>
<tr>
<td>MTYS</td>
<td>MILY</td>
<td>MVTI</td>
<td>LETP</td>
</tr>
<tr>
<td>MPVH</td>
<td>MLLA</td>
<td>YFLF</td>
<td>GFLL</td>
</tr>
<tr>
<td>MHVH</td>
<td>MLLP</td>
<td>YYYG</td>
<td>PDLG</td>
</tr>
<tr>
<td>MVVL</td>
<td>MLLS</td>
<td>YPGY</td>
<td></td>
</tr>
<tr>
<td>MVVL</td>
<td>MYGV</td>
<td>YYGP</td>
<td></td>
</tr>
<tr>
<td>MVVL</td>
<td>MVGV</td>
<td>YAMY</td>
<td></td>
</tr>
</tbody>
</table>

*M, norleucine.

Table 3.1 Selected Sequences for *E. coli* PDF*
CHAPTER 4

SCREENING COMBINATORIAL LIBRARIES FOR OPTIMAL SUBSTRATES
OF PEPTIDE DEFOMYLASE BY MASS SPECTROMETRY

4.1 Introduction

As we already discussed in the first chapter, combinatorial chemistry has become a powerful tool for the discovery and optimization of ligands (e.g., peptides, oligonucleotides, and small organic molecules) that bind to a variety of enzymes and protein receptors. On the other hand, screening of combinatorial libraries for efficient catalysis (e.g., identification of effective catalysts for a chemical reaction or the optimal substrates of an enzyme) has been a considerably more challenging problem [115]. Because in a catalytic event, the catalyst binds to a substrate(s), converts the substrate(s) into a product(s), and then dissociates from the product(s), it leaves behind no record of the interaction that occurred between the catalyst and the substrate/produces. In the case of identifying optimal enzyme substrates, the challenge lies in the separation/identification of a reaction product(s) from a complex mixture of potential substrates.

A number of combinatorial library methods have been reported for the systematic
identification of optimal enzyme substrates. The more commonly used methods include phage display [23-27], selective labeling of reaction products with radioactive or fluorescent tags [32, 33], quenched fluorescence [17-19], and enzyme-linked assays [22]. However, these methods were each developed for a specific enzyme or a class of enzymes (e.g., endoproteases) and are not generally applicable to other enzymes. Although phage display has been widely used to determine the substrate specificity of endoproteases, it does not work with other enzymes such as aminopeptidases, acylases, and protein phosphatases. The latter enzymes also require posttranslationally modified substrates (e.g., phosphopeptides), which cannot be bio-synthesized by the ribosome. The quenched fluorescence method suffers from a similar drawback, and its application has so far been limited to endoproteases. Finally, the methods that involve selective derivatization of an enzymatic reaction product with a radioactive/fluorescent tag or a linking enzyme often require the exclusion of certain functional groups from the initial library, resulting in incomplete libraries.

In this chapter, we present a novel methodology for the rapid identification of optimal enzyme substrates from a combinatorial library by mass spectrometry. In this method, we have overcome the problem of product identification by partially labeling the substrates with a heavier isotope (heavy/normal isotope = 1:1), so that each member of the substrate library appears as a doublet peak in a mass spectrum. Enzymatic reaction removes the functional group that contains the isotopic label and the product(s) appears as a singlet peak(s) in the spectrum, allowing for its unambiguous identification. Here we demonstrate the validity of this methodology by determining the substrate specificity of peptide deformylase (PDF), an antibacterial drug target under study in one of our
laboratories. This method should be readily applicable to a variety of enzymes or
synthetic catalysts that catalyze the removal, addition, or substitution of a functional
group from/onto a substrate.

4.2 Experimental Procedures

4.2.1 Materials

All materials for peptide synthesis including resins, protected amino acids, and
coupling reagents were purchased from Advanced ChemTech (Louisville, KY, USA).
Other chemicals were from Aldrich (Milwaukee, WI, USA). PDF enzyme was purified as
previously described [116].

4.2.2 Synthesis of The Peptide Library

A tripeptide library, formyl-X₁X₂R, was constructed with 19 different amino acids at
the two randomized positions. The amino acids used were Ala, Arg, Asp, Asn, Glu, Gln,
Gly, His, He, Leu, Lys, Phe, Pro, Ser, Thr, Trp, Tyr, Val, and (Nle). Cysteine and
methionine were excluded from the library due to their tendency to oxidation and Nle
was used as an isosteric replacement of Met. Initial studies showed that Met-containing
peptides consistently gave weak signals, probably due to oxidation of Met into
methionine sulfoxide, which overlaps with the corresponding Phe-containing peptides.
An arginine was added at the C-terminus to improve aqueous solubility and ionization
efficiency. Library synthesis was performed on 1.0 g of Wang resin (0.9 mmol/g) using
standard Fmoc/HBTU/HOBt chemistry. The randomized positions were generated by the
Split-Pool synthesis method [2-4]. The resin was equally divided into 19 portions and
placed into the 19 1mL reacting vessels. A different Fmoc-amino acid (5-fold excess) was placed into each vessel and the coupling reaction was allowed to proceed for 3 h at room temperature. The coupling reaction was repeated once to ensure completion reaction. After removal of the N-Fmoc group, the second random position was generated in the same manner. Finally, the N-termini of the peptides were formylated by using a 1:1 (mol/mol) mixture of HCOOH and DCOOD (10-fold excess) in the presence of dicyclohexylcarbodiimide and dimethylaminopyridine in CH$_2$Cl$_2$. The reaction was complete in 3 h (as monitored by ninhydrin tests). Deprotection of peptide side chains and cleavage of the peptides off the solid support were effected with a cocktail containing 4.75 mL of trifluoroacetic acid, 0.1 mL of anisole, 0.1 mL of ethanedithiol, and 0.25 mL of thioanisole for 2 h at room temperature. The crude peptide solution was dried under a gentle flow of nitrogen gas and the residual solid was triturated three times each with 10 mL of diethyl ether. The resulting peptides were dried in air and stored at –20 °C.

4.2.3 Library Screening and MS Sequencing

Typically, ~20 mg of the peptide library were dissolved in 200 µL of ethanol and diluted in ~40 mL of a buffer containing 1.5 mM ammonium carbonate (pH 7.0). The solution was centrifuged at 3000 rpm for 10 min and clear supernatant was divided into four equal aliquots (10 mL each). One of the aliquots was kept as control, whereas the other three aliquots were each treated with 2 µL of Escherichia coli Co(II)-PDF (23 µg/µL). The PDF reactions were allowed to proceed at room temperature for 5, 10, and 40 min, respectively, before being quenched by heating at 100 °C for 15 min (the enzyme is typically inactivated within the first 30 s). The resulting solution was quickly frozen in
liquid nitrogen and lyophilized for 20 h to obtain a white powder. The powder (~5 mg) was dissolved in 50 µL of a solvent mixture containing 50:50:0.5 (v/v) methanol/water/acetic acid prior to ESI-MS analysis. Library screening was performed on a Bruker Apex Ile 7-tesla Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Billerica, MA, USA). The sample was infused in an external microelectrospray source [117, 118] at a rate of 15 µL/h. Typical electrospray ionization (ESI) conditions were: needle voltage 2.2 kV, capillary exit 100 V, skimmer 5 V and heated capillary temperature 90 °C. Ions were accumulated in a linear hexapole ion trap for 0.2 s and then transferred to a 3"-Penning trap (1.5 V trapping voltage) by the electrostatic ion transfer optics and trapped using the sidekick technique. Typical base pressure for the instrument was 1.5 × 10^{-10} Torr. A MIDAS data station [119] controlled all experiments. Product peaks were identified by manual inspection of the expanded spectrum and searching for the presence of singlet peaks.

MS/MS spectra of selected product peaks were acquired on a Micromass Q-TOF2 spectrometer. Ions within a 1-Da window surrounding a desired product peak were selected for fragmentation analysis through collision induced dissociation (CID). For some of the product ions, there were other peaks derived from the initial library within the 1-Da window and, therefore, the product ions selected for secondary MS analysis were contaminated to varying degrees by species from the initial library. However, the presence of these contaminants did not affect the sequence assignment of the product peaks.
4.3 Results

We chose PDF as a model system to demonstrate the validity of this methodology. PDF is essential for protein maturation in bacteria by hydrolytic removal of the N-terminal formyl group from newly synthesized proteins and is an attractive target for antibacterial drug design [120]. Efficient deformylation by PDF requires at least a dipeptide and, for E. coli PDF, the most preferred substrate has the consensus sequence of f-MX(F/Y/K/R)Y (X= any amino acid) [22]. To demonstrate the methodology, a 361-member tripeptide library, X₁X₂R (X₁ and X₂= Nle or any of the 18 natural amino acids except for Cys and Met), was synthesized and an isotopically labeled formyl group (HCO/DCO ~1:1) was added to the N-termini of all peptides. Due to the existence of natural isotopes (e.g., ¹³C) which contribute to the M +1 peak, the HCO/DCO ratio was slightly adjusted so that the M and M +1 peaks had approximately equal intensity. In a mass spectrum, each member of the resulting library should give a doublet separated by 1.0063 Da, corresponding to the monoisotopic and the deuterium-labeled peptides, respectively (Figure 4.1a). However, upon treatment of the library with PDF, a deformylated product should appear as a singlet, due to loss of the mass-degenerate formyl group (Figure 4.1b). Thus, PDF product peaks are readily identified from a mass spectrum by searching for the appearance of singlet peaks. The identified product peaks are verified by loss/reduction of the corresponding substrate peaks at m/z m + 28 and m + 29, and then selected for MS/MS sequencing through collision-induced fragmentation (Figure 4.1c).

