Solution Structure and Electrostatic Properties of an SH2 Domain/Phosphopeptide Complex

by

Alex U. Singer

A thesis submitted in conformity with the requirements of the Degree of Doctor of Philosophy, Graduate Department of Medical Genetics and Microbiology, University of Toronto

© Copyright by Alex U. Singer 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
Solution Structure and Electrostatic Properties of an SH2 Domain/Phosphopeptide Complex -- by Alex U. Singer, for the Degree of Doctor of Philosophy, Department of Medical Genetics and Microbiology, University of Toronto, 1998

Abstract

SH2 domains are small (~100 amino acid) protein recognition domains found in numerous proteins involved in signal transduction which bind to sites of tyrosine phosphorylation with high affinity in a sequence-dependent manner. We have focused on the SH2 domains of phospholipase C-γ (PLC-γ), which provide a link between activated growth factor receptors via binding through its two SH2 domains and the production of the second messengers IP₃ and DAG. The interaction of PLC-γ with the platelet derived growth factor receptor (PDGFR) is at sequences about Tyr 1021 of the PDGFR, and disruption of this interaction results in decreased cell growth following growth factor stimulation. Binding studies using degenerate phosphopeptide libraries suggest that this interaction involves the C-terminal (PLCC), and not N-terminal (PLCN) SH2 domain of PLC-γ. Thus we have studied this interaction involving the PLCC SH2 domain and a 12 amino acid phosphopeptide representing sequences about Tyr 1021 using heteronuclear NMR techniques.

I was involved in the cloning and purification of this SH2 domain and preparation of NMR samples of this protein/peptide complex. A full structural determination was performed on this complex in collaboration with Dr. Steve Pascal. During structure determination, I defined the conformation of the phosphopeptide in this complex, as well as demonstrating protein-peptide contacts. Protein/peptide NOEs involving ρTyr resonances defined a large positively-charged pocket containing four arginine residues which bound this residue. NOEs could not define contacts with the pTyr phosphate group, and we used large downfield chemical shift changes of guanidinium group resonances to do so. pH
titration studies demonstrated that the pTyr phosphate group is bound in the -2 charge state with several residues held in place to facilitate pTyr binding by a complex hydrogen bonding network. A large hydrophobic cavity on the SH2 domain surface bound six residues C-terminal to pTyr, and in particular, the Ile +1 and Pro +3 residues were deeply buried. Thus a combination of NMR techniques involving NMR assignment, structure determination and pH titration studies provided significant insights into the specific binding of SH2 domains.
Acknowledgments

The work I describe in this thesis is part of a collaboration involving the labs of Drs. Tony Pawson, Julie D. Forman-Kay and Lewis E. Kay. The following is a clarification of my role in these studies. The studies described in Chapters 3 and 5 are exclusively my own. In Chapter 2, the cloning and purification of the PLCC SH2 domain was my work, though the assistance of Dr. Gerry Gish, a research associate in the Pawson lab, must be acknowledged. The remaining work which I describe in that chapter was performed by myself, with the exception of the backbone assignment, which also involved the work of Dr. Forman-Kay and, to a lesser extent, Dr. Gish. In Chapter 4, I was exclusively involved in assignment of protein-peptide NOEs, while Dr. Steve Pascal, a post-doctoral fellow in the labs of Drs. Forman-Kay and Kay, performed the bulk of the assignment and NOE determination for the SH2 domain in this protein-peptide complex. Structure calculations on this protein-peptide complex were run by both myself and Dr. Pascal. The work described in Chapter 6 with regard to chemical shifts of arginine resonances of the SH2 domain and SH2 domain/peptide complex was a collaboration between myself, Dr. Toshio Yamazaki, a former post-doctoral fellow in the lab of Dr. Kay, and Dr. Pascal. The arginine stereoassignment and determination of NOEs involving arginine guanidinium groups is strictly mine.

I would like to thank my supervisors, Dr. Julie Forman-Kay and Dr. Tony Pawson for their support (financial and otherwise), enthusiasm, encouragement and training they provided during the time in which I was in their laboratories. From them I learned a considerable amount both in terms of the fields of signal transduction and structural biology, but also about the discipline required to become a successful scientist.

All pulse sequences were provided for by Dr. Lewis Kay, with the exception of the Arg-specific NOESY sequence, which was written by Dr. Yamazaki. Many of the NMR experiments were run by Drs. Kay, Yamazaki, and Dr. Ranjith Muhandiram, or they
assisted in running them. Phosphopeptides were provided both by the laboratory of Dr. Steve Shoelson, at the Joslin Diabetes Centre in Boston, Mass., or by Dr. Gerry Gish. I would like to thank Dr. Niel Farrow for a Monte Carlo routine for error analysis for the pH titration data in Chapter 5.

During the time in the lab, I received much in terms of help, scientific discussions and friendship. Dr. Gerry Gish was invaluable throughout the first year or so, in which I was involved in molecular biology and biochemistry. Dr. Steve Pascal was a great help during the next phase of NMR work and structure determination. In addition, I would like to thank, in no particular order, Genevieve Gasmi, Ouwen Zhang, Voula Kanelis, Randy Willis, Henry Mok, Kevin Gardner, Catherine Zhaowlin for their valuable discussion and friendship during the course of my stay in these laboratories, as well as Vicki Lay, who has been a source of friendship and support throughout.

Finally I would like to thank my parents for their support and patience throughout my years of study and their encouragement to pursue a scientific career.
Table of Contents

Abstract i
Acknowledgments iii
Table of Contents v
List of Figures x
List of Tables xv
Abbreviations xvii

Chapter 1: Introduction
I. Overview 1
II. Signal Transduction Pathways 2
   A. Growth Factor Receptors 3
      1. General 3
      2. Structural Biology 4
   B. SH2 Domains 7
      1. General 7
      2. Structural Biology 15
   C. PTB Domains 16
   D. Phospholipase C 17
      1. Role in Signal Transduction 17
      2. Primary Structure 18
      3. Structural Biology of Phospholipase C Isozymes 24
         a. PLC-δ Catalytic Domain 24
         b. The PH Domain 28
         c. The SH3 Domain 31
   E. Interaction of Domains in Signaling Molecules 31
### III. Introduction to Nuclear Magnetic Resonance (NMR)

A. Basic NMR Parameters
- 1. Introduction to the NMR Experiment  
- 2. Chemical Shift  
- 3. Scalar Coupling  
- 4. Nuclear Overhauser Effect (NOE)

B. Structure Calculations

C. NMR Spectra: From One to Four Dimensions

### IV. Summary

---

### Chapter 2: Purification and Preliminary Characterization of the PLCC SH2 Domain

I. Introduction  

II. Results  

A. Purification  
- 1. Sub-cloning and Conditions for Protein Expression  
- 2. GST-PLCC SH2 Domain Purification  
- 3. Purification of pET-derived PLCC SH2 Domain  

B. Aggregation State of the Protein  

C. Initial NMR Spectra  

D. Backbone Assignment  

E. Binding Kinetics  

III. Summary

---

### Chapter 3: Structure of the pY1021 Phosphopeptide as it is Bound to the PLCC SH2 Domain

I. Introduction  

II. Materials and Methods
III. Results
   A. Assignment 117
   B. Peptide Structure 124
   C. Stereospecific Assignment, \( \chi_1 \) Restraints and Proline Rings 129
IV. Summary 145

Chapter 4: SH2 Domain-Phosphopeptide Contacts and Structure Calculations
I. Introduction 146
II. Experimental Methods 147
III. Results and Discussion 151
   A. Initial Experiments 151
   B. The Structure of the PLCC SH2 Domain/pY1021 Complex 153
      1. Overall Features 153
      2. Contacts to Asp -1 and pTyr 163
      3. Ile +1 174
      4. Pro +3 185
      5. Ile +2 188
      6. Residues Leu +4 to Asp +6 190
   C. The PLCC SH2 Domain/pY992 Complex 195
IV. Summary 196

Chapter 5: pH Titration Studies of an SH2 Domain-Phosphopeptide Complex
I. Introduction 197
II. Materials and Methods 201
III. Results and Discussion 203
   A. Histidine pK\(_a\) Values in the SH2 Domain-Peptide Complex 203
   B. Histidine pK\(_a\) Values in the Free PLCC SH2 Domain 213
C. pTyr Phosphate $pK_a$ 219
D. One-bond $^1H-^{15}N$ HSQC Spectra 222
E. Amide Exchange with Solvent 227

IV. Summary 232

Chapter 6: Investigations of Arginines and pTyr in the PLCC SH2 Domain/pY1021 Complex

I. Introduction 234

II. Materials and Methods 236

III. Results and Discussion 238
   A. Arginine Resonances of the PLCC SH2 Domain: Sample Conditions 238
   B. Backbone Amide Chemical Shift Changes 248
   C. Arg 37 Stereoassignment 256
   D. NOEs Involving the Arginine Guanidinium Group Protons 264
   E. Solvent Protection of the Arginine Guanidinium Groups 274
   F. Role of the Arginine Residues 277

IV. Summary 280

Chapter 7: Summary and Future Directions

I. Summary 281

II. Future Directions 282
   A. Electrostatic Interacions 282
      1. Hydrogen-bonding Restraints 283
      2. Inclusion of Electrostatic Energy into Structure Calculations 283
      3. Chemical Shift Restraints 284
   B. Time-averaged or Ensemble-averaged NOEs 285
   C. Mapping the Binding Site and Experimental Drug Design 286
D. Mutational Studies

E. Phosphopeptide Dynamics

F. Larger Fragments of PLC-γ

G. Summary

Appendix 1: Lineshape Calculations with Mathematica

Appendix 2: Stereoassignment and Restraints about the Dihedral Angle

χ1 for the pTyr Residue in the PLCC SH2 Domain/pY1021 Complex

Using the Program Stereosearch

Appendix 3: Measurement of J from Phase-Sensitive COSY Spectra

Appendix 4: Structural and Sequence Alignment of the BG Loop

References:
### List of Figures

**Chapter 1: Introduction**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Tertiary Structure of Tyrosine Kinase Domains</td>
<td>6</td>
</tr>
<tr>
<td>Figure 2</td>
<td>SH2 Domain-Containing Proteins</td>
<td>9</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Sequence Alignment of SH2 Domains</td>
<td>11</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Two Growth Factor Receptors and their Sites of Tyrosine Phosphorylation</td>
<td>14</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Phospholipase C Isozymes</td>
<td>20</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Tertiary Structure of the PLC-δ Catalytic Domain</td>
<td>27</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Tertiary Structure of the PLC-δ PH Domain</td>
<td>30</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Tertiary Structure of the PLC-γ SH3 Domain</td>
<td>33</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Coupling Constants in Proteins</td>
<td>41</td>
</tr>
<tr>
<td>Figure 10</td>
<td>NOEs and Secondary Structures in Proteins</td>
<td>45</td>
</tr>
<tr>
<td>Figure 11</td>
<td>NOEs and Structure Precision in Proteins</td>
<td>47</td>
</tr>
<tr>
<td>Figure 12</td>
<td>NMR Spectra: One and Two Dimensions</td>
<td>51</td>
</tr>
<tr>
<td>Figure 13</td>
<td>$^{15}$N-$^1$H HSQC Spectra of Folded and Unfolded Proteins</td>
<td>54</td>
</tr>
<tr>
<td>Figure 14</td>
<td>NMR Spectra: From Two to Four Dimensions</td>
<td>56</td>
</tr>
</tbody>
</table>

**Chapter 2: Purification and Preliminary Characteristic of the PLCC SH2 Domain**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Primary Sequence of PLCC SH2 Domain Constructs</td>
<td>63</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Induction of Expression of PLCC SH2 Domain Constructs</td>
<td>66</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Purification Scheme Used for the PLCC SH2 Domain Cloned into the pGEX-KT Vector</td>
<td>69</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Purification of PLCC SH2 Domain from the GST-PLCC SH2 Domain Fusion Protein</td>
<td>72</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Purification Scheme Used for PLCC SH2 Domains Expressed</td>
<td></td>
</tr>
</tbody>
</table>
from pET-11d Vectors

Figure 6 -- Purification of the PLCC SH2 Domain Expressed from the pET Expression Vector 1. Phosphocellulose Chromatography

Figure 7 -- Purification of the PLCC SH2 Domain Expressed from the pET Expression Vector 2. Phenyl Sepharose and Mono S Chromatography

Figure 8 -- Aggregation State of the PLCC SH2 Domain

Figure 9 -- $^{1}$H-$^{15}$N HSQC of the PLCC SH2 Domain

Figure 10 -- Backbone Assignment of the PLCC SH2 Domain

Figure 11 -- Exchange Kinetics of the PLCC SH2 Domain/pY1021 Complex

Figure 12 -- Exchange Kinetics of the PLCC SH2 Domain/pY992 Complex

Chapter 3: Structure of the pY1021 Phosphopeptide as it is Bound to the PLCC SH2 Domain

Figure 1 -- NMR Spectra: Filtered and Edited Experiments

Figure 2 -- Pulse Sequence Employed For the 2D $^{13}$C,$^{15}$N F$_{1}$/F$_{2}$-Filtered NOESY

Figure 3 -- Assignment of the pY1021 Resonances When Bound to the PLCC SH2 Domain

Figure 4 -- Extended Structure for the pY1021 Peptide in Complex with the PLCC SH2 Domain

Figure 5 -- Summary of Sequential NOEs in the pY1021 Peptide

Figure 6 -- pY1021 NOEs I

Figure 7 -- pY1021 NOEs II

Figure 8 -- Stereoassignment of H$\beta$ Protons of the pTyr Residue

Figure 9 -- Conformations About $\chi_1$ for Ile Residues

Figure 10 -- Proline Ring Conformations in the pY1021 Peptide
Chapter 4: SH2 Domain-Phosphopeptide Contacts and Structure Calculations

Figure 1 -- Tertiary Structure of the PLCC SH2 Domain/pY1021 Complex 156
Figure 2 -- Sequence Alignment and Secondary Structure of SH2 Domains 159
Figure 3 -- Surface and Electrostatic Potential of the PLCC SH2 Domain 161
in the PLCC SH2 Domain/pY1021 Complex 162
Figure 4 -- Protein-Peptide NOEs as a Function of Peptide Residue 165
Figure 5 -- Half Filtered NOESY: NOEs to Asp -1 and pTyr 168
Figure 6 -- Protein-Peptide NOEs Involving Asp -1 and pTyr 170
Figure 7 -- Protein-Peptide NOEs Involving the pTyr Residue: Summary by Individual pTyr Resonance 172
Figure 8 -- Half Filtered NOESY -- NOEs to Ile +1 and Pro +3 177
Figure 9 -- Protein-Peptide NOEs Involving Ile +1 to Pro +3 180
Figure 10 -- The Ile +1 Binding Site 182
Figure 11 -- The Pro +3 Binding Site 187
Figure 12 -- Protein-Peptide NOEs Involving Leu +4 to Asp 6 193

Chapter 5: pH Titration Studies of an SH2 Domain-Phosphopeptide Complex

Figure 1 -- Potential Titratable Groups in the PLCC SH2 Domain/pY1021 Complex Over the pH Range of 5-8 199
Figure 2 -- Representative 1H-15N HMBC Spectra of the Histidine Imidazole Ring 206
Figure 3 -- 1H-15N HMBC Spectra 208
Figure 4 -- Histidine 15Nδ1 Chemical Shifts for the PLCC SH2 Domain/pY1021 Complex as a Function of pH with Calculated pKa Values 211
Figure 5 -- Hydrogen-Bonding Network Including His βD4, Arg βB5 and the pTyr Residue 215
Figure 6 -- Demonstration of a Low pKa Value for His βD4 in Free and
pY1021 Complexed SH2 Domain

Figure 7 -- Low pK_a Value for the pTyr Phosphate in the PLCC SH2 Domain/pY1021 Complex

Figure 8 -- Arginine H_E Exchange Broadening Behavior as a Function of pH

Chapter 6 -- Investigations of Arginines and pTyr in the PLCC SH2 Domain/pY1021 Complex

Figure 1 -- ^15N-^1H HSQC Spectra

Figure 2 -- Arg Guanidinium Group Bidentate Hydrogen-Bonding Interactions

Figure 3 -- Arg ^1H_E and ^15N_E Chemical Shift Differences

Figure 4 -- Phosphate-Oxygen Hydrogen Bond Restraints

Figure 5 -- Backbone NH Chemical Shift Differences

Figure 6 -- Backbone ^15N Chemical Shift Differences

Figure 7 -- Geometry of the Arg Guanidinium Group

Figure 8 -- Arg 37 Stereoassignment

Figure 9 -- Intra-residue Arg Guanidinium Group NOEs

Figure 10 -- Inter-residue Arg Guanidinium Group NOEs: Arg-specific 2D ^15N NOESY

Figure 11 -- Inter-residue Arg Guanidinium Group NOEs: 3D ^13C NOESY

Figure 12 -- Summary of Arg 37 and His 57 Inter-residue NOEs

Figure 13 -- Superposition of Residues Arg 18, Arg 37 and Arg 39 from the Most Recent Structure Calculations

Figure 14 -- Guanidinium Group Solvent Protection

Appendix 3: Measurement of J from Phase Sensitive COSY Spectra

Figure 1 -- Measurement of J from Phase Sensitive COSY Spectra
Appendix 4: Structural and Sequence Alignment of the BG Loop

Figure 1 — Sequence Alignment of the BG Loop of a Number of SH2 Domains 309
List of Tables

Chapter 1: Introduction
Table 1: Comparison of PLC-γ SH2 Binding Sites 23

Chapter 2: Purification and Preliminary Characterization of the PLCC SH2 Domain
Table 1: NH and $^{15}$N Chemical Shifts for the Free and pY1021-Bound PLCC SH2 Domain 91
Table 2: Linewidth Data for the pY1021 Complex 98
Table 3: Linewidth Data for the pY992 Complex 103

Chapter 3 -- Structure of the pY1021 Phosphopeptide as it Bound to the PLCC SH2 Domain
Table 1: Comparison of Chemical Shifts of the Free and PLCC SH2 Domain-complexed pY1021 Phosphopeptide 122
Table 2: Coupling Constants Measured from the Double-Filtered PE-COSY of the PLCC SH2 Domain/pY1021 Complex (Hz) 135

Chapter 4 -- SH2 Domain-Phosphopeptide Contacts and Structure Calculations
Table 1: Structural Statistics for the PLCC SH2 Domain/pY1021 Solution Structure Determination 154
Table 2: ID50 and $\Delta(\Delta G)$ Values for Interactions Between Peptides and the PLCC and PLCN SH2 Domains (taken from Kay et al., 1997 and Steve Shoelson, personal communication) 175
Table 3: Comparison of Sequence of the PLCN SH2 Domain with the PLCC SH2 Domain at Residues Involved in +1 Binding 184
Chapter 5: pH Titration Studies of an SH2 Domain-Phosphopeptide Complex: Unusual Hisidine and Phosphate $pK_a$ Values