Figure 4.2 shows the mass spectra of the 361-member library, obtained in positive ion mode on the 7-Tesla ESI FTICR mass spectrometer, before (top) and after treatment with
PDF (bottom). The control experiment without enzyme treatment (Figure 4.2, top spectrum) was carried out to verify that all of the doublet peaks present were attributable to labeled peptides. As expected, all of the peptides in the initial library existed as doublet peaks. Due to the presence of isobaric peptides of different sequences (e.g., GAR vs. AGR, GTR vs. ASR, and AYR vs. FSR), the number of observed doublets was significantly less than 361. Based on the calculated molecular masses, the library should produce 141 distinct doublets in the m/z range 317-576, most of which were indeed observed and well resolved in the spectrum. The observed difference in peak intensity was likely due to different ionization efficiencies and aqueous solubility as well as the presence of isobaric peptides. After treatment of the library (5 mg) with E. coli Co(II)-PDF (46 ng) for 40 min, a number of singlet peaks emerged and were readily visible in the overall spectrum (the labeled peaks in Figure 4.2). Manual inspection of the expanded spectrum (two sections of are shown in Figure 4.3a) revealed the appearance of 22 peaks of various m/z ratios (Table 4.1). Since these singlets were not present in the initial library and exhibited increasing intensities as the enzyme incubation time was increased (from 5 min to 10 min and then to 40 min), they were assigned as the product peaks. Consistent with this interpretation, the corresponding substrate peaks (at m/z m + 28 and m + 29) had reduced intensity or completely disappeared (Figure. 4.3a and data not shown). In the vast majority of cases, the product peaks were well resolved from the substrate peaks. The smallest mass difference between an observed product and an initial library member was 0.0029 Da, between product MER (calculated m/z 417.2462 Da) and substrate DCO-GRR (calculated m/z 417.2433 Da), which require a mass resolving power of >144000 for their separation. Since the mass resolving power we observed was
104,000 at m/z 500, we were unable to clearly separate these two peaks. MER and another product (MDR at m/z 403.230) also had identical masses with substrates f-MSR and f-MTR, respectively. However, these products were still unambiguously identified by the observation that the peaks at m/z 403.245 and 417.261 had significantly higher intensities than the corresponding M + 1 peaks. Since the limited PDF reaction was carried out in solution under competitive conditions, the most preferred substrates should react first. Therefore, the observed singlet peaks should be derived from the most preferred substrates of PDF.

Eight relatively abundant singlets (or product ions) (m/z 359.25, 375.25, 385.27, 387.29, 401.30, 435.29, 451.29, and 474.30) were selected for MS/MS sequencing analysis on the ESI Q-TOF spectrometer (Figure 4.3b). Possible amino acid composition was generated from the observed accurate molecular mass as determined by FTICR MS during library screening. All of these peptides contained an N-terminal amino acid with the residue mass of 113.08, which was assigned as a norleucine. Earlier studies had previously shown that E. coli PDF strongly prefers an N-terminal Met or Nle. It has high activity toward certain N-formylphenylalanyl peptides but lacks significant activity toward N-formylleucyl and N-formylisoleucyl peptides [22, 121]. The penultimate residues of these peptides were determined as Ala, Ser, Pro, Val, Leu (or Ile, Nle), Phe, Tyr, and Trp, respectively (Table 4.1). The sequences of the other 14 products were assigned on the basis of their accurate molecular masses. Eight of these peptides were consistent with having a Nle as the N-terminal residue and Thr, Asn, Asp, Gln, Lys, Glu, His, or Arg as the penultimate residue. The last 6 peptides contained a Phe at their N-termini and Glu, Ser, Thr, Trp, Tyr, and Val at their penultimate. MS/MS analysis also
revealed that the peak at m/z 435.29 contained a mixture of two isobaric peptides, MFR (major) and FMR (minor) (Figure 4.3b). The N-phenylalanyl peptides in general had lower intensity than the Norleucyl peptides (Figure 4.2). When the library was treated with PDF for a shorter time (5 or 10 min), the N-phenylalanyl peptides were absent in the spectra, whereas most of the N-norleucyl peptides were already present (data not shown). These observations are consistent with the N-formylphenylalanyl peptides being poorer substrates of the enzyme. Thus, our data suggest that PDF strongly prefers a norleucine and, to a lesser extent, a phenylalanine as the N-terminal residue, whereas it has little selectivity at the penultimate position. This is in excellent agreement with earlier literature data [22, 113] and demonstrates the effectiveness of this method in the identification of optimal enzyme substrates from a combinatorial library.

4.4 Discussion

Information on the substrate specificity of an enzyme has at least three important applications. First, it helps identify the physiological substrate(s) of the enzyme and therefore the cellular function of the enzyme. Second, the specificity data facilitate the design of specific inhibitors against that enzyme. Finally, optimal substrates of an enzyme provide more effective kinetic assays for the enzyme. Therefore, in recent years, considerable efforts from many laboratories have been devoted to develop combinatorial library methods to systematically study the substrate specificity of various enzymes [16-27]. However, as discussed earlier, none of the existing methods are general methods that can be effectively applied to different classes of enzymes. In fact, for several important
classes of enzymes such as protein serine/threonine phosphatases and methionine aminopeptidases, there has been no effective methodology available to determine their substrate specificity.

We have developed a novel method for the rapid identification of optimal enzyme substrates by direct screening of combinatorial libraries using mass spectrometry. As discussed earlier, the greatest challenge in screening efficient catalysis is how to differentiate a reaction product from a complex library of substrates. We have overcome this difficulty by partially labeling all of the substrates with a heavier isotope so that they appear as doublets in a mass spectrum, whereas a product appears as a singlet due to enzymatic removal of the isotopic label. We envision that this strategy should also work in the reverse direction, i.e., all of the substrates are singlets (no isotopic label) but are converted into doublets as a result of enzymatic addition/substitution of a partially isotopically labeled group onto the substrates. Thus, our method should be generally applicable to any enzyme or synthetic catalyst that catalyzes the removal, addition, or substitution of a functional group from/onto a substrate. Another advantage of our method is that the enzymatic reaction is carried out in the solution phase, thus eliminating any potential impact of the solid support on enzyme specificity. Finally, the current method does not require the exclusion of any amino acid from the library. Cysteine was positions excluded from the library in this work to simplify sample handling. However, libraries containing all 20 natural amino acids should be easily screened by this method in the-presence of reducing agents.

In this work, the library was analyzed on a 7-tesla FTICR mass spectrometer. Although the high resolving power of FTICR MS (104,000 at m/z 500) was very helpful in
identifying the products ions, we believe that library screening on a conventional mass spectrometer with a resolving power of 10000-20000 (e.g., ESI-TOF MS or MALDI-TOF MS) should also be possible. Since product identification is based on the peak shape (singlet vs. doublet), a product peak overlapped with a substrate peak can still be identified by the uneven intensities of the pair of peaks (as demonstrated by the identification of products MDR and MER in this work). Obviously, a small product peak could potentially be buried under an intense substrate peak and escape detection. One possible solution is to increase the enzymatic reaction time so that the product peak becomes more prominent. Alternatively, one can fractionate the enzyme-treated library into less complex mixtures prior to MS analysis. Another factor that complicates the product identification is the presence of natural isotopes that contribute to the M + 1 peak (e.g., $^{13}$C). This problem can be alleviated by using isotopic labels that produce peaks at M + 2 or M + 3 (e.g., HC$^{18}$O-), where the natural isotopic peaks are much smaller.

It is worth noting that Benner and co-workers have previously used FTICR MS to screen a 16-member library for the optimal substrate of glutathione-S-transferase [73]. Their strategy was to directly compare the mass spectrum of the enzyme-treated library with that of the starting library (before enzyme treatment), and identify reaction products by the appearance of new peaks in the mass spectrum. As such, the Benner method is limited to small libraries in which none of the starting library members are isobaric with the product(s), a condition unlikely to be met by most larger libraries. For example, in the 361-member library used in this work, there are at least two starting library members (f-MSR and f-MTR) that are isobaric with two of the products (MDR and MER, respectively). Also, because the Benner method requires comparison with the starting
library for product combinatorial libraries for identification, it mandates very high reproducibility in obtaining the mass spectra. Such reproducibility may be difficult to achieve if HPLC fractionation of a large library is necessary to resolve any overlapping peaks, as it is highly unlikely that exactly the same set of peptides (from a complex library) would end up in the same fraction in two separate HPLC runs. Unlike the Benner method, our method identifies a product based on its peak shape (singlet vs., doublet) and, therefore, it does not require comparison with the starting library. Consequently, in our method, the mass spectrometer may be readily interfaced with various on-line separation techniques (e.g., HPLC), and the library may be screened in a continuous, automated fashion. As shown above, our method is capable of identifying products even when they are isobaric with members of the starting library. Although it has been demonstrated on a relatively small library in this work, the method should work with larger libraries when coupled with liquid chromatography or other separation techniques. Its application to other enzymes and more diverse libraries is currently under way in our laboratory.
Figure 4.1 Scheme showing library screening by mass spectrometry. Members of the starting library each appear as a pair of doublets at mass/charge M and M+1, where a product appears as singlet at mass/charge M-28 in the mass spectrum.
Figure 4.2 ESI FTICR mass spectra (positive ion mode) of the 361-member peptide library before enzymatic treatment (top) and after treatment with E. coli. Co(II)-PDF (46ug) for 40 min (bottom). Peaks numbered 1-16 (according to the numbering in Table 4.1) indicate some of the more prominent product peaks identified. Peaks labeled with a * are due to unknown contaminants.
Figure 4.3 (a) Expanded spectrum of the box regions of Figure 4.2, showing the decrease in intensity of substrate peptides (f-MFR and f-FMR at $m/z$ 463.2 (peak a) and 464.2 (peak a+1) and the appearance of a product peak b (MFR and FMR) at $m/z$ 435.2 as a result of PDF treatment. (b) A Q-tof MS/MS spectrum of the product peak b ($m/z$ 435.2). M, norleucine.
<table>
<thead>
<tr>
<th>entry No</th>
<th>sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>observed mass</th>
<th>calculated mass</th>
<th>intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MAR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>359.252</td>
<td>359.241</td>
<td>1.6</td>
</tr>
<tr>
<td>2</td>
<td>MSR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>375.248</td>
<td>375.236</td>
<td>1.9</td>
</tr>
<tr>
<td>3</td>
<td>MPR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>385.269</td>
<td>385.256</td>
<td>2.3</td>
</tr>
<tr>
<td>4</td>
<td>MVR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>387.285</td>
<td>387.272</td>
<td>5.5</td>
</tr>
<tr>
<td>5</td>
<td>MTR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>389.265</td>
<td>389.251</td>
<td>2.6</td>
</tr>
<tr>
<td>6</td>
<td>M(I/L/M)R&lt;sup&gt;b&lt;/sup&gt;</td>
<td>401.302</td>
<td>401.287</td>
<td>17</td>
</tr>
<tr>
<td>7</td>
<td>MNR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>402.261</td>
<td>402.246</td>
<td>1.2</td>
</tr>
<tr>
<td>8</td>
<td>MDR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>403.245</td>
<td>403.230</td>
<td>1.7</td>
</tr>
<tr>
<td>9</td>
<td>MQR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>416.278</td>
<td>416.262</td>
<td>2.6</td>
</tr>
<tr>
<td>10</td>
<td>MKR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>416.314</td>
<td>416.298</td>
<td>1.0</td>
</tr>
<tr>
<td>11</td>
<td>MER&lt;sup&gt;c&lt;/sup&gt;</td>
<td>417.261</td>
<td>417.246</td>
<td>3.7</td>
</tr>
<tr>
<td>12</td>
<td>MHR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>425.279</td>
<td>425.262</td>
<td>0.91</td>
</tr>
<tr>
<td>13</td>
<td>MFR&lt;sup&gt;b&lt;/sup&gt;/FMR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>435.289</td>
<td>435.272</td>
<td>9.0</td>
</tr>
<tr>
<td>14</td>
<td>MRR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>444.322</td>
<td>444.305</td>
<td>0.53</td>
</tr>
<tr>
<td>15</td>
<td>MYR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>451.285</td>
<td>451.267</td>
<td>3.8</td>
</tr>
<tr>
<td>16</td>
<td>MWR&lt;sup&gt;b&lt;/sup&gt;</td>
<td>474.303</td>
<td>474.283</td>
<td>6.0</td>
</tr>
<tr>
<td>17</td>
<td>FVR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>421.272</td>
<td>421.256</td>
<td>1.0</td>
</tr>
<tr>
<td>18</td>
<td>FSR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>409.235</td>
<td>409.220</td>
<td>0.37</td>
</tr>
<tr>
<td>19</td>
<td>FTR&lt;sup&gt;c&lt;/sup&gt;</td>
<td>423.246</td>
<td>423.236</td>
<td>0.21</td>
</tr>
<tr>
<td>20</td>
<td>FER&lt;sup&gt;c&lt;/sup&gt;</td>
<td>451.249</td>
<td>451.231</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 4.1. The Most Abundant Products from PDF Treated Library
CHAPTER 5

SCREENING COMBINATORIAL LIBRARIES FOR OPTIMAL ENZYME SUBSTRATES OF PROTEIN TYROSINE PHOSPHATASE SHP-1 BY MASS SPECTROMETRY

5.1 Introduction

In Chapter 4, we have discussed a novel combinatorial method, which we now refer to as “enzyme-catalyzed loss of isotope peak signal enhancement (ECLIPSE)”, for the identification of optimal enzyme substrates. In this method, distinction between substrates and products is achieved by partially labeling the substrates with a heavier isotope (heavy/normal isotope = 1:1), so that each member of the substrate library appears as a doublet in a mass spectrum. Enzymatic reaction removes the functional group that contains the isotopic label and the products appear as singlets in the spectrum, allowing for their unambiguous identification. We now report the application of this method to the determination of the subsite specificity of protein tyrosine phosphatase SHP-1. In this case, one of the non-bridging phosphate oxygen atoms of pY is labeled with 50% $^{18}$O. Thus, each pY peptide should appear as a doublet in a mass spectrum, separated by 2.0043 Da. However, catalytic turnover by a PTP removes the phosphoryl
Reversible phosphorylation of proteins on tyrosyl residues is one of the key events that mediate the execution and regulation of many cellular processes. A proper level of phosphorylation is critical for these processes and is controlled by the opposing functions of protein tyrosine kinases (PTKs) and protein tyrosine phosphatases (PTPs). Indeed, the imbalance of these two activities has been associated with a number of human diseases and conditions [122]. A large number of PTKs and PTPs have been identified but their precise mechanisms of action in vivo have been largely unknown. A major challenge in this field has been the determination of the in vivo targets/substrates of these PTKs and PTPs.

One approach to identifying potential substrates of PTKs and PTPs is to study their in vitro substrate specificity using purified proteins or synthetic peptides, followed by database searches using the consensus peptide sequence(s). Substrate specificity data would also be very useful in the design of selective inhibitors to modulate cellular signaling processes. Several laboratories have used synthetic phosphotyrosyl (pY) peptides corresponding to known phosphorylation sites in proteins to define the substrate specificity of PTPs [123-128]. While these studies have clearly demonstrated the existence of primary sequence specificity (i.e., PTPs recognize the linear sequence flanking the pY residue), they have yet to produce a consensus sequence for any PTP.
This is because the PTP active site makes contacts with 3–5 residues on either the N- or C-terminal side of pY [129, 130]. A complete characterization of each PTP would require the synthesis and testing of a prohibitively large number of single peptides ($20^6$–$20^{10}$). Therefore, more recent efforts have been focused on various combinatorial approaches, in which peptide libraries are synthesized simultaneously and screened against PTPs of interest. However, a major difficulty of the combinatorial approaches is to find a reliable method to separate/identify a reaction product (a tyrosyl peptide) from a complex mixture of substrates (pY peptides). Cheung et al. have employed $\alpha$-chymotrypsin to discriminate between tyrosyl and pY peptides in a resin-bound library, based on the observation that $\alpha$-chymotrypsin cleaves the dephosphorylated products after the tyrosyl residue but not the unreacted pY peptides [33]. The drawback of this method is that it requires the exclusion of all aromatic and basic amino acids (e.g., Y, F, W, H, K, and R) from the library. A solution-phase library approach has also been reported independently by two laboratories [62, 63]. In this method, peptide libraries containing nonhydrolyzable pY analogues were incubated with PTP1B and the resulting enzyme-peptide complexes were separated from unbound peptides by affinity or size-exclusion chromatography. The bound peptides were then analyzed by electrospray ionization mass spectrometry (ESI-MS). The major drawback of this method is that it only selects for high-affinity peptides to a PTP, which may or may not be a good substrate of the enzyme.

Using our ECLIPES method, we can systematically examine the substrate specificity of protein tyrosine phosphatase by screening a pY peptide library in solution phase, therefore, eliminating the drawbacks of using pY analogues and exclusion of all aromatic and basic amino acids.
5.2 Experimental Procedures

5.2.1 Materials

Peptide synthesis reagents and resins were purchased from Advanced ChemTech (Louisville, KY). \(^{18}\)O-Enriched water (95% \(^{18}\)O) was purchased from Isotech Inc. (Oxford, OH). All other chemicals were obtained from Aldrich or Sigma. The catalytic domain of SHP-1, SHP-1 (\(\Delta\)SH2), was purified from a recombinant Escherichia coli strain as previously described [131]. Protein concentrations were determined by Bradford assay using bovine serum albumin as standard. Concentrations of pY peptides were determined by complete hydrolysis with SHP-1 (\(\Delta\)SH2) and measuring the absorbance increase at 282 nm (\(\varepsilon = 826 \text{ M}^{-1}\text{cm}^{-1}\) at pH 7.0).

5.2.2 Synthesis of t-Butyl N\(^{\alpha}\)-[(Fluoren-9-yl)methoxylcarbonyl]-L-tyrosinate (1)

See Appendix B

5.2.3 Synthesis of \(^{18}\)O-Labeled N\(^{\alpha}\)-[(Fluoren-9-yl)methoxylcarbonyl]-O-[bis(allyloxy) phosphoryl]-L-tyrosine (4)

See Appendix B

5.2.4 Peptide Synthesis

The pY peptide library RNNXXpYA-NH\(_2\) was constructed on 1.0 g of Rink resin (0.7 mol/g) using standard Fmoc/HBTU/HOBt solid-phase peptide chemistry and \(^{18}\)O-labeled pY building block 4. The two randomized positions were prepared by the Split-Pool synthesis method [2-4] with 19 different amino acids (cysteine and methionine are
excluded from the library and norleucine is included as methionine replacement). The coupling reactions were carried out with 5 equiv of Fmoc-protected amino acids for 3-4 h and repeated once to ensure complete reaction. After removal of the N-terminal Fmoc group with 20% piperidine in DMF, the resin-bound peptides were treated for 14 h with a cocktail containing diethylamine (10 equiv), HCOOH (10 equiv), Pd(PPh3)₄ (0.05 equiv), and PPh₃ (0.1 equiv) in THF to remove the O-allyl groups. Deprotection of peptide side chains and cleavage off the resin (~300 mg) were carried out with 5 mL of TFA plus a small amount of anisole (100 µL), ethanedithiol (150 µL), and thioanisole (250 µL) for 2 h at room temperature. TFA and other volatile substances were removed under a gentle flow of nitrogen, and the residue was triturated for 5 times with diethyl ether. The resulting peptides were obtained as a white solid.

Individual pY peptides were synthesized in a similar manner using the Rink resin and unprotected N-Fmoc-pY (Advanced ChemTech). The identity of all peptides was confirmed by MALDI mass spectrometric analysis. With the exception of peptide RNNTQpYA-NH₂, which was purified by HPLC before enzymatic assays, all other peptides showed good purity (>80%) by HPLC and/or MALDI analysis and were used in PTP assays without further purification.

5.2.5 Library Screening

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

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

5.2.6 Mass Spectrometry

The mass spectra of the peptide library were acquired on a Bruker Apex Ile 7-tesla Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (Bellerica, MA) under the positive ion mode. Samples were infused in an external microelectrospray source [117] at a rate of 15 µL/h. Typical experimental conditions at ESI source were the following: needle voltage at 2000–2300 V, capillary exit at 200 V, skimmer at 5 V and heated capillary temperature at 85 °C. The experiment event sequence and their corresponding parameters were as described below. Ions were accumulated in a linear hexapole ion trap for 1.0 s and then transferred to a 3-inch Penning trap with 1.5 V trapping voltage. The broadband ion cyclotron excitation was conducted by chirp mode with excitation amplitude of 1.0 V. A MIDAS data station [119] was used to control the experiment event sequence and to acquire data. Typically, 1 M broadband time-
domain data were collected repetitively by ~1000 acquisitions in order to improve the signal to noise ratio by averaging all the acquisition signals. Hamming apodization was applied to the data before it was processed by Fast Fourier Transform to yield the FTICR MS spectrum. The product peaks were identified by manual inspection of the expanded spectrum and searching for the presence of singlet peaks.

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

5.2.7 On-Line HPLC-MS Analysis of Starting Peptide Library

A 150 × 4.6 mm C18 column from Vydac (Hesperia, CA) was directly coupled to a Micromass Q-TOF2 mass spectrometer (Manchester, UK). Gradient elution was performed with (A) water + 0.5% formic acid and (B) Acetonitrile + 0.5% formic acid as mobile phase. The gradient started at 10% B and went to 100% B in 60 min. The flow rate was 1 mL/min. The fraction coming out of the column was split and only about 1 µl/min of the flow was introduced into the mass spectrometer by electrospray for detection. Typically, 100 µl of saturated pY peptide library solution was injected into HPLC.
5.2.8 PTP Assay

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

5.3 Results

5.3.1 Library Design and Synthesis

A 361-member peptide library RNNXXpYA-NH$_2$ ($X = \text{Nle}$ or any of the 18 natural amino acids except for Cys and Met), in which the two residues immediately N-terminal to pY were randomized, was designed to evaluate the contribution of the $-1$ and $-2$ residues to PTP substrate recognition and catalysis. Methionine was replaced with norleucine (M) to avoid any oxidation at the side chain during experiments. The N-terminal RNN motif was added to improve the solubility of the library peptides. The arginine would also enhance the ionization efficiency of the peptides during MS analysis. An alanine was placed at the $+1$ position to avoid any unfavorable interactions between a peptide and the PTP active site (Figure 5.2). To facilitate library screening by mass spectrometry, the pY residue was partially labeled with $^{18}\text{O}$ at one of the non-bridging
phosphate oxygen atoms ($^{16}\text{O}/^{18}\text{O} = 1:1$) (Figure 5.1).