Table 1: Phosphate $^{31}P$, Histidine $^{15}N_{\delta 1}$, and Amide $^1H$ and $^{15}N$ pH Titration Data
Abbreviations:

C2 domain — Ca\(^{2+}\)-binding \(\beta\)-barrel of 120 amino acids whose structure was first described in a fragment of synaptotagmin

CN-NOESY — 3D NOESY which simultaneously records the \(^{13}\)C and \(^{15}\)N chemical shifts of the attached protons involved in the NOE interaction

COSY — correlated spectroscopy

DAG — diacyl glycerol

DHFR — \textit{E. coli} dihydrofolate reductase

EF hand — helical Ca\(^{2+}\)-binding domain

EGFR — epidermal growth factor receptor

FGFR — fibroblast growth factor receptor

GST — \textit{Schistosoma japonicum} glutathione S-transferase

HCCH-COSY — 4D COSY which measures the proton and \(^{13}\)C chemical shifts of the coupled spins

HMQC — heteronuclear multiple quantum coherence

HOHAHA — homonuclear Hartmann-Hahn

\(^{15}\)N-HOHAHA — 3D HOHAHA experiment recording the amide \(^{15}\)N, NH and scalar coupled proton spins

HSQC — heteronuclear single quantum coherence

IP\(_3\) — inositol 1,4,5 triphosphate

IR — insulin receptor

LB — Luria Broth

MAPK — mitogen-activated protein kinase (a Ser/Thr kinase)

MD — molecular dynamics

NOE — Nuclear Overhauser Effect

NOESY — Nuclear Overhauser Effect Spectroscopy
$^{13}$C-NOESY – 3D NOESY experiment in which the $^{13}$C chemical shift of one of the NOE partners is recorded

$^{15}$N-NOESY – 3D NOESY experiment in which the $^{15}$N chemical shift of one of the NOE partners is recorded

PDGFR – platelet derived growth factor receptor

PE-COSY – primitive E-COSY

PH Domain – plextrin homology domain

PIP2 – phosphatidylinositol-4,5-bisphosphate

PKC – protein kinase C (a Ser/Thr Kinase)

PLC – phospholipase C

PLCC – C-terminal SH2 domain of phospholipase C-γ

PLCN – N-terminal SH2 domain of phospholipase C-γ

PTB Domain – phosphotyrosine-binding domain

pTyr – phosphotyrosine

pY1021 – phosphopeptide representing sequences about Tyr 1021 of the PDGFR, with sequence DNDpYIIPLPDPK

pY992 – phosphopeptide representing sequences about Tyr 1021 of the EGFR, with sequence ADEpYLIPQQGF

RMSD – root mean squared deviation

SH2 Domain – Src homology 2 Domain

SH3 Domain – Src homology 3 domain

TB – Terrific Broth

TIM Barrel – β-barrel fold observed first in the structure of triose phosphate isomerase

TK Domain – tyrosine kinase domain

VDW – van der Waals

22 – fragment of PLC-γ containing both SH2 domains

xviii
22Y3 -- fragment of PLC-γ containing both SH2 domains, the SH3 domain and sequences between

P22Y3H -- fragment of PLC-γ containing the second PH domain and all intervening sequences
Chapter 1: Introduction

I. Overview

In the following chapters, I will be presenting my studies of the solution structure, dynamics and electrostatic properties of a src homology 2 (SH2) domain in complex with a high affinity phosphopeptide. Studies predominantly involve the C-terminal SH2 domain of the bovine phospholipase C-γ1 (PLC-γ) enzyme (referred to as the PLCC SH2 domain) in complex with a 12-mer phosphopeptide derived from sequences about Tyr 1021 of the platelet derived growth factor receptor (PDGFR)(referred to as the pY1021 phosphopeptide). The thesis is organized into the following chapters:

1. An introduction to SH2 domains, the enzyme PLC-γ, and their roles in signal transduction, as well as a discussion on NMR techniques used for structure determination.
2. Discussion of the cloning and purification of this SH2 domain and initial NMR characterization of the PLCC SH2 domain, the PLCC SH2 domain/pY1021 complex, and another complex of the PLCC SH2 domain with an 11-residue high-affinity phosphopeptide derived from sequences about the epidermal growth factor receptor (EGFR)(referred to as the pY992 phosphopeptide).
3. Resonance assignment and structure of the pY1021 phosphopeptide bound to the PLCC SH2 domain.
4. Characterization of NOEs between the PLCC SH2 domain and the pY1021 phosphopeptide, and results of structure calculations. As the results of this chapter are NOE-based, the studies of the electrostatic interactions required for pTyr recognition are dealt with in Chapter 6.
5. Characterization of the electrostatic properties of the PLCC SH2 domain/pY1021 complex by determination of pKₐs of several titratable groups over the pH range of 5-8.
6. Characterization of the electrostatic interactions required for pTyr recognition. This involves determination of hydrogen bonds to the pTyr phosphate group by chemical shift studies, as well as determination of NOEs of a number of Arg residue guanidinium groups shown to be important for pTyr recognition.

7. Summary of the work in the previous five chapters, plus possibilities for future work.

This chapter is an introduction both to the system under study as well as the NMR techniques described in the following chapters. The biological importance of the system under study is discussed in terms of the molecular biology, biochemistry and structural biology of growth factor receptors, SH2 domains, a number of other protein recognition domains involved in signal transduction, and the PLC enzymes. This is followed by a brief introduction to the basic parameters of NMR (chemical shift, coupling constant and NOE), with a discussion of 2D, 3D and 4D NMR and structure calculations.

II. Signal Transduction Pathways

Signal transduction is the process by which cells react to external signals such as growth factors, resulting in cell growth and differentiation. In the first part of this chapter, I review the initial steps of signal transduction, concentrating on the receptor tyrosine kinases. Then I discuss their downstream effectors, namely SH2-domain containing proteins and the role of SH2 domains. In addition, a number of other protein recognition domains involved in signal transduction are introduced. Since the SH2 domain under study is from PLC-γ, I also discuss the structure and function of the PLC isozymes. Results of molecular biological, genetic and biochemical techniques are discussed, as well as information gleaned from structural biology.
A. Growth Factor Receptors

1. General

The first step in signal transduction involves recognition of a growth factor (or other extracellular stimulus) by its receptor. There are many different classes of receptors, however one of the largest families of receptors are those which contain intrinsic tyrosine kinase (TK) activity. Tyrosine kinase activity is induced upon receptor binding, resulting in autophosphorylation and phosphorylation of target molecules, which in turn become substrates for binding of SH2 domain-containing proteins.

The receptor tyrosine kinases have widely divergent extracellular regions responsible for binding cytoplasmic ligands (reviewed by Fantl et al., 1993). In contrast, the intracellular region with its tyrosine kinase domain is more highly conserved. Upon ligand binding, the growth factor receptor is thought to dimerize (Schlessinger, 1988). Dimerization of the extracellular domain then leads to activation of the tyrosine kinase domain and cross-phosphorylation of the receptor (reviewed in Heldin, 1995).

It was initially thought that receptor tyrosine kinases would have downstream substrates, however often the best substrates for receptor tyrosine kinases are the receptors themselves. Autophosphorylation sites on the receptor allow binding of a number of proteins through their SH2 domains (Anderson et al., 1990; Moran et al., 1990), which are small regulatory domains of approximately 100 amino acids. SH2 domains were initially characterized by a comparison of the primary structure of cytoplasmic tyrosine kinases (such as the src proto-oncogene), in which three regions of homology were found (Sadowski et al., 1986). The most C-terminal region, the SH1 (or src homology 1) domain, (approximately 250 amino acids) contains the tyrosine kinase domain. The other regions of sequence homology, which make up the central SH2 (src homology 2) and N-terminal SH3 (src homology 3) domains (approximately 100 and 60 amino acids,
respectively), contain no homology to sequences of known catalytic activity. However, it is now known that these domains mediate protein recognition and targeting.