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

5.3.2 MS Analysis of Untreated Library

Theoretically, the peptide library should contain 361 distinct peptides with a molecular-mass range between 830.3 and 1090.4 Da. Due to mass degeneracy (e.g., peptides RNNDFpYA-NH$_2$ and RNNFDpYA-NH$_2$ share the same exact mass), however, these peptides have only 141 unique masses. FTICR-mass spectrometer, like other trapping mass analyzers, has a limit for the maximum ions that can be stored in the trap without significantly affecting its performance. The ion-ion interactions or “space charge effect” ought to be minimized by limiting the population of the ions in the trap. The signal to noise ratio can improved through signal averaging by summing a number of time domain transients prior to the Fourier transformation. Therefore, in order to obtain high resolution and high accuracy mass spectra for the 361-member peptide library, ~1000 time-domain transients were acquired over 2 hours, summed and subsequently transformed to frequency domain signal by Fourier transform. Analysis of the untreated library on a 7 T FTICR mass spectrometer (average resolving power ~100,000 at m/z 900) indicated that virtually all of the peptides with distinct masses were well resolved in the spectrum. Because of the isotopic enrichment with $^{18}\text{O}$, each peptide appears as a
doublet of equal intensity (m/z m and m+2) (Figure 5.3A). Manual inspection of the spectrum revealed a total of >130 unique doublet peaks, representing >90% of all possible signals. The observed difference in peak intensity was likely due to different ionization efficiencies, the presence of isobaric peptides, and the uneven amplitude of the excitation pulse. A few singlet peaks of relatively low intensities were also visible in the spectrum. These were caused by background hydrolysis of the pY side chain during library construction and screening. Fortunately, the extent of background hydrolysis was small as compared to PTP-mediated hydrolysis, so that the screening was not adversely affected.

5.3.3 Identification of Optimal Substrates of SHP-1

The peptide library was completely dissolved in an ammonium carbonate buffer and treated with SHP-1(ΔSH2) for various length of time (3.5–40 min). Since all of the peptides in the library were present at the approximately same concentration, the most preferred substrates should react first. Further, the enzymatic reaction was carried out under very dilute conditions (the concentration of each peptide was ~0.3 μM); therefore, SHP-1 was likely operating under the k_cat/K_M conditions for most of the substrates (vide infra). Dephosphorylation by SHP-1 removed the isotopic label, resulting in products as singlet peaks in the mass spectrum, which can be readily identified and sequenced by tandem mass spectrometry.

MS analysis of the SHP-1-treated library revealed a total of 23 singlets of significant intensities at various m/z values (Table 5.1 and Figure 5.3B–D). Most of these peaks were already clearly visible after 3.5 min of PTP treatment. In contrast, a few weak
singlets appeared only after 40 min of reaction; these peaks were not further studied. The 23 singlets were verified as products by the observed loss/reduction of the corresponding substrate peaks at m/z m+80 and m+82. Their possible amino acid compositions were inferred from the observed accurate molecular masses. Their actual sequences were then determined by MS/MS analysis on an ESI Q-TOF instrument. For example, a strong singlet peak of m/z 878.458 emerged in the enzyme treated library (Figure 5.4A and Table 5.1, entry 12). The molecular mass of this product is consistent with peptides containing sequence EZ (Z = leucine, isoleucine, or norleucine) or ZE at the randomized positions. MS/MS experiment indicated that the major product had the sequence RNNEZYA-NH₂ (Figure 5.4B). The intensities of the fragment ions corresponding to sequence RNNZEYA-NH₂ were much weaker and not significantly above the noise level. Unfortunately, due to mass degeneracy, it was not possible to determine whether all three sequences (EL, EI, and EM) were present or in what proportions they were present in the product. Note that due to the presence of an arginine at the N-termini of the peptides, the b fragment ions dominated the CID spectrum, significantly simplifying the spectral interpretation. All together, good MS/MS spectra were obtained for 16 product peaks. These spectra suggested that the major products all had either aspartic or glutamic acid at the –2 position (Table 5.1). For the DD peptide, no MS/MS experiment was necessary. For the rest of peptides (EP/PE, ET/TE, EQ/QE, SF/FS, DY/YD and FZ orZF), weak product signals and/or the presence of other intense peaks of similar m/z values prevented unambiguous sequence assignment. Based on the consensus of the sequenced products, we assume that the former products were EP, ET, and EQ, respectively. Therefore, the consensus sequence for the catalytic domain of SHP-1 is (D/E)XpY (X = any amino acid
except for Arg, Lys, His, or Gly). This is consistent with the previous observation that PTPs generally prefer acidic over basic residues at positions N-terminal to pY [123-128].

5.3.4 Identification of Poor Substrates of SHP-1

The disfavored substrates were identified by treating the pY library with excess enzyme for extended periods of time and analyzing the remaining doublet peaks in the mass spectrum (data not shown). Fifty-two sequences (27 doublet peaks) showed no detectable product formation after overnight incubation and therefore represent the least reactive substrates of SHP-1 (class I in Table 5.2). Each of these sequences contains at least one basic residue (arginine or lysine) at the randomized positions. Another 35 sequences (20 doublet peaks) showed only very small product peaks after overnight incubation with SHP-1. These sequences were also considered as poor substrates of SHP-1 (class II in Table 5.2). Again, these sequences turn to have basic (e.g., Lys, Arg, and His) or polar, neutral residues (e.g., Gln, Asn, Thr) at one of the randomized positions. Thus, SHP-1 strongly disfavors basic residues at the –1 and –2 positions.

5.3.5 Rank Order of Selected SHP-1 Substrates

The analysis described above revealed that SHP-1 generally prefers peptides of the consensus sequence (D/E)XpY. However, it does not provide further information about the relative reactivity of the 23 selected sequences. This is due to the fact that the intensity of a peak in an MS spectrum depends on many factors including sample concentration, ionization efficiency of the molecule, and the electrospray conditions. Therefore, a product peak of the highest intensity does not necessarily indicate that the
corresponding substrate is the best substrate. In order to rank order the 23 sequences, the
time course for their formation was monitored. Thus, the pY peptide library was treated
with SHP-1 for varying length of time (3.5, 10, and 40 min) before it was analyzed by
FTICR-MS. Comparison of the spectra from different reaction times showed that for
most of the singlet peaks, their intensity increased with time (Figure 5.3A–D). Next, the
intensity of each singlet peak was normalized relative to an internal standard and plotted
against time to generate its reaction progress curve. In this work, peptide
RNNZR(RZ)pYA-NH₂ (m/z 985.473) was used as the internal standard since it had no
detectable reaction during the experimental period. Since all of the peptides were present
in the initial library at approximately the same concentration, the shape of the reaction
progress curve provides an indicator of the potency of that substrate. For example, the
intensity of peptide RNNDEYA-NH₂ (m/z 880.402) approached the plateau value at 10
min, whereas the intensity of peptides RNNEZYA-NH₂ (m/z 878.458) and
RNNEN(DQ)YA-NH₂ (m/z 879.416) further increased substantially after 10 min (Figure
5.4 A, C). This indicates that at 10 min, peptide RNNDEpYA-NH₂ was almost
completely reacted, whereas RNNEZpYA-NH₂ and RNNEN(DQ)pYA-NH₂ were not,
suggesting the former as the better substrate. Note that even after 40 min, no significant
product signal was observed for sequences QN/NQ (m/z 878.423), KN/NK (m/z
878.460), GW/WG (m/z 879.422), DK/KD (m/z 879.444), SR/RS (m/z 879.455), or
FP/PF (m/z 880.443) in the expanded region (Figure 5.4A), indicating that these peptides
are very poor substrates of SHP-1.

The reaction progress curves of all 23 substrates are illustrated in Figure 5.5. It is
apparent from the strong curvature of their reaction progress curves that peptides
containing acidic residues at both –1 and –2 positions (e.g., DD, DE, and EE) are the most preferred substrates of SHP-1 (class 1). Furthermore, DD and DE appear to be slightly more reactive than EE. The next group of peptides (class 2) contain mostly aspartate at the –2 position and hydrophobic residues at the –1 position (Z, Y, F, W, V, A) (Table 5.1). Their reaction progress curves exhibited some curvature, indicating that a significant fraction of these substrates was reacted during screening. The majority of class 3 peptides contained an E at the –2 position and various amino acids at the –1 position. Their reaction progress curves are essentially linear over the 40-min period. Overall, it appears that SHP-1 has slight preference for an aspartate over glutamate at the –2 position.

5.3.6 Kinetic Analysis of Selected pY Peptides

Five of the selected peptides (with sequences DE, DD, EE, DF, and EF) were individually synthesized and tested against SHP-1(△SH2). These peptides showed \( k_{cat}/K_M \) values in the range of \( 1.6\times10^4 \) M\(^{-1}\) s\(^{-1} \) (Table 5.3). These are among the most efficient peptide substrates known for SHP-1 [123, 131]. The kinetic data are consistent with the ranking order predicted by the time-course experiments. For example, the data confirmed our prediction that SHP-1 slightly prefers an aspartate to glutamate at the –2 position (i.e., DE>EE, DF>EF). The MS/MS spectrum for peptide DF/FD showed that the predominant species was DF (fragment ion ratio was 6:1 for DF/FD) (data not shown). We thus synthesized the FD sequence (RNNFDpYA-NH\(_2\)) and found that the DF sequence was indeed 2-fold more reactive than the FD sequence (Table 5.3). Peptide RNNTQpYA-NH\(_2\), which showed a very weak product peak at m/z 865.422 after 40 min,
had a $k_{cat}/K_M$ value of $8670 \text{ M}^{-1} \text{s}^{-1}$. Two of the predicted poor substrates, RNNGKpYA-NH$_2$ and RNNKKpYA-NH$_2$, had $k_{cat}/K_M$ values of $640$ and $460 \text{ M}^{-1} \text{s}^{-1}$, respectively. Therefore, sequence variation at the –1 and –2 positions results in at least 100-fold difference in activity. To gain further insight into the cause of this difference, we synthesized an all-alanine peptide, RNNAApYA-NH$_2$, which had a $k_{cat}/K_M$ value of $2020 \text{ M}^{-1} \text{s}^{-1}$. These data suggest that a good sequence combination promotes the formation of productive $E\cdot S$ complex and catalysis, whereas a poor sequence interferes with binding to the enzyme active site and slows down the reaction.