2. Structural Biology

Enzymatic activity of receptor tyrosine kinases is tightly regulated, such that activation occurs only following ligand binding, receptor dimerization and autophosphorylation. The mechanism for activation of the kinase domain by dimerization has been hypothesized based on the crystal structure of the unphosphorylated tyrosine kinase (TK) domains of the insulin receptor (IR-TK) (Hubbard et al., 1994) and fibroblast growth factor receptor (FGFR-TK) (Mohammadi et al., 1996). The overall topology of both TK domains is very similar to that of protein kinases specific for serine and threonine residues such as the cAMP-dependent protein kinase (Knighton et al., 1991), with a large and small domain, such that both the substrate and ATP can bind in the cleft formed between these two domains (see Figure 1a). TK domains contain a flexible loop (termed the "activation loop") with a number of sites of Tyr phosphorylation, and autophosphorylation of this loop must occur for the TK to be catalytically active. The two TK structures show two different mechanisms by which the unphosphorylated activation loop inhibits TK activity. In the IR-TK, a portion of the activation loop occupies the ATP-binding site and residues 1161-1163 lie in the substrate-binding site, while in the FGFR-TK, Pro 663 of the activation loop is positioned such that the substrate Tyr cannot bind in the active site (see Figure 1b). It is speculated that these loop conformations have intrinsic flexibility (Hubbard et al., 1994), so that the loop can become a substrate for transphosphorylation by a nearby kinase molecule. Such a nearby kinase molecule could be present due to receptor dimerization via ligand binding. Once this Tyr is phosphorylated, specific electrostatic interactions of the P_i group may stabilize another conformation of the
Figure 1 -- Tertiary Structure of Tyrosine Kinase Domains. A. Tertiary structure of the FGFR TK domain (figure taken from Mohammadi et al., 1996) showing the two domains, between which is a cleft into which ATP fits. Between strand β8 and helix αEF is a loop, termed the activation loop, which contains two Tyr residues known to act as substrates for phosphorylation. B. Space-filling model of the FGFR TK and insulin receptor (IR) TK (Hubbard et al., 1994) (figure taken from Mohammadi et al., 1996). The backbone of the activation loops is displayed as a worm model, showing that regions of the activation loop (colored orange) occupy a similar position in the two molecules. A portion of the IR-TK activation loop (coloured green) however occupies additional space within the substrate-binding cavity and interferes with ATP binding. The activation loop of the FGFR-TK does not interfere with ATP binding, and ATP and non-hydrolysable analogues (such as AMP-PCP) can bind in this site.
activation loop, allowing ATP and the peptide or protein substrate to enter into the active site and up-regulate the tyrosine kinase. Recently, the crystal structure of the TK domain of the cytoplasmic tyrosine kinase lck uniquely phosphorylated at this site has been solved (Yamaguchi & Hendrickson, 1996), and the pTyr phosphate group was found to be part of a hydrogen-bonding network involving residues in the activation loop and catalytic loop which places the catalytic residues in a proper position for enzymatic activity. Thus Tyr phosphorylation at this position in both receptor and cytoplasmic TKs may also play a role in stabilization of the active site.

B. SH2 Domains

1. General

SH2 domains are found in a wide range of proteins predominantly involved in signal transduction. SH2 domain-containing proteins can be divided into two classes (see Figure 2), namely those with catalytic activities such as cytoplasmic tyrosine kinases (src, abl) and phosphatases (syp or PTP1C), ras GTPase-activating protein (GAP) and PLC-γ, and those without an identifiable catalytic activity, containing only SH2 and other protein recognition domains. Such proteins are thought to act by bridging interactions between proteins and have been referred to as "molecular adaptors" (Koch et al., 1991). Examples of such SH2 domain-containing adaptor proteins include sem-5/GRB-2, SHC, and crk.

In Figure 3, an alignment of SH2 domain sequences (taken from Marengere & Pawson, 1992) is shown, in which five regions of sequence similarity are obvious. Within these regions of sequence similarity are a number of well-conserved residues, including a Trp defining the N-terminus of the SH2 domain in region I, an Arg in region I, the so-called "FLVR" sequence in region II and a His in region III. The two highly-conserved Arg residues have been shown in crystal and NMR structures to be involved in pTyr
Figure 2 -- SH2 Domain-Containing Proteins. A number of SH2 domain-containing proteins (figure taken from Pawson & Gish, 1992), showing number and position of the SH2 domains and other domains in the primary sequence. The proteins are sub-divided into those containing an enzymatic activity and those which do not, acting as adaptors to bridge interactions between signalling molecules.
**SH2/SH3 ENZYMES**

- **Src**: 3 SH2 KINASE
- **Abl**: 3 SH2 KINASE
- **Syk**: SH2 SH2 KINASE
- **PTP1C**: SH2 SH2 PTPase
- **PLCγ**: PLC SH2 SH2 3 PLC
- **GAP**: SH2 SH2 GAP
- **Vav**: DBL 3 SH2 3

**ACTIVITY**

- **Src**: TYROSINE KINASE
- **Abl**: TYROSINE KINASE
- **Syk**: TYROSINE KINASE
- **PTP1C**: PHOSPHOTYROSINE PHOSPHATASE
- **PLCγ**: PHOSPHOLIPASE C
- **GAP**: Ras GTPase ACTIVATION
- **Vav**: GUANINE NUC. EXCHANGE

**SH2/SH3 ADAPTORS**

- **p85 PI3K**: 3 SH2 SH2
- **c-Crk**: SH2 3 3
- **SHC**: GLY/PRO SH2
- **Nck**: 3 3 3 SH2
- **Sem-5/GRB2**: 3 SH2 3
- **ISGF3α**: 3 SH2

**TARGETS**

- **p85 PI3K**: PI 3′-KINASE
- **Sem-5/GRB2**: Ras PATHWAY
- **ISGF3α**: ISGF3γ
Figure 3 -- Sequence Alignment of SH2 Domains. Sequence alignment of SH2 domains (figure taken from Marengere & Pawson, 1992) showing five regions of sequence similarity in SH2 domains.
binding (i.e. Waksman et al., 1993), and mutations at these residues as well as the conserved His residue have been shown to weaken or abolish pTyr binding in some SH2 domains (Marengere & Pawson, 1992, Mayer et al., 1992). The Trp residue has been shown to play a structural role in the packing of the hydrophobic core.

SH2 domains were originally shown to bind to sites of tyrosine phosphorylation on growth factor receptors (Anderson et al., 1990; Moran et al., 1990). This binding was later shown to be sequence specific, in that SH2 domains from different signaling proteins bind to different sites of phosphorylation. This phenomenon has been observed both in vitro and in vivo. For example, EGFR has numerous sites of tyrosine phosphorylation at the C-terminus of the protein (see Figure 4); Tyr 992 is a binding site for PLC-γ (Rotin et al., 1992) and a minor site of binding for SHC (Batz et al., 1994), Tyr 1068 is a major binding site for GRB-2 (Batz et al., 1994), Tyr 1086 a binding site for abl (Zhu et al., 1994) and a minor site for GRB-2 (Batz et al., 1994), and Tyr 1173 is a major site for SHC (Batz et al., 1994). PDGFR has a number of sites of tyrosine phosphorylation N- and C-terminal to the kinase domain, as well as in a 100-amino acid region between the kinase domain termed the "kinase insert"; it has been shown that residue Tyr 716 is a binding site for GRB-2 (Ann-Kristin Arvidsson et al., 1994), Tyr 740 and Tyr 751 are binding sites for p85 (Kazlauskas et al., 1992), Tyr 751 is a binding site for nck (Nishimura et al., 1993), Tyr 771 for GAP, Tyr 1009 for syp (Valius & Kazlauskas, 1993) and Tyr 1021 for PLC-γ (Valius & Kazlauskas, 1993). With respect to PLC-γ SH2 domains, high-affinity binding sites have been found on other receptor tyrosine kinases such as FGFR (Mohammadi et al., 1991), trk (Obermeier et al., 1993), hepatocyte growth factor receptor (HGFR) (Ponzetto et al., 1994) and ret (Borrello et al., 1996).

To understand sequence specificity, the primary sequences of high-affinity SH2 domain-binding sites were compared. The sequence pYV/MXM was conserved among high-affinity binding sites for p85 SH2 domains (Cantley et al., 1991; Excobedo et al., 1991), and it was speculated that SH2 domain specificity lies within the first three residues
Figure 4: Two Growth Factor Receptors and their Sites of Tyrosine Phosphorylation. Schematic diagram of EGFR and PDGFR showing their sites of Tyr phosphorylation, their position relative to the tyrosine kinase domain, and the SH2 domain-containing and PTB (pTyr binding) domain-containing proteins which have been shown to bind to each particular site of phosphorylation.
Legend:

- Kinase Domain
- Extracellular Domain
- Membrane-Spanning Domain
- Tyr Phosphorylation Region
- Plasma Membrane
- SH2 Domain
- PTB Domain
- pTyr Residue
C-terminal to pTyr. Binding studies of SH2 domains to phosphopeptide libraries degenerate at positions +1, +2 and +3 relative to the pTyr were performed, followed by peptide sequencing of the bound peptides (Songyang et al., 1993). Different SH2 domains were shown to possess different degrees of specificity, some showing weak specificity (e.g. abl), while others showed a tightly-defined specificity. Most SH2 domains were found to prefer hydrophobic amino acids at the +3 position. However, SH2 domains have been subdivided into at least three classes based on preference for hydrophilic amino acids (type I, e.g. cytoplasmic tyrosine kinases) or hydrophobic amino acids (type II, e.g. vav, and type III, e.g. p85, PLC-γ) at the +1 position (Songyang et al., 1994). However, there is considerable variability in the specificity of members of class I and class III.

SH2 domain-containing proteins often act as links between receptor activation and later events in signal transduction. A number of SH2 domain-containing proteins have enzymatic activities which lead to production of second messengers. For example, p85 acts as a regulatory subunit for phosphatidylinositol-3-kinase, which phosphorylates inositol phospholipids at the 3' position, and PLC-γ hydrolyses phosphatidylinositol 4,5 bis-phosphate (PIP2) to the second messengers diacylglycerol (DAG) and inositol triphosphate (IP3). In addition, a number of SH2 domain-containing proteins also contain SH3 domains, which act by binding proline-rich sequences on proteins. Proline-rich sequences on the guanine nucleotide exchange factor sos have been shown to be a target for SH3 domain-containing proteins, specifically the small adapter protein GRB-2 (Olivier et al., 1993; Rozakis-Adcock et al., 1992). Sos in turn activates the small G protein ras, which then activates a Ser/Thr kinase cascade involving the kinases raf, mek and MAPK.

2. Structural Biology

Several structures of SH2 domains have been solved by X-ray crystallography or in solution by 3D heteronuclear NMR techniques. Structures of SH2 domains in isolation
which have been determined include the solution structures of the N-terminal SH2 domain of p85 (Booker et al., 1992) and the abl SH2 domain (Overduin et al., 1992a; Overduin et al., 1992b) as well as the crystal structure of the src SH2 domain (Waksman et al., 1993). Structures of SH2 domain-phosphopeptide complexes have also been determined (Breeze et al., 1996; Eck et al., 1993; Hatada et al., 1995; Lee et al., 1994; Narula et al., 1995; Nolte et al., 1996; Pascal et al., 1994; Rahuel et al., 1996; Waksman et al., 1992; Waksman, et al., 1993; Zhou et al., 1995). In these structures, the N- and C-termini of the SH2 domain are reasonably close in space, consistent with their modular nature. This allows the domain to be inserted into a protein without disrupting the overall structure or fold of the protein to which it is being inserted. This structural feature is common to all the protein recognition domains being discussed in this chapter.

Detailed discussion of structures of SH2 domains and SH2 domain-phosphopeptide complexes is found in later chapters, where comparison is made to the structure of the PLCC SH2 domain/pY1021 complex, which is the focus of this thesis.

C. PTB Domains

Another pTyr-binding (PTB) domain has been described which shares no structural or sequence homology with the SH2 domain. This domain, originally observed in the N-terminal region of SHC but which is found in a number of proteins involved in signal transduction (Bork & Margolis, 1995), is between 100 and 160 amino acids in length (Kavanaugh & Williams, 1994). Like SH2 domains, PTB domain binding is also sequence specific, however binding studies have indicated that sequence specificity lies primarily in residues N-terminal to the pTyr residue. Specifically, the sequence NPXY (where Y is the pTyr residue) has been shown to be important in binding (Butzer et al., 1995; Gustafson et al., 1995). In addition, there is affinity for hydrophobic residues N-terminal to the NPXY motif, and the position and type of hydrophobic amino acid forms
the basis of peptide specificity (van der Geer et al., 1996). Recently, the solution structures of the SHC (Zhou et al., 1995) and IRS-1 (Zhou et al., 1996) PTB domain have been solved in solution, and no structural homology to SH2 domains is observed. Surprisingly PTB domains are observed to have a strong degree of structural homology to the PH domain (to be described later) in that both contain a β-sandwich of two nearly orthogonal β-sheets and a C-terminal α-helix. The Ca atoms of the PH domain of pleckstrin and the PTB domain of SHC superimpose with an rmsd of 1.9 Å in the region of structural homology (Zhou et al., 1995). In addition to the observation of structural homology, binding of the PTB domain to phosphoinositol-containing lipids has been demonstrated in vitro (Zhou et al., 1995), which is also a function of PH domains.

The binding mechanism of PTB domains to SH2 domains is also very different in terms of the secondary structure of the bound peptide, and conservation of residues involved in pTyr binding. Further discussion concerning the structure of the PTB domain-peptide complexes will be included in later chapters in order to compare their structures to SH2 domains.

D. Phospholipase C

1. Role in Signal Transduction

One pathway in signal transduction involves events following activation of PLC-γ and other isozymes of PLC. PLC hydrolyses phosphatidylinositol 4,5 bisphosphate (PIP$_2$) to diacylglycerol (DAG) and inositol 1,4,5 triphosphate (IP$_3$), which then act as second messengers for downstream pathways in signal transduction processes. DAG binds and activates the Ser/Thr kinase protein kinase C (PKC). In turn, PKC phosphorylates and activates other proteins, including the Ser/Thr kinase raf (Kolch et al., 1993; Sozeri et al., 1992). IP$_3$ is bound by IP$_3$- and ryanodine receptors in the
endoplasmic reticulum, which then release intracellular stores of bound Ca$^{+2}$ (Berridge, 1993). PKC also helps to stimulate Ca$^{+2}$ release by phosphorylation of the IP3-receptor, which is important in opening the Ca$^{+2}$ channel (Cameron et al., 1995). Upon opening of Ca$^{+2}$ channels, the intracellular Ca$^{+2}$ concentration rises from approximately 10$^{-7}$ to 10$^{-5}$ M. Free Ca$^{+2}$ is necessary for the activity of a number of enzymes involved in signal transduction such as MAPK, PKC, calmodulin and the protein phosphatase calcineurin. Calmodulin is thought to be the most important protein for mediating the Ca$^{+2}$ signal (for a review, see James et al., 1995). It contains two domains which each bind Ca$^{+2}$, but the amino-terminal domain binds with a lower affinity (10$^{-6}$ M) than the carboxy-terminal domain (10$^{-7}$ M). Following Ca$^{+2}$ chelation, calmodulin can bind to a number of proteins and stimulate their activities.

Release of Ca$^{+2}$ from intracellular stores is tightly regulated in this system. Two proteins which are activated by increased Ca$^{+2}$ concentration, calmodulin and calcineurin, bind to a complex containing the IP3 receptor and the immunophilin molecule FKPB12 (Cameron et al., 1995). The association of the phosphatase calcineurin with the IP3 receptor then leads to dephosphorylation of the IP3 receptor at the PKC site, and closing of the Ca$^{+2}$ channel. Thus, the association of a kinase and a phosphatase with the IP3 receptor leads to oscillations in Ca$^{+2}$ concentration in the cell.

2. Primary Structure

The three isozymes of PLC, PLC-β, -γ and -δ (Rhee, 1991), contain independently-folded domains responsible for substrate binding, catalysis and regulation of enzymatic activity (see Figure 5). The amino acids responsible for the enzymatic activity lie within the X and Y sequences common to all PLC-γ isoforms. Initial results of deletion mapping and limited proteolysis experiments of PLC-δ (Cifuentes et al., 1993; Ellis et al., 1993) suggested that the regions of X and Y necessary for catalytic activity are approximately 30
Figure 5 -- Phospholipase C Isozymes. Schematic of the primary structure of the three PLC isozymes, showing the catalytic domains, protein recognition domains and sites of tyrosine phosphorylation that have been identified. Note that the catalytic regions are described by the older nomenclature of X and Y (Rhee, 1991) despite the recent structural evidence that the catalytic regions consist of three separate domains (Essen et al., 1996). The symbols are as follows: PH is the PH domain, G is a region interacting with Gα subunits, X and Y represent the two regions of sequence similarity necessary for catalytic activity, P and H represent the two halves of a PH domain in PLC-γ, N2 and C2 represent the N-terminal and C-terminal SH2 domains respectively, 3 represents the SH3 domain, and the three balls and sticks on PLC-γ represent sites of tyrosine phosphorylation.
and 40K respectively.

In addition to the X and Y regions, all isozymes of PLC contain at least one PH domain based on sequence homology (Parker et al., 1994). The role of PH domains in signal transduction is not fully understood, however structures of PH domains show a binding pocket rich in lysine and arginine residues suggested (Yoon et al., 1994) and later shown in vitro (Harlan et al., 1994) to bind PIP2, IP3 and other anionic phospholipids and phospholipid headgroups. However, it has also been suggested that PH domains are involved in interaction with βγ subunits of heterotrimeric G proteins (Koch et al., 1993; Touhara et al., 1994).

The PH domain of PLC-δ binds to PIP2 and IP3 with high affinity (Kd = 1.7 μM in lipid vesicles and 210 nM, respectively) and IP3 acts as a competitive inhibitor for PIP2 binding (Cifuentes et al., 1994). Binding of the PH domain to PIP2 can function to anchor PLC-δ to the membrane and to allow PIP2 hydrolysis to proceed via a "scooting" or "processive" manner, such that a number of PIP2 molecules are hydrolyzed before the enzyme dissociates from the membrane surface (Cifuentes et al., 1993). A similar role can be envisioned for the PH domains of the other two PLC isozymes, however such a role for the second PH domain of PLC-γ may be simplistic because this PH domain is split by an insert of more than 250 amino acids containing sequences for two SH2 domains, one SH3 domain and two sites of tyrosine phosphorylation. The N-terminal region of PLC-δ containing the PH domain has also been shown to be responsible for dimerization of this isozyme (Ellis et al., 1993).

PLC-δ activity does not seem to be regulated in any manner, but the other two isozymes of PLC are activated by external stimuli such as growth factors. PLC-β is activated by the Gq class of α-subunits of trimeric G-proteins, with residues 908-1142 of PLC-β being responsible for regulation (Wu et al., 1993). PLC-γ is activated by growth factors via its two SH2 domains, and the sequence specificity of these SH2 domains dictates the sites of Tyr phosphorylation on various receptors to which PLC-γ binds. It has
been shown in vitro and in vivo that Tyr 1021 and Tyr 992 are the principal binding sites for PLC-γ on the PDGFR (Larose et al., 1993; Valius et al., 1993) and EGFR (Rotin et al., 1992; Vega et al., 1992), respectively. Furthermore, in vivo experiments indicate that both the Tyr 1021 site and sites involved in p85 binding are important for DNA replication and cell division following PDGFR stimulation of HepG2 cells (Valius & Kazlauskas, 1993). However, the importance of PLC-γ in the signal transduction pathway may vary among cell types, as it has been found that the Y→F mutation of the principal PLC-binding site of flg had no effect on DNA replication and growth following serum stimulation in the cell lines tested (Mohammadi et al., 1992; Peters et al., 1992).