5.4 Discussion

PTPs were initially thought as promiscuous “housekeeping” enzymes that simply oppose the action of PTKs. Recent work indicates that PTPs comprise a large family whose members play active roles in a wide variety of cellular processes and that PTPs exhibit exquisite substrate specificity in vivo. There is now growing evidence that PTPs achieve their impressive selectivity through combinations of specific targeting strategies (e.g., SH2 domains) and intrinsic specificity of their catalytic domain [133]. Thus, information on the substrate sequence specificity (intrinsic specificity) of PTPs will help identify their physiological substrates and the cellular functions of these enzymes. In addition, the specificity data will facilitate the design of specific inhibitors against the PTPs. Such inhibitors would provide potential therapeutic agents as well as invaluable tools for deciphering the cellular processes in which PTPs are involved. Optimal substrates would also provide more effective kinetic assays for these enzymes. Therefore, in recent years, considerable efforts have been made by many laboratories to define the
substrate specificity of PTPs using both combinatorial approaches [33, 62, 63] and by assaying individual pY peptides [123-128]. However, as discussed above, the existing methods each suffer from certain drawbacks and consequently, there is only limited knowledge about the specificity of a few PTPs.

In this work, we have applied the ECLIPSE method, which was developed recently by our laboratories and discussed in Chapter 4, to determine the substrate specificity of SHP-1 at the –1 and –2 positions. Screening of a 361-member pY peptide library revealed that the catalytic domain of SHP-1 has a consensus sequence of (D/E)XpY (X = any amino acid other than Arg, Lys, His, or Gly), with the highest activity toward peptides containing aspartate or glutamate at both –1 and –2 positions. The selected peptides are among the most efficient substrates reported so far for SHP-1 [123-131]. Since SHP-1 can make contacts with 3–5 residues on either side of the pY residue [130], we expect that still more efficient substrates will be found by optimizing the sequences at other positions. The ECLIPSE method also identifies the least reactive substrates. For SHP-1, the poorest substrates generally have basic residues at one or both of these positions. Many PTPs have basic residues near their active sites; the positively charged peptides are expected to have unfavorable electrostatic interactions with the enzyme active/binding site. Among the 361 peptides analyzed, there is at least a 100-fold difference in reactivity (compare peptides DE vs KK in Table 5.3). This difference is likely to be much larger if the other important positions are optimized. Thus, SHP-1 has relatively strong sequence selectivity toward pY peptides. Interestingly, the substrate specificity of SHP-1 at the –1 and –2 positions is similar to that of a number of other PTPs, such as PTP1B [62, 63], LAR [33, 134], Yersinia PTP [135, 136], and PTP1 [136].

86
Like PTP1B, SHP-1 also demonstrates a remarkable degree of plasticity in accommodating both acidic and large hydrophobic residues at the −1 position [137]. The similar specificity at the −1 and −2 positions suggests that different PTPs perhaps recognize their in vivo substrates by engaging in specific interactions at other positions.

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

Perhaps one might ask whether isotopic labeling is necessary in this method (or
why not simply compare the starting and the enzyme-treated libraries and look for new peaks formed?). Benner and co-workers have indeed used FTICR-MS to screen a very small library (16 members) for the optimal substrate of glutathione-S-transferase, by directly comparing the mass spectra of enzyme-treated and untreated libraries and identifying reaction products by the appearance of new peaks in the enzyme-treated spectrum [73]. While simple to perform, their method is limited to small libraries in which none of the substrates are isobaric with the product(s). For larger libraries, there is a much greater probability of product peaks overlapping with substrate peaks; separation into smaller sub-libraries would be necessary prior to MS analysis. However, we are not aware of any current separation technique that is capable of fractionating a complex combinatorial library into exactly the same fractions in two separate runs. This would make it difficult to directly compare the spectra of starting and enzyme-treated libraries.

Our method has no such limitation. Since ECLIPSE identifies a product based on its peak shape (singlet vs doublet), comparison with the starting library is unnecessary. If the isobaric peptides in the library can not be resolved in the mass spectrum, one can run the library through a chromatographic column to separate the isobaric peptides according to their retention time. These isobaric peptides can be detected separately by mass spectrometry to allow us to examine their isotopic ratio. The monoisotopic peptide and the peptide containing $^{18}$O would have virtually same retention time. Therefore, at any given time during the separation, the peptide from the starting peptide library should come out of the column and be detected as a doublet (M peak and M+2.0043Da), as shown in Figure 5.6. The peptides coming out column between 15 min and 16 min were detected as the doublet peaks, pairing with their M+2 peak.
ECLIPSE should be capable of screening much larger libraries by interfacing the mass spectrometer with various separation devices (e.g., HPLC). Unfortunately, FTICR-MS is not technically ideal for on-line coupling to HPLC because of the slow acquisition of data. Nevertheless, coupled with on-line HPLC separation, ECLIPSE could enable us to study the PTPs specificity with Q-tof, which has much lower resolution than FTICR-MS.

A product peak completely overlapped with a substrate peak can still be identified by the uneven intensities of the doublet. For example, in our previous work of screening a peptide deformylase library, products MDR and MER shared the same molecular formulas with substrates formyl-MSR and formyl-MTR, respectively [138]. The two products were still unambiguously identified by ECLIPSE. It is also worthwhile to point out that the unique feature for product identification should also greatly facilitate the automation of this screening process. For instance, one can develop computer algorithm that read raw data of peak list from the mass spectrum and output the singlet peaks by examining the ratio of M peak and M+2 peak [139].

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

In conclusion, we have demonstrated ECLIPSE as an effective method to systematically evaluate the substrate specificity of PTPs. The results show that the catalytic domain of SHP-1 is rather selective for peptides of the consensus sequence (D/E)XpY. This method should be widely applicable to a variety of other enzymes. Its application to other enzymes and more diverse libraries to assess the specificity of SHP-1 at other subsites is already under way in our laboratories.
Figure 5.1 Scheme showing library screening by ECLIPSE. Each member of the starting library appears as a pair of doublets at $m/z$ m and m+2, whereas a product appears as a singlet at $m/z$ m-80 in the mass spectrum.
Figure 5.2 The design of the pY peptide library for the identification of optimal enzyme substrates of protein tyrosine phosphatases by mass spectrometry-based screening.
Figure 5.3 ESI-FTICR mass spectra (positive ion mode) of the 361-member pY peptide library before enzymatic treatment (A) and after treatment with SHP-1 (8.0 µg) for 3.5 (B), 10 (C), and 40 min (D). Peaks numbered 1–23 (according to the numbering in Table 1) indicate the most prominent product peaks identified.
Figure 5.4 (A) Expanded spectra for the m/z 878–881 region in Figure 5.3, showing the appearance of product peaks with increasing SHP-1 reaction time. a, peptide RNNDZYA-NH₂; b, peptide RNNDQ(EN)YA-NH₂; c, peptide RNNDEYA-NH₂. Peak a+1 is due to the existence of natural isotopes in a (e.g., ¹³C). (B) A Q-TOF MS/MS spectrum of the product peak a (m/z 878.4). (C) Plot of peak intensities for products a–c in part A as a function of time (reaction progress curves). Z, Leu, Ile, or Nle.
Figure 5.5 Reaction progress curves for the 23 most abundant SHP-1 products.
Figure 5.6 On-line HPLC-MS analysis of starting peptide Library.
<table>
<thead>
<tr>
<th>Entry No.</th>
<th>sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Observed m/z</th>
<th>MS/MS&lt;sup&gt;b&lt;/sup&gt;</th>
<th>rank&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DA</td>
<td>822.390</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>EA</td>
<td>836.408</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>DP</td>
<td>848.408</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>DV</td>
<td>850.425</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>DT</td>
<td>852.403</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>EP</td>
<td>862.424</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>DZ/EV</td>
<td>864.442</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>DN</td>
<td>865.397</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>DD</td>
<td>866.381</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>ET</td>
<td>866.419</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>SF/FS</td>
<td>870.428</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>EZ</td>
<td>878.458</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>EN/DQ</td>
<td>879.416</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>DE</td>
<td>880.402</td>
<td>+</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>EQ</td>
<td>893.428</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>EE</td>
<td>894.412</td>
<td>+</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>a</sup>Z, leucine, isoleucine, or norleucine.  
<sup>b</sup>+, MS/MS performed; -, no MS/MS performed.  
<sup>c</sup>1-3, classes 1-3 peptides.

Table 5.1 Most Abundant Products from the SHP-1-Treated Library
<table>
<thead>
<tr>
<th>Entry No.</th>
<th>sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Observed m/z</th>
<th>MS/MS&lt;sup&gt;b&lt;/sup&gt;</th>
<th>rank&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>FZ/ZF</td>
<td>896.481</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>DF</td>
<td>898.424</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>EF</td>
<td>912.440</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>DY</td>
<td>914.420</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>EY</td>
<td>928.439</td>
<td>+</td>
<td>3</td>
</tr>
<tr>
<td>22</td>
<td>DW</td>
<td>937.435</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>EW</td>
<td>951.453</td>
<td>+</td>
<td>3</td>
</tr>
</tbody>
</table>

<sup>a</sup>Z, leucine, isoleucine, or norleucine.  
<sup>b</sup>+, MS/MS performed; -, no MS/MS performed.  
<sup>c</sup>1-3, classes 1-3 peptides.