In both the PDGFR and EGFR sites, it is of interest whether the N-terminal, the C-terminal or both SH2 domains of PLC-γ bind to these phosphorylation sites. In experiments using degenerate phosphopeptide libraries (Songyang et al., 1993), differences in specificity between the PLCN SH2 domain and PLCC SH2 domain were observed (see Table 1). The clearest difference in specificity lies in position +2, in which there is a preference for a hydrophobic amino acid in the C-terminal SH2 domain, and for a negatively-charged amino acid in the N-terminal SH2 domain. Thus the PLCC SH2 domain is expected to bind preferentially to pY992 and pY1021, since Ile residues are found at the +2 position in both cases.

The presence of an SH3 domain in PLC-γ is also important. SH3 domains are found in many proteins containing SH2 domains, as well as a number of cytoskeletal proteins and proteins involved in signal transduction (Koch et al., 1991). These domains bind to proline-rich regions of target proteins, as first demonstrated by Cicchetti et al. (1992) and Ren et al. (1993), such as those present in guanine nucleotide exchange proteins (GNRPs) for ras or other small G-proteins (Cicchetti, et al., 1992; Olivier, et al., 1993; Rozakis-Adcock et al., 1992). SH3 domain binding regions have also been found on the tubulin-associated GTPase dynamin (Gout et al., 1993) and a sodium channel (Rotin et al., 1994). The SH3 domain of PLC-γ has been shown to bind proline-rich sequences in dynamin.
Table 1: Comparison of PLC-γ SH2 Binding Sites

I. C-terminal SH2 domain

Phosphopeptide library (Songyang et al., 1993)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Binding Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>YV (3.7) I (2.4) P (2.5)</td>
<td></td>
</tr>
<tr>
<td>I (2.9) L (2.0) V (2.2)</td>
<td></td>
</tr>
<tr>
<td>L (1.7) I (2.2)</td>
<td></td>
</tr>
</tbody>
</table>

PDGFR-β
*PDGFR-α
EGFR (human)
*let-23
*erb-B2 (human)
*human basic FGFR consensus

<table>
<thead>
<tr>
<th>Phosphopeptide library</th>
<th>Binding Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>flg</td>
<td>1012 VQPNEDNDYIIPLD PKP</td>
</tr>
<tr>
<td>bek</td>
<td>1009 EQLRSLADGYIIPDPDIDP</td>
</tr>
<tr>
<td>FGFR-3</td>
<td>988 MDDVDADEYLIPOGFFS</td>
</tr>
<tr>
<td>FGF-4</td>
<td>1232 GSTAQEDNSYLIPKTEVQ</td>
</tr>
<tr>
<td>met</td>
<td>1121 PSETDGYVAPLTCPQX</td>
</tr>
<tr>
<td>PLC-γ</td>
<td>XTXXXEYVIVEVEX</td>
</tr>
</tbody>
</table>

II. N-terminal SH2 domain

Phosphopeptide library (Songyang et al., 1993)

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Binding Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>YL (3.8) E (2.8) L (3.2)</td>
<td></td>
</tr>
<tr>
<td>I (3.2) D (1.8) I (2.7)</td>
<td></td>
</tr>
<tr>
<td>V (2.4) V (2.4)</td>
<td></td>
</tr>
</tbody>
</table>

flg
*bek
trk
ret

<table>
<thead>
<tr>
<th>Phosphopeptide library</th>
<th>Binding Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>flg</td>
<td>757 RVILTSNQELYLDLSPLDQ</td>
</tr>
<tr>
<td>bek</td>
<td>759 ILTLTTNEEYLDLSQPLEQ</td>
</tr>
<tr>
<td>trk</td>
<td>776 QALAQAPPVYLDVLG</td>
</tr>
<tr>
<td>ret</td>
<td>535 KRRDYLDLAASTPSDSL</td>
</tr>
</tbody>
</table>

Note: Those sequences which have not been shown to bind experimentally but are predicted based on sequence similarity are marked by an asterisk. Sequences are from Aroian et al. (1990), Borrello et al. (1996), Claesson-Welsh et al. (1989), Mohammadi et al. (1991), Obermeier et al. (1993), Partanen et al. (1991) Songyang et al. (1993), Ullrich et al. (1984).
(Seedorf et al., 1994); this interaction may be important in cytoskeletal changes which occur during signal transduction.

PLC-γ contains a number of sites of Tyr phosphorylation which have been demonstrated to enhance enzymatic activity following phosphorylation (Kim et al., 1991). In particular, Tyr 783, which lies in an approximately 40 amino-acid region linking the C-terminal SH2 domain and the SH3 domain, is important in activation of the enzyme. The mechanism is unknown, but it has been shown that the actin-binding protein profilin inhibits PLC-γ activity by competing for binding of PIP₂, and phosphorylation of PLC-γ can overcome this inhibition (Goldschmidt-Clermont et al., 1991). The PIP₂-binding site on profilin overlaps the actin-binding site, so when profilin is released from binding to PIP₂ by phosphorylated PLC-γ, it can in turn bind to actin, coupling PLC-γ activation to cytoskeletal changes. Another aspect of PLC-γ activation by tyrosine phosphorylation may involve intra-molecular interactions, in that it has recently been shown that the C-terminal SH2 domain binds to the phosphorylated Tyr 783 site (Gerry Gish and Neil Farrow, personal communication).

3. Structural Biology of Phospholipase C Isozymes

a. PLC-δ Catalytic Domain

A number of pathogenic prokaryotes code for phosphoinositol (PI)-specific PLC enzymes which act as potential virulence factors (Heinz et al., 1995 and references therein). These prokaryotic PLC enzymes are much smaller than the mammalian enzymes, being approximately 300 amino acids in size. However, within the X domain of mammalian PLC isozymes is a region of approximately 120 amino acids which shares homology with prokaryotic PLC (Cheng et al., 1995). Based on this homology, a number of residues in
PLC-δ which are conserved among mammalian, yeast and prokaryotic PLC isozymes have been shown by site-directed mutagenesis to be essential for catalysis (Cheng et al., 1995; Ellis et al., 1995). The crystal structure of the *B. cereus* phospholipase C in complex with myo-inositol has been solved (Heinz, et al., 1995). The protein was found to contain nine strands and eight helices arranged in a manner similar to the (βα)8 barrel originally described in triose phosphate isomerase and referred to as a TIM barrel. The modification from a strict TIM barrel involves the fourth β-strand being followed by two additional β-strands rather than an α-helix. Within the C-terminus of the β-barrel is a small cleft into which myo-inositol can fit and where the catalytic residues including a pair of well-conserved His residues are found.

Recently, the structure of a fragment of the rat PLC-δ1 protein containing the complete X and Y regions was solved (Essen et al., 1996; Grobler et al., 1996). Surprisingly, this fragment consists of three separate domains, an EF-hand domain containing four EF-hands (residues 133-281 of PLC-δ), a TIM barrel as observed in the prokaryotic PLC structure (residues 299-606), and a Ca+2-binding β-barrel first described by the structure of a fragment of synaptotagmin I, termed a C2 domain (residues 626-756) (see Figure 6). The TIM barrel includes sequences from the X and Y regions such that the boundary between the two sequence domains is formed by a flexible loop of approximately 40 amino acids. IP3 was found to bind in a cavity in the TIM barrel formed by residues from the X and Y regions. All hydroxyl and phosphate groups of IP3 make extensive interactions with residues in the TIM barrel with the exception of the 6-hydroxyl group. The C2 domain has no known catalytic role, however binding of lanthanide ions (analogous to Ca+2) opens up a cavity in the C2 domain large enough to fit the polar head group of an anionic lipid molecule (Grobler et al., 1996). Thus by binding to polar head groups of membrane lipids, the C2 domain is believed to help fix the orientation of the catalytic domain for binding and hydrolysis of inositol-containing phospholipids (Essen et al., 1996). The EF hand domain, as well as the TIM barrel and C2 domains all bind at
Figure 6 -- Tertiary Structure of the PLC-δ Catalytic Domain. The following figure is taken from Essen (1996), showing the backbone fold of the three domains which are present in the X and Y regions of PLC-δ.
least one Ca\textsuperscript{2+} ion, however no role for the EF hand in PIP\textsubscript{2} hydrolysis could be discerned from this study.

b. The PH Domain

Several structures of isolated PH domains have been reported, including the PH domains of dynamin (Ferguson et al., 1994), pleckstrin (Yoon et al., 1994) and \(\beta\)-spectrin (Maclas et al., 1994). Binding studies have shown that PH domains bind negatively-charged lipid headgroups, as demonstrated by the structure of the PH domain of PLC-\(\delta\) in complex with IP\textsubscript{3} (Ferguson et al., 1995). This structure will be discussed both in terms of its importance for PLC activity and with respect to PH domains in general. It is of note that the interaction of the PLC-\(\delta\) PH domain with IP\textsubscript{3} is the strongest PH domain/lipid or lipid-headgroup interaction described; PH domain interactions are frequently of low affinity (\(K_d \sim 30\) \(\mu\)M) in comparison to the high affinity PLC-\(\delta\) PH domain/IP\textsubscript{3} interaction (\(K_d = 210\) nM, Ferguson et al., 1995).

In Figure 7a is shown the overall backbone fold of the PH domain of PLC-\(\delta\) in complex with IP\textsubscript{3} (taken from Ferguson et al., 1995). The structure may be described as a seven-stranded \(\beta\)-sheet containing a four-stranded and a three-stranded sheet which lie orthogonal to each other. A C-terminal \(\alpha\)-helix is also present capping one side of the \(\beta\)-sheet structure. In addition, there are two small \(\alpha\)-helices, one at the N-terminus and another between strands \(\beta_5\) and \(\beta_6\). These two small helices are not involved in binding IP\textsubscript{3} and are not found in the structures of other PH domains solved to date. In the PH domain of \(\beta\)-spectrin a small \(\alpha\)-helix is observed between strands \(\beta_3\) and \(\beta_4\); this is significant because the second PH domain of PLC-\(\gamma\) contains an insertion of >250 amino acids containing two SH2 domains, an SH3 domain and an approximately 40 amino acid region containing sites of Tyr phosphorylation. IP\textsubscript{3} is buried deeply in a pocket formed by loops between strands \(\beta_1\) and \(\beta_2\) and between \(\beta_3\) and \(\beta_4\). Numerous hydrogen bonds are
Figure 7 -- Tertiary Structure of the PLC-δ PH Domain. A. Interaction of the PLC-δ PH domain with an inositol-containing phospholipid in the plasma membrane showing the orientation in the plasma membrane and the backbone fold of the PH domain. Specific contacts between the PH domain and IP3 are shown in part B. The following figures were taken from Ferguson et al. (1995).
present between residues in this binding pocket and the 1-, 4- and 5-phosphate groups (see Figure 7B).

c. The SH3 Domain

The SH3 domain of PLC-γ has been solved by heteronuclear NMR techniques (Kohda et al., 1993), but shows no features unusual among the SH3 domain structures solved to date. Several SH3 domains have been solved by NMR and X-ray crystallography, in isolation (Booker et al., 1993; Borchert et al., 1994; Kohda et al., 1994; Koyama et al., 1993; Musacchio et al., 1992; Noble et al., 1993; Yu et al., 1993) and in complex with high-affinity Pro-rich peptides (Feng et al., 1994; Musacchio et al., 1994; Wittekind et al., 1994; Yu et al., 1994). The overall fold of the PLC-γ SH3 domain is shown in Figure 8A; the SH3 domain contains two three-stranded β-sheets oriented orthogonally to each other. Within this structure is a hydrophobic surface rich in aromatic amino acids containing many of the residues conserved among SH3 domains, which forms the Pro-rich peptide binding site (see Figure 8b). The role of these aromatic residues in proline recognition has also been shown by site-directed mutagenesis (Lim & Richards, 1994). The proline residues of the peptide adopt a left-handed type II poly-Pro (PPII) helix conformation, in which two of the Pro residues in each turn of the helix contact the protein. Outside of the hydrophobic binding pocket are a pair of loops which are strongly acidic. The presence of basic residues on the SH3 domain ligands just outside the Pro-rich sequence are important in determining specificity and directionality of peptide binding.

E. Interaction of Domains in Signaling Molecules

We have studied the PLCα SH2 domain in isolation, but it is important to consider that PLC-γ, like many signal transduction proteins, is a multi-domain protein. There has
Figure 8 -- Tertiary Structure of the PLC-γ SH3 Domain. Figures taken from Kohda et al. (1993). A. Overall fold of the PLC-γ SH3 domain. Residues from strand β2, the β5-β6 loop, strand β7 and the β7-β8 loop contain a number of conserved hydrophobic amino acids which form a hydrophobic patch as illustrated in another view by the stereo pair in part B, which has been shown to be the binding site for the Pro-rich peptides in structures of other SH3 domain/peptide complexes.
been interest as to whether these domains interact with each other or are independent, and a number of investigations have been performed on larger peptides containing at least two of these domains. In a number of cases, the separate domains do not interact to a significant extent in solution. For example, the chemical shifts of a peptide containing the SH2 and SH3 domains of abl are practically identical to the sum of a spectrum containing the chemical shifts of the SH2 domain and the SH3 domain in isolation (Gosser et al., 1995). Similarly, many of the chemical shifts of PLCC SH2 domain can be observed in a construct containing both SH2 domains of PLC-γ (Neil Farrow, personal communication). This is in contrast with the crystal structure of a construct containing the SH2 and SH3 domains of lck which demonstrates an inter-molecular association of the SH3 domain with a proline-rich loop region in the SH2 domain (Eck et al., 1994). Differences in binding affinity of SH2 domain and SH3 domain molecules when the other is present in cytoplasmic Tyr kinases indicate that this interaction may be biologically relevant (Panchamoorthy et al., 1994). In addition, Tyr phosphorylation of crk has been shown to lead to an intramolecular association via its SH2 domain (Rosen et al., 1995).

Recently, three crystal structures of complete or nearly complete signaling molecules indicate extensive interaction between domains. The crystal structure of GRB-2 (see Figure 2) has been reported (Maignan et al., 1995), and crystal packing interactions yield a dimeric GRB-2 in which the C-terminal SH3 domain of one molecule contacts the SH2 domain of another. The dimerization surfaces are not close to the binding surfaces of either the SH2 domain or SH3 domain, so this dimerization does not alter phosphopeptide or Pro-rich peptide binding. In addition, the crystal structures of the nearly complete src (Xu et al., 1997) and hck (Sicheri et al., 1997) molecules containing SH3, SH2 and TK domains have recently been reported and indicate interaction of the SH2 domain with the C-terminal Tyr-phosphorylated tail and interaction of the SH3 domain with the linker between the SH2 domain and TK domain. Though the SH3 domain interaction is far from the active site of the TK, this network of hydrophobic interactions repositions a helix involved
in TK catalysis, inhibiting activity. Thus, removal of this SH3 domain interaction significantly enhances TK activity (Moarefi et al., 1997).

III. Introduction to Nuclear Magnetic Resonance (NMR)

A. Basic NMR Parameters

1. Introduction to the NMR Experiment

When a sample such as a protein is placed in an external magnetic field, NMR-active nuclei (such as \(^1\)H, \(^{13}\)C and \(^{15}\)N nuclei) possess angular momentum or spin, and the frequency at which they rotate is quantized, i.e. is limited to spin states \((-I, -I+ 1, -I+2, ..., I-2, I-1, I)\), where \(I\) is the spin quantum number of the nucleus of interest. Spin 1/2 nuclei are particularly advantageous because there are only two spin states, \(-1/2\) and \(1/2\). Spin 1/2 nuclei include \(^1\)H, \(^{13}\)C, \(^{15}\)N, \(^{19}\)F, and \(^{31}\)P.

The two spin states of spin 1/2 nuclei can align with (\(\alpha\), spin 1/2) or against (\(\beta\), spin -1/2) an external magnetic field. The energetics of this system can be described by the equation:

\[
U = -\gamma h B_o
\]

where \(U\) is the energy difference between the two spin states, \(\gamma\) is the magnetogyric ratio of the nucleus, and \(B_o\) is the strength of the magnetic field. Thus the energy difference is proportional to the strength of the magnetic field, i.e. the stronger the field, the larger the energy difference between the two spin states. In addition, the energy difference between two nuclei is proportional to \(\gamma\), which is a constant for each nucleus, while the signal intensity is proportional to \(\gamma^3\). Since the magnetogyric ratio of the proton is approximately 4 and 10 times larger than those of the \(^{13}\)C and \(^{15}\)N nucleus, the energy difference
between the \( \alpha \) and \( \beta \) states for \(^1\text{H}\) nuclei is 4 and 10 times larger than \(^{13}\text{C}\) and \(^{15}\text{N}\) nuclei, respectively, resulting in \(^1\text{H}\) nuclei being 64 and 1000 times more sensitive than \(^{13}\text{C}\) and \(^{15}\text{N}\) nuclei, respectively. However, with even very large magnetic fields, the energy difference, and thus the population difference between the two spin states, is small. Observation of these small population differences in NMR spectroscopy requires a high sample concentration, higher than that needed for other spectroscopic techniques such as CD or fluorescence.

When a sample is placed in a static magnetic field, the nucleus aligns either in the direction of or opposite to the external magnetic field; however, no transitions between the two spin states occur. Absorption of energy and transitions between the two spin states can be accomplished by applying an oscillating magnetic field in a direction orthogonal to the external magnetic field. The absorption intensity can then be plotted as a function of frequency leading to a one-dimensional spectrum.

2. Chemical Shift

From the discussion so far, one would assume that all protons resonate at the same frequency. In fact, electrons interact with the external field and set up an opposing field which in turn affects the magnitude of the magnetic field felt by the nucleus. Since different nuclei in a molecule are within different electronic environments, the magnitude of the opposing field varies from nucleus to nucleus. The effect of the opposing field is described by the equation

\[
\text{H}_{\text{eff}} = H(1 - \sigma)
\]

where \( \text{H}_{\text{eff}} \) is the effective magnetic field, \( H \) is the external magnetic field and \( \sigma \) is the screening constant which describes the effect of the electronic environment. Practically, as
different nuclei "see" different local magnetic fields depending on their chemical environment, they absorb electromagnetic radiation at slightly different frequencies. This difference in frequency is defined as the chemical shift,

\[ \delta = \left( \frac{v - v_{\text{ref}}}{SF} \right) \times 10^6 \]

where SF is the spectrometer frequency, v is the frequency of the nucleus of interest, v_{\text{ref}} is the frequency of absorption of a reference, and \( \delta \) is the chemical shift of the nucleus, expressed in units of ppm, or parts per million.

In amino acids and proteins, chemical shifts are dominated by three effects. First, electrons in \( \pi \)-bonds have large effects on the chemical shift of nuclei, and these effects are especially strong in conjugated double bond systems such as aromatic rings, in which the \( \pi \) electrons are thought to circulate like an electrical current above and below the plane of the ring. Second, the proximity of charged groups to the nucleus of interest is known to affect chemical shift. Finally, chemical bonds which show an asymmetric charge distribution, or bond anisotropy, can effect the chemical shift of nearby nuclei.