Table 5.1 Most Abundant Products from the SHP-1-Treated Library (continued from last page)
<table>
<thead>
<tr>
<th>Class I&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Class II&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>RG/GR</td>
<td>KG/GK</td>
</tr>
<tr>
<td>RA/AR</td>
<td>KS/SK</td>
</tr>
<tr>
<td>RS/SR</td>
<td>KV/VK</td>
</tr>
<tr>
<td>RV/VR</td>
<td>KT/TK</td>
</tr>
<tr>
<td>RT/TR</td>
<td>KN/NK</td>
</tr>
<tr>
<td>RZ/ZR</td>
<td>KD/DK</td>
</tr>
<tr>
<td>RN/NR</td>
<td>KK</td>
</tr>
<tr>
<td>RD/DR</td>
<td>KE/EK</td>
</tr>
<tr>
<td>RQ/QR</td>
<td>KH/HK</td>
</tr>
<tr>
<td>RK/KR</td>
<td>KF/FK</td>
</tr>
<tr>
<td>RE/ER</td>
<td>KY/YK</td>
</tr>
<tr>
<td>RH/HR</td>
<td>KW/WK</td>
</tr>
<tr>
<td>RF/FR</td>
<td>RW/WR</td>
</tr>
<tr>
<td>RR</td>
<td></td>
</tr>
</tbody>
</table>

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

Table 5.2. Poor Substrates of SHP-1
<table>
<thead>
<tr>
<th>peptide</th>
<th>class</th>
<th>(k_{cat}/K_M) ((\times 10^3 \text{ M}^{-1}\text{s}^{-1})^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNNDEpYA-NH2</td>
<td>1</td>
<td>42.7 ± 3.4</td>
</tr>
<tr>
<td>RNNDDpYA-NH2</td>
<td>1</td>
<td>38.1 ± 3.3</td>
</tr>
<tr>
<td>RNNEEpYA-NH2</td>
<td>1</td>
<td>22.4 ± 1.1</td>
</tr>
<tr>
<td>RNNDFpYA-NH2</td>
<td>2</td>
<td>21.9 ± 1.2</td>
</tr>
<tr>
<td>RNNFDpYA-NH2</td>
<td></td>
<td>11.3 ± 1.6</td>
</tr>
<tr>
<td>RNNEFpYA-NH2</td>
<td>3</td>
<td>16.5 ± 0.9</td>
</tr>
<tr>
<td>RNNTQpYA-NH2</td>
<td></td>
<td>8.67 ± 0.48</td>
</tr>
<tr>
<td>RNNAApYA-NH2</td>
<td></td>
<td>2.02 ± 0.08</td>
</tr>
<tr>
<td>RNNKKpYA-NH2</td>
<td>I (poor)</td>
<td>0.46 ± 0.09</td>
</tr>
<tr>
<td>RNNGKpYA-NH2</td>
<td>I (poor)</td>
<td>0.64 ± 0.07</td>
</tr>
</tbody>
</table>

\(^a\)Data reported are the mean ± SD from at least three independent sets of experiments.

Table 5.3. Kinetic Properties of pY Peptides
APPENDIX A

This part of the work was independently conducted by Dr. In-Ja L. Byeon.

2.2.2 NMR experiments

NMR experiments were performed on a Bruker DMX-600 or DRX-800 spectrometer at 20 °C. The protein concentration was 0.5 mM. The samples contained 10 mM sodium phosphate, 1 mM DTT, and 1 mM EDTA in 95% H₂O/5% ᵃH₂O or 100% ᵃH₂O at pH 6.5. Intermolecular (FHA2-Rad9 pTyr peptide) NOEs were identified from 3D ¹³C-edited (f₂), ¹³C/¹⁵N-filtered (f₃) NOESY and the intramolecular NOE assignments of the pTyr peptide were obtained using 2D ¹³C/¹⁵N filtered NOESY using the complex sample ([¹³C/¹⁵N-FHA2]:[pTyr peptide]~1:2). The intramolecular NOEs for the FHA2 domain in the complexed form were identified using 3D ¹⁵N-edited NOESY and ¹³C-edited NOESY using the complexed sample. A mixing time of 100 ms was used for all the NOESY data. The NOESY data were processed and analyzed using XWIN-NMR 2.6 (Bruker) or Felix 95.0 (Molecular Simulations Inc.). Structural calculations were conducted using a simulated annealing method within X-PLOR using the distance restraint derived from the identified NOEs and H-bonds as well as the backbone torsion angle restraints derived from the TALOS program. For the complex structure calculation, three additional distance restraints (4.3(±0.3) Å for the C⃗ of Arg605, Arg617, and Arg620 to the P atom
of pTyr) were included, since the NMR titration data and the complex structural calculation without these three restraints (clearly indicated that these three Arg side-chains are involved in binding (approximately 4-5 Å from the Arg N\textsuperscript{η} atom to the phosphate group O atom). The structures were analyzed by X-PLOR, PROCHECK, and MOLMOL. For both the free and the pTyr peptide-complexed FHA2 structures, the final 19 structures with lowest energy were selected from a total of 60 calculated ones. Peptide-binding experiments were performed by recording a series of 2D \(^{15}\)N-HSQC spectra on uniformly \(^{15}\)N-labeled protein samples with different concentrations of peptides. The pH of the samples was maintained at pH 6.5 throughout the titration.
APPENDIX B

DETAILS ON THE SYNTHESIS OF THE $^{18}$O-LABLED PY AMINO ACID

This part of the work was independently conducted by Dr. Hua Fu (Figure B.1).

5.2.2 Synthesis of t-Butyl N$^\alpha$-[(Fluoren-9-yl)methoxycarbonyl]-L-tyrosinate (1).

Fluorenlylmethyl chloroformate (5.45 g, 21.0 mmol) was slowly added to a stirred solution of L-tyrosine t-butyl ester (5.0 g, 21.0 mmol) in 70 mL of dioxane and 60 mL of 10 % Na$_2$CO$_3$ in an ice/water bath. The mixture was stirred for 4 h in the ice/water bath and then 12 h at room temperature. The mixture was extracted with ethyl acetate ($2 \times 200$ mL), and the organic phase was combined and evaporated under reduced pressure. The crude product was purified by silica gel chromatography (1:1 ether/hexane) to obtain a white solid (8.2 g, 85% yield). $^1$H NMR (250 MHz, CDCl$_3$): $\delta$ 7.27-7.78 (m, 8H, Ar of Fmoc), 7.01 (d, 2H, Ar of Tyr), 6.73 (d, 2H, Ar of Tyr), 5.27 (d, 1H, NH of Tyr), 4.14-4.62 (m, 4H, CH$_2$ and CH of Fmoc, CH of Tyr), 3.70 (s, 1H, OH of Tyr), 3.00-3.03 (m, 2H, CH$_2$ of Tyr), 1.43 (s, 9H, t-butyl).
5.2.3 Synthesis of $^{18}$O-Labeled N$^\alpha$-[(Fluoren-9-yl)methoxycarbonyl]-O-[bis(allyloxy)phosphoryl]-L-tyrosine (4).

1H-Tetrazole (0.44 g, 6.0 mmol) was added to a solution of bis(allyloxy)diisopropaminophosphine (1.68 g, 6.0 mmol) and t-butyl N-Fmoc-L-tyrosinate (1) (1.84 g, 4.0 mmol) in anhydrous THF (15 mL) and the resulting solution was stirred under Ar atmosphere for 45 min at room temperature. A solution of iodine (1.52 g, 6.0 mmol) and 0.57 g (30 mmol) water ($^{16}$O/$^{18}$O = 1:1) in THF (5 mL) was added at –58 °C. After 2 h an aqueous solution of 10% Na$_2$S$_2$O$_5$ (15 mL) was added to the mixture. The solution was stirred for 10 min at room temperature, transferred to a separatory funnel, and extracted with ethyl acetate (40 mL). The organic phase was washed with 5% NaHCO$_3$ (2×30 mL) and dried over Na$_2$SO$_4$. After silica gel chromatography (2:1 hexane/ethyl acetate), 1.78 g of a white solid (compound 3) was obtained (72% yield). $^1$H NMR (250 MHz, CD$_3$COCD$_3$): δ 7.16-7.87 (m, 12H, Ar of Fmoc and Tyr), 5.94-5.98 (m, 2H, OCH$_2$CH=CH$_2$), 5.20-5.41 (m, 4H, OCH$_2$CH=CH$_2$), 4.54-4.66 (m, 5H, OCH$_2$CH=CH$_2$, NH of Tyr), 4.16-4.43 (m, 4H, CH$_2$ and CH of Fmoc, CH of Tyr), 2.91-3.12 (m, 2H, CH$_2$ of Tyr), 1.41 (s, 9H, t-butyl); $^{31}$P NMR (CD$_3$COCD$_3$): δ -4.98.

Compound 3 (1.78 g, 2.87 mmol) was treated with 8 mL of trifluoroacetic acid overnight at room temperature. After the removal of trifluoroacetic acid by rotary evaporation, the residue was dissolved in 20 mL of diethyl ether. The solution was extracted with 5% NaHCO$_3$ solution (3 × 10 mL). The combined aqueous phase was washed with diethyl ether (3 × 10 mL), acidified to pH 2 by the addition of 30% HCl, and extracted with CH$_2$Cl$_2$ (3 × 15 mL). The organic phase was dried over Na$_2$SO$_4$ and
the solvent was evaporated under reduced pressure to afford 1.6 g of compound 4 (99% yield). $^1$H NMR (250 MHz, CDCl$_3$): $\delta$ 10.51 (br. 1H, COOH), 7.08-7.74 (m, 12H, Ar of Fmoc and Tyr), 5.84-5.95 (m, 2H, OCH$_2$CH=CH$_2$), 5.21-5.36 (m, 4H, OCH$_2$CH=CH$_2$), 4.43-4.63 (m, 5H, OCH$_2$CH=CH$_2$ and NH), 4.16-4.43 (m, 4H, CH$_2$ and CH of Fmoc, CH of Tyr), 2.81-3.18 (m, 2H, CH$_2$ of Tyr); $^{13}$C NMR (62.5MHz, CDCl$_3$): 199.42, 173.84, 156.29, 149.87, 149.76, 144.26, 144.11, 141.71, 133.71, 132.30, 132.19, 131.27, 128.13, 127.49, 125.49, 120.47, 120.38, 119.33, 69.59, 69.51, 67.39, 54.98, 53.87, 47.55, 37.43. $^{31}$P NMR (CDCl$_3$): $\delta$ -6.20; ESI-MS: [M+Na]$^+$ m/z 586.588 (1:1).
Figure B.1 Synthesis of partially $^{18}$O-labeled pY building block 4.
APPENDIX C

SCREENING A PEPTIDE INHIBITOR LIBRARY FOR PTP1B INHIBITOR

C.1 Introduction

PTP1B, a best characterized member of protein tyrosine phosphatase super family [140, 141], had been suggested as a drug target for treatment of type II diabetes [142]. The development of specific and potent PTP1B inhibitor has become the intense pursuit of both academic and industrial labs. Many PTP1B inhibitors have been discovered over the past 10 years [143]. Recently, Fu et al. in our lab have discovered a new class of PTP1B inhibitor with aldehyde functional group [144]. Among the aldehydes tested, 4-Carboxycinnamaldehyde (I, Figure C.1) was found to be the most potent inhibitor not only of PTP1B (IC₅₀=970 µM) but also of other protein tyrosine phosphatases such as SHP-1 [131] and VHR [145]. It was found by Fu et al. that the formation of an enamine between the aldehyde group of the inhibitor and the guanidine group of Arg-221 in the PTP1B active site provides the basis for the reversible inhibition. Furthermore, the attachment of the tripeptide Gly-Glu-Glu to the para position of cinnamaldehyde (Cinn-GEE, 2, Figure C.1) substantially increased the potency against PTP1B to Kᵢ of 5.4 µM.
This suggests that the peptide portion of the inhibitor makes specific interactions to the amino acid residues around the active site, which improves the binding affinity to the enzyme. Cinnamaldehyde provides us with a general core structure from which more potent and specific PTP1B inhibitors may be derived. To optimize the affinity contribution from the peptide portion of inhibitor, we screened a peptide library with cinnamaldehyde group attached to the N terminal of the peptides.

C.2 Experimental Procedures

C.2.1 Materials

All materials for peptide synthesis including resins, protected amino acids, and coupling reagents were purchased from Advanced ChemTech (Louisville, KY, USA). Other chemicals were from Aldrich (Milwaukee, WI, USA). GF-5 desalting gel from Pierce (Rockford, IL) was used for size exclusion chromatography. The exclusion limit is 5000 Daltons and particle size ranges from 40 to 100 microns. Plastic column (10 mL) from BioRad (Hercules, CA) was used to pack the size exclusion column. The peptide microtrap from Michrom BioResources (Auburn, CA) was used to concentrate the peptides after size exclusion chromatography. The catalytic domain of PTP1B (residues 1-321) was expressed in *E. coli* and purified to near-homogeneity according to literature procedures [146]. The Rink resin had a loading capacity of 0.7 mmol/g. Ethyl 4-[2-(1,3-Dioxolane-2-yl)-ethenyl]benzoyl Carbonate (3, Figure C.1) was provided by Hua Fu and Junguk Park. Buffer A contains 50 mM HEPES, 50 mM NaCl, 1mM EDTA and 5mM tris(carboxyethyl)phosphine (TCEP).
C.2.2 Peptide Library Synthesis

A pentapeptide library, NH2-X1X2X3RS-NH2, was constructed with 3 different amino acids (Gly, β-Ala and Pro) at the X1 randomized positions, 6 different amino acids (Thr, Leu, Asn, Lys, Glu, and Phe) at the X2 randomized positions, and 6 different amino acids (Val, Thr, Asp, Gln, Arg and Tyr) at the X3 randomized positions. Library synthesis was performed on 1.0 g of Rink resin (0.7 mmol/g) using standard Fmoc/HBTU/HOBt chemistry. The randomized positions were generated by the split-pool synthesis method [2-4]. The resin was divided into equal portions and placed into the 5 mL reacting vessels. A different Fmoc-amino acid (5-fold excess) was placed into each vessel and the coupling reaction was allowed to proceed for 3 h at room temperature. The coupling reaction was repeated once to ensure completion reaction. After removal of the N-Fmoc group, the second random position was generated in the same manner. Finally, ~130 mg (0.4 mmol) of anhydride 1 in 5 mL of CH2Cl2 and 200 µl of N-methylmorpholine were added to the resin (1/6 of the total resin, ~0.13m mol) suspended in 5 mL of anhydrous DMF. The mixture was shaken for 8 h at room temperature. Ninhydrin tests indicated complete acylation of the N-terminal amine. The solvents were drained, and the resin was washed with DMF (5 × 5 mL) and CH3OH (3 × 5 mL). Deprotection of side-chain as well as aldehyde protecting groups and cleavage from the resin were carried out with a cocktail containing 7 mL of 90% TFA in water, 0.1 mL of anisole, and 0.15 mL of thioanisole for 3 h at room temperature. TFA, H2O, and other volatile chemicals were removed under a gentle flow of nitrogen, and the residue was triturated several times with ~ 40 mL diethyl ether. Peptide inhibitor library was obtained as a brownish solid (yield ~100 mg).
C.2.3 Size Exclusion Isolation Of Inhibitor-Protein Complex

Typically, 10 mg of the peptide inhibitor library was dissolved in 100 µl of DMSO to make a stock solution (~125 mM). 2 µl of stock was diluted in 400 µl of buffer A, followed by the addition of the 20 µl of PTP1B (250 µM). The mixture was incubated in room temperature for about 3 hours before proceeding to the next step. Two control experiments were conducted at the same time. In one control (control 1), 2 µl of stock was diluted in 420 µl of buffer A without adding PTP1B whereas in another control (control 2), 20 µl of PTP1B (250 µM) was diluted in 400 µl of buffer A.

Three size exclusion columns were packed with GF-5 desalting gel, each having ~7 mL bed volume. The columns were washed extensively with double distilled water and 0.8 mL of buffer A before use. The inhibitor-protein mixture as well as the two control solutions was loaded to the columns respectively. The columns then were centrifuged at rpm of 2000. The solutions that flew through the column were collected. The inhibitor-protein mixture and protein solution (control 2) was assayed with Bradford method to test the presence of protein. The absorbance of the inhibitor solution (control 1) was monitored at 214 nm to test the presence of peptide inhibitors.

1 µl of TFA was added to each of three solutions to denature the protein before peptide trap was used to concentrate the peptide inhibitors that went through the size exclusion column by binding to PTP1B. Typically, the peptide trap was washed with 300 µl of 50 % Acetonitrile and equilibrated with 0.1% TFA water solution before 100 µl of the solution was loaded onto the trap. After washing the trap with 300 µl of 0.1% TFA, the
peptide inhibitors were eluted with 50 μl of 95% Acetonitrile. The resulting solutions were analyzed by mass spectrometry.

C.2.4 Mass Spectrometry Analysis

The concentrated samples were first analyzed with Micromass Q-TOF2 mass spectrometer in the positive ion mode (Manchester, UK). Typically, the sample was diluted twofold with 50% methanol before it was introduced into spectrometer by electrospray at a rate of ~1 μL/min. The peptide inhibitors of interest were fragmented through collision-induced dissociation (CID) with argon. The fragment ions were typically accumulated over a period of 5 minutes to yield the desired CID spectra.

Alternatively, a 1 μl aliquot of the concentrated solution was mixed with 2 μl of saturated solution of 2, 5-dihydroxybenzoic acid in water and 1 μl of the resulting mixture was applied to the spectrometer plate. MALDI mass analysis was performed on a Bruker Reflex III MALDI-TOF mass spectrometer in the positive ion mode.

C.2.5 Liquid Chromatography-Mass Spectrometry (LC-MS)

The solution without concentrating was directly used to conduct Liquid Chromatography-Mass Spectrometry (LC-MS) experiment. A 50 μm diameter capillary C18 column from New Objective (Woburn, MA) was coupled to Micromass Q-TOF2 mass spectrometer. Typically, 10 μl of the solution was injected into the column. The MS/MS experiments on selected peptide ions were conducted automatically using the software control.
C.2.6 Peptide Synthesis And PTP Inhibition Assays

Five peptide inhibitors were synthesized by the same method used in the library synthesis. Stock solutions of crude peptide inhibitors were prepared in dimethyl sulfoxide (DMSO), and their concentrations were calculated from the known inhibitor masses and solvent volumes. A typical reaction (total volume 1 mL) contained 50 mM HEPES (pH 7.4), 1 mM EDTA, 1 mM tris(carboxyethyl)phosphine (TCEP), 50 mM NaCl, 5% (v/v) DMSO, 0.1-0.2 µM PTP1B, and 0-50 µM inhibitor. After incubation of the enzyme with the inhibitor overnight at room temperature, the reaction was initiated by the addition of 1.0 mM \( p \)-nitrophenyl phosphate (\( p \)NPP). The reaction progress was monitored at 405 nm on a Perkin-Ellmer UV-Vis spectrophotometer.

C.3 Results And Discussion

C.3.1 Design And Synthesis Of The Peptide Inhibitor Library

Fu et al. have discovered that attachment of the tripeptide Gly-Glu-Glu to the para position of cinnamaldehyde resulted in an inhibitor (Cinn-GEE) of substantially increased potency against PTP1B (\( K_i = 5.4 \) µM). This let us believe that with certain peptide sequence attached to cinnamaldehyde, we could obtain even more potent inhibitor. With this in mind, we have designed a combinatorial peptide library that contains 108 members covering a range of characteristic amino acids at these three positions, Cinn-\( X_1X_2X_3 \). The amino acids were chosen in such a way that the mass degeneracy of the library members was minimized. Consequently, with the determination of molecular weight by electrospray mass spectrometry (ESI-MS), we can infer the peptide sequence. To
improve the solubility of library, arginine and serine were added to the C terminal of the peptide, resulting in the final form of the library as Cinn-X₁X₂X₃SR-NH₂. The library should produce 96 distinct molecular peaks in the m/z range of 676-819 in the ESI-MS spectrum, most of which were indeed observed and well resolved in the spectrum (data not shown).

C.3.2 Size Exclusion Isolation Of Inhibitor-Protein Complex

Size exclusion has long been used to separate the macromolecules according to their size difference. Zuckermann et al. first utilized this approach to separate the antibody-peptide complex from the rest of a peptide library with subsequent analysis of the peptide ligands by High Performance Liquid Chromatography (HPLC) [147]. Dunayevskiy et al. coupled ESI-MS to the size exclusion isolation to identify the ligands released from the complex [64]. However, separating the protein-ligand complex would disturb the binding equilibrium, causing the dissociation of the complex. Therefore, the separation of the protein-ligand complex from the free ligands has to be fast enough to preserve the complex. By centrifugation, the protein-ligand complex and the protein go through the size exclusion column in seconds whereas the free ligands (small molecules) are trapped inside the column. The ligands that bind to the protein can be subsequently released from the complex by denaturing the protein, and subjected to further analysis for identification.

To prevent the free ligands from eluting through the size exclusion column, certain precautions ought to be taken. First, the bed volume of the column has to be large enough compared to the loading volume. A ratio of ~15:1 (bed volume vs. loading volume) was chosen in this experiment. Secondly, the concentration of the inhibitor solution should be
kept below certain level, in this case, ~0.3 mM. A series of control experiments with different conditions were conducted to find the best conditions. The presence of the peptide inhibitors in the eluting solution was tested by measuring the absorbance of the solution at 214 nm. The absorbance higher above the background would indicate the presence of the peptide inhibitors. However, the concentration of the inhibitor could not be kept too low to be detected.

On the other hand, it was found that the proteins were never 100% recovered after the elution by centrifugation. In fact, under certain conditions (large bed volume, low protein concentration), the protein recovery was extremely low, as indicated by Bradford assay. This could be due to the non-specific binding of the protein to the column. Without enough protein eluting from the column, there would not be enough peptide inhibitors released from their binding protein. Therefore, a proper bed volume and protein concentration ought to be chosen. The conditions described in the experimental procedures were reached through many trials.

C.3.3 Mass Spectrometry Analysis

Note that the amount of peptide inhibitors released from the binding PTP1B was very low but enough for mass spectrometric detection. With the binding stoichiometry of 1:1 ratio, the total amount of peptide inhibitor released from PTP1B was 10 nmol. If we assume ~10 peptide inhibitors significantly bind to PTP1B, each inhibitor would present at 1 nmol level. Keep in mind that this calculation is based on 100% protein recovery after centrifuging elution and no dissociation of protein-inhibitor complex, which is seldom the real circumstance. Therefore, the estimated peptide inhibitors of interest
would present at subnanomole range, which is not very difficult for mass spectrometric detection. However, the presence of large amount of protein and its degradation products (small peptides) made it rather problematic to detect the peptide inhibitors at low level.quantity. As shown in ESI-MS spectrum of the eluting solution (Figure C.2), in the mass range of peptide inhibitors (m/z 676-819), many of prominent peaks were multiple-charged peptide ions, which came from the degradation of the protein. The presence of these high level peptides greatly suppressed the signal of the peptide inhibitor, making the identification of the peptide inhibitor nearly impossible. To alleviate the problem, MALDI-MS was used to produce single-charged peptide ions. In the MALDI-MS spectrum (Figure C.3), the number of peaks shown in the range between m/z 676 and m/z 819 was significantly reduced as a result of eliminating multiple-charged ions in this range. A number of peaks appear at the m/z that corresponded to calculated molecular weight of some library members. Peaks at m/z of 678.33, 747.43, 765.36 and 779.41 corresponded to library member GTT, BTR/BEQ, BFQ, and GFR respectively. However, because MALDI-MS lacks the capability of conducting tandem mass spectrometry experiment, the identities of these peptides have not been further confirmed.

**C.3.4 Liquid Chromatography-Mass Spectrometry (LC-MS)**

Coupling on-line Liquid Chromatography to the ESI-MS allows us to generate simplified mass spectrum at a given time window, therefore eliminating the interference of the protein degradation product and making the identification peptide inhibitor much easier. In fact, the peptide inhibitors mostly eluted from the capillary column in the first 30 min whereas the larger peptides from the protein degradation eluted mostly after 30
min. In order to confirm the identities of the peptide inhibitors, tandem mass spectrometry experiments were performed automatically on a number of singly charged ions that correspond to the calculated molecular weight of the library member. Unfortunately, the quality of the CID spectra was not good enough to interpret the sequence information. It was due to the low level of the peptide inhibitors and the short time that was allowed to accumulate CID spectrum in an on-line separation. Nevertheless, the CID spectra for most of the peptide ions did suggest the presence of the cinnamaldehyde portion on the peptide (data not shown).

**C.3.5 PTP1B Inhibition Assays**

Five peptide inhibitors were synthesized individually. Their crude products were used directly for the inhibition assay to test their potency. Inhibitor cinn-GEESR-NH$_2$ was synthesized as a comparing standard. Among the five peptides tested (Table C.1), BTR inhibitor showed comparable inhibition potency as that of GEE standard whereas GTD and GFD were surprisingly found to activate the PTP1B activity at 20 µM concentration. GTY did not show any noticeable effect on the activity of PTP1B at 20 µM concentration.

**C.4 Conclusion**

The size exclusion isolation of protein-ligands complex from free small molecules has been demonstrated by a number of groups. However, they have not yet used this method to screen a chemical library with decent size. The difficulty of applying this method to large chemical library lies in the fact that the ligands that released from the binding
protein are in such a low level that the other contaminating species (protein degradation, column materials) could greatly interfere the identification of the ligands. This problem could be circumvented by selectively enriching the low level ligands and separate them from other interfering species. For example, when synthesizing the peptide inhibitor library, we can add a biotin affinity tag on every member in the library and use affinity avidin purification to selectively enrich and separate the library members from other species.

In this work, the identities of the peptide inhibitors detected in the mass spectra were mostly inferred through the matching of their measured molecular weight to the calculated molecular weight. Due to their low level quantity, CID spectra of these peptide inhibitors can not give clear peptide sequence information. Four peptides inhibitors were chosen from the library to test their inhibition potency. Only one of them, cinn-BTRSR, showed comparable inhibition potency to that of the existing inhibitor, cinn-GEESR.
Figure C.1 Structure of the three compounds 1, 2, 3.
Figure C.2 ESI-MS spectrum of the elution solution
Figure C.3 MALDI-MS spectra of the eluting solution (b) and control 1 (a).
Table C.1 The inhibition test of the peptide inhibitors chosen from the library.

<table>
<thead>
<tr>
<th>Peptide inhibitor</th>
<th>Percent of the PTP activity at 20 µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cinn-GTDSR</td>
<td>142%</td>
</tr>
<tr>
<td>Cinn-GFDSR</td>
<td>155%</td>
</tr>
<tr>
<td>Cinn-GTYSR</td>
<td>101%</td>
</tr>
<tr>
<td>Cinn-BTRSR</td>
<td>62%</td>
</tr>
<tr>
<td>Cinn-GEESR</td>
<td>76%</td>
</tr>
</tbody>
</table>
D.1 Research Goals

Microbial resistance to existing antibiotics has becoming a serious threat to human health. To counter the resistant pathogens, we need to develop novel antibiotics that attack new targets in the pathogens. Peptide deformylase (PDF), an essential enzyme involved in protein synthesis and maturation, is such a new target [120]. Several academic labs (including our own) and over a dozen of pharmaceutical companies are designing inhibitors against the bacterial PDF. Our lab has recently shown that PDF is also present in some eukaryotes such as Plasmodium falciparum, the parasite that causes malaria in much part of the third world, and that PDF inhibitors can stop the parasite growth [148]. Unfortunately, due to the perceived lack of market value, drug companies are reluctant to invest in the development of antimalarial drugs. One of our goals is to develop potent inhibitors against the malarial PDF (PfPDF) as potential antimalarial drugs. Another goal is to develop a novel methodology to rapidly screen synthetic compounds.
D.2 Inhibitor Design

PDF is a metalloenzyme, which contains an essential Fe$^{2+}$ ion in the active site. All of the reported inhibitors take advantage of this feature by binding directly to the metal ion. The N-formyl-hydroxylamine (FHA) moiety is a particularly attractive metal ligand because it closely mimics the substrate structure (Figure D.1a), binds to Fe$^{2+}$ with high affinity, and is stable in an animal body [149]. We choose also FHA as the metal ligand. The overall structure of the designed inhibitor mimics a preferred substrate, f-ML-pNA, with the FHA moiety mimicking the formylmethionyl group in the substrate (P$_1'$ residue) (Figure D.1b). We will use four different side chains (R$_1$) at this position. At the second amino acid position (P$_2'$ position), we will use amino acid glutamine to facilitate the synthetic operations. An earlier study from this lab has shown that PDF does not have strong selectivity at the P$_2'$ position and that a glutamine at this position gives good activity [22]. At the P3' position, we will use 30 different amines (R$_2$) to mimic the third amino acid in the substrate. The combination of 4 R$_1$ and 30 R$_2$ groups should produce 120 different compounds.

D.3 Combinatorial Synthesis of PfPDF Inhibitors

Synthesis and testing of 120 compounds individually would cost enormous manpower and resources. We will synthesize the compounds using a combinatorial approach, in which hundreds to millions of different compounds are generated simultaneously by combining a small set of building blocks. These compounds are then screened against the proper target to isolate the ones with the desired properties. Finally, the structures of the selected ones ("hits") are identified using various analytical tools. The main advantage of
combinatorial chemistry is that a large number of compounds can be rapidly synthesized in a short time, significantly shortening the time it takes to find a hit. Not surprisingly, since its inception in early 1990's, this technique has quickly become a major component of the pharmaceutical/biotech industry. The disadvantage of the approach is that, because each compound can only be synthesized at minute amount, screening and identifying the compounds is enormously challenging, a task no less than finding a needle in a haystack. Thus, a second goal of this project is to develop a generally effective method for screening and identifying compounds from combinatorial libraries.

The 120 compounds will be combinatorially synthesized using a well-established split-pool synthesis method [2-4]. The details of the synthesis will not be provided here due to the fact that the synthesis component is already under control and therefore is not the focus of my future work.

D.4 Library Screening and Hit Identification

The 120 compounds will be screened by mass spectrometry, a technique known for its extreme sensitivity. However, mass spectrometry per se cannot tell which of the 120 compounds has the highest affinity to PfPDF. To solve this problem, we will apply the partial isotopic labeling concept we have previously developed [138]. Specifically, during the last step of inhibitor synthesis (the addition of a formyl group), the compound library will be equally split into two portions. A formyl group (H-C=0) is added to the first portion, whereas a deuterated formyl group (D-C=0) is added to the other (Figure D.2). The Deuterated inhibitor library is passed through a column immobilized with PfPDF, whereas the other copy is passed through a blank column. The flow-through solutions
from the two columns are combined and analyzed in a mass spectrometer. If an inhibitor does not bind PfPDF at all, both copies of the inhibitor (HCO and DCO) should emerge from the columns. Because the two inhibitors differ only in one hydrogen isotope, they have essentially the same properties. In a mass spectrum, they will appear as a doublet of equal intensity and are separated by one mass unit (the mass difference between H and D). On the other hand, if an inhibitor binds to PfPDF, all or part of the deuterated inhibitor will be retained by the PfPDF column. In a mass spectrum, this inhibitor will produce either a singlet (if all of the deuterated inhibitor is retained by the PfPDF column) or a doublet of unequal intensities (if part of the inhibitor is retained by the column). Thus, one can use the intensity ratio of the two peaks \((m+1)/m\) to rank order the inhibitors according to the binding affinity; the smaller the ratio, the tighter the binding. Once the most potent inhibitor is identified, its structure is determined by isolating the peak and performing tandem mass spectrometric experiments, which are automated on our instruments. All of the compounds are designed to have unique masses. Finally, the identified potent inhibitor(s) are individually synthesized on a large scale and tested against PfPDF.

D.5 Expected Results

Given its simplicity and our experience with the isotopic labeling technique, we are confident that the new screening method will work as planned. This would provide a very useful technique for other practitioners of combinatorial chemistry. We are less certain whether a potent inhibitor will emerge from the designed library. However, we should at
least identify the more potent ones from the 120-member library. These hits will provide valuable clues for designing a second-generation library.

D.6 Synthesis Failure

Unfortunately, the majority of synthesized inhibitor library members do not have the right molecular weight as determined by ESI-MS. Therefore, this new screening method was not further pursued.
Figure D.1 (a) Structure of the inhibitor, (b) Core structure of the designed compounds. R1 and R2 group are depicted in next page respectively.
R1 =

R2 =

Figure D.1 (a) Structure of the inhibitor, (b) Core structure of the designed compounds
Figure D.2 The inhibitor library screening by ESI-MS
REFERENCES


31. Lebl, M.; Krcháč, V.; Salmon, S. E.; Lam, K. S. Methods (San Diego) 1994, 6, 381.


16, 505.


90. Liao, H.; Byeon, I. J.; Tsai, M. D. J. Mol. Biol. 1999, 294, 1041.


120. Pei, D. Emerging Therapeutic Target 2001, 5, 23.


139. Freitas, M. A. unpublished results.