In the following chapters, I discuss chemical shift changes in a number of situations with respect to the structure of the PLCC SH2 domain/pY1021 complex and dissect the contributions of each of the three effects on the chemical shift changes observed. With respect to the effect of anisotropic chemical bonds, the C=O and C-N bonds of the peptide backbone contribute strongly, and since the position of these bonds depends on the \( \phi \) and \( \varphi \) backbone dihedral angles, chemical shift differences are observed depending on secondary structure. For example, H_\( \alpha \) protons are shifted downfield if the backbone adopts a \( \beta \)-strand conformer but are shifted upfield if the peptide backbone adopts an \( \alpha \)-helical structure. The chemical shifts of \(^{13}\text{C}_\alpha \) and \(^{13}\text{C}_\beta \) nuclei have also been found to be strongly dependent on the conformation of the peptide backbone (Spera & Bax, 1991). The combined effect of these chemical shift differences for H_\( \alpha \), \(^{13}\text{C}_\alpha \) and \(^{13}\text{C}_\beta \) nuclei
from random coil values can be used to predict secondary structure using the Chemical Shift Index, or CSI (Wishart & Sykes, 1994).

Since chemical shift is strongly effected by charged or ionizable groups, the effect of chemical shift will vary depending on the ionization state of the chemical group. Thus the chemical shift of a nearby group can be measured as a function of pH and fit to a sigmoidal curve in order to determine the pK\textsubscript{a} of this group. This approach is described in a later chapter (see Chapter 5) in a study of the electrostatics of the PLCC SH2 domain/pY1021 complex.

Finally, when the pY1021 phosphopeptide binds to the PLCC SH2 domain, a number of chemical shift changes occur. With respect to pTyr binding, this may be due to effects of the close approach of the peptide backbone of the pTyr, -1 or +1 residue, binding of the aromatic ring of the pTyr, or effects due to the charge and bond anisotropy effects of the phosphate group. In Chapter 6, I discuss an approach to dissecting the individual effects of the chemical shift changes observed.

3. Scalar Coupling

NMR-active nuclei can be affected by the spin state of other nuclei separated by 1 to approximately 5 chemical bonds via scalar coupling. For homonuclear proton spectra, the magnitude of this coupling (J-coupling) becomes very small when the protons are separated by more than three chemical bonds (usually <1 Hz). In terms of the appearance of the NMR spectrum, scalar coupling leads to the resonance being split into multiple lines (I+1 lines for coupling to I distinct nuclei), and the difference in frequency between these lines is the coupling constant (J) expressed in units of Hz.

The magnitude of the coupling constant gives important structural information. In particular, the magnitude of the coupling constant for nuclei separated by three chemical bonds is related by a trigonometric function to the dihedral angle between the two atoms.
(termed the Karplus equation). Numerous Karplus curves have been reported in the literature, and the magnitude and shape of these curves has been shown to be dependent on the electronegativities of substituents on the carbon or nitrogen atoms to which the atoms of interest are attached (Haasnoot et al., 1980). There is a dependence on the coupling constant between the NH and Hα proton ($^3J_{HN\alpha}$) and the backbone dihedral angle $\phi$, and the Karplus curve demonstrating this relationship is shown in Figure 9a. From this curve, we see that the coupling constant is near the minimum when the backbone adopts an $\alpha$-helical conformation, and is near the maximum when the backbone adopts an extended or $\beta$-strand conformation. Karplus curves have also been parameterized for the coupling constant between the Hα and two Hβ protons and the dihedral angle $\chi_1$, and this is shown in Figure 9b. Typically, side chain dihedral angles occupy one or a mixture of the three staggered rotamers. These rotamers about the $\chi_1$ dihedral angle can be distinguished by the presence of two small, or one small and one large, coupling constant involving the Hα and each of the Hβ protons (see Chapter 3).

By measurement of the coupling constant, one can determine if the dihedral angle occupies one of the staggered rotamers, and if it does, one can determine the value for the dihedral angle of interest and use this information as a restraint in NMR structure calculations. In the PLCC SH2 domain/pY1021 complex, Dr. Steve Pascal used a mixture of coupling constant and intra- and sequential NOEs to determine $\chi_1$ restraints (Pascal et al., 1994), while Lewis Kay measured $J_{C\alpha C\delta}$ to determine $\chi_2$ dihedral angle restraints for Leu and Ile residues (Kay et al., 1996). I also used coupling constant information to obtain $\chi_1$ restraints for a number of pY1021 residues, as well as to obtain information about the conformation of proline rings in the pY1021 peptide when bound to the PLCC SH2 domain.
Figure 9 -- Coupling Constants in Proteins. The following figures are taken from Sutcliffe (1983). A. Karplus Curve for the three-bond coupling constant $J_{\text{HNH}}$ as well as the equation for this curve. Values for this coupling constant are shown to be maximum when the backbone dihedral angle adopts a $\beta$ conformation, and minimum when the backbone dihedral angle adopts an $\alpha$ conformation. B. The Karplus equations for the three-bond coupling constant involving the $H_\alpha$ proton and the two $H_\beta$ protons, and the coupling constants (and short-range NOEs) which are observed among the three staggered rotamers which occur about the dihedral angle $\chi^1$. 
\[ ^3J_{H_{\alpha}H_{\alpha}} = 6.4 \cos^2(\phi - 60^\circ) - 1.4 \cos(\phi - 60^\circ) + 1.9 \]

B

\[ ^3J_{H_{\alpha}H_{\alpha}} = 9.5 \cos^2(\chi_1) - 1.6 \cos(\chi_1) + 1.8 \]
\[ ^3J_{H_{\alpha}H_{\alpha}} = 9.5 \cos^2(\chi_1 - 120^\circ) - 1.6 \cos(\chi_1 - 120^\circ) + 1.8 \]

<table>
<thead>
<tr>
<th>Conformation</th>
<th>( g^- )</th>
<th>( t )</th>
<th>( g^+ )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi_1 )</td>
<td>60°</td>
<td>180°</td>
<td>-60°</td>
</tr>
</tbody>
</table>

\[ ^3J_{H_{\alpha}H_{\beta2}} \text{ (Hz)} \]
\[ ^3J_{H_{\alpha}C=O} \text{ (Hz)} \]

- NOE(\( \alpha-\beta2 \))
  - \( \leq 5 \)
  - Strong
  - Weak
- NOE(\( \alpha-\beta3 \))
  - \( > 10 \)
  - Strong
- NOE(\( C=O-\beta2 \))
  - Weak
  - Strong-Medium
- NOE(\( C=O-\beta3 \))
  - Strong-Medium
  - Strong
  - Weak
4. Nuclear Overhauser Effect (NOE)

Through-space interactions occur by means of relaxation processes in which the magnetization of one spin is exchanged with another close in space via fluctuating magnetic fields created by one proton as it tumbles in solution. This process of magnetization transfer is referred to as dipolar coupling and is the basis for the NOE or Nuclear Overhauser Effect. The rate of magnetization transfer, or the cross-relaxation rate $\sigma_{ij}$ between two spins $i$ and $j$, is related to the distance between the two protons by the relation

$$\frac{r_{ij}}{r_{kl}} = \left(\frac{\sigma_{ij}}{\sigma_{kl}}\right)^{1/6}$$

where $r_{ij}$ and $r_{kl}$ represents the distance between the protons $i$ and $j$, and $k$ and $l$ respectively. This relation depends on the two vectors $i-j$ and $k-l$ having the same rate of tumbling, or correlation time, in solution. $\sigma_{ij}$ in principle is related to the magnitude of the NOE, however this relation is complicated by leakage of energy to the external environment (or lattice) and additional transfers to third and fourth spins. As the time for NOE buildup (or the mixing time) is increased, the leakage of energy to the lattice and transfers to other spins become important such that the $r^{-6}$ relationship breaks down. This process of secondary and even tertiary NOEs occurring following the first NOE interaction is referred to as spin diffusion, i.e. the magnetization diffuses or transfers through the whole molecule. However, at intermediate mixing times, the NOE intensity can be used semi-quantitatively to obtain distance ranges. This is done by classifying the intensity of the various NOEs as strong, medium or weak, and defining distance ranges for each category. For example, a strong NOE may represent a distance of $< 2.7$ Å. Weak NOEs define distances which are generally not greater than 5-6 Å. Of course, NOEs cannot define distances less than the van der Waals distance between two protons, which is defined as 1.8 Å (Nilges et al., 1988). With the use of NOEs in this semi-quantitative manner, the NOE has become the basis for structure determination via solution NMR, and can be used
for sequential assignment, determination of secondary structure, and finally, tertiary structure determination.

As protons between sequential amino acids are close in space, often strong NOEs are observed between them. Which protons are closest, however, varies depending on the values of the backbone torsion angles \( \phi \) and \( \varphi \), and thus the secondary structure of the molecule (see Figure 10). In \( \alpha \)-helices, the NH of residue \( i \) and the NH of residue \( i+1 \) are close in space (written as \( dNH_i;NH_{i+1} \)), while in \( \beta \)-strands, the strongest sequential NOE is \( dC\alpha_iNH_{i+1} \). A number of longer-range NOEs can also help define the secondary structure. For example, \( (i, i+3) \) and \( (i, i+4) \) NOEs can be observed in \( \alpha \)-helices, while long-range NOEs between pairs of backbone protons can be observed between amino acids which make up individual strands of \( \beta \)-sheets.

Short-range NOEs can be used to help define the overall topology of the molecule in terms of its fold; however, the complete structure can only be obtained by assignment of many short, medium and long-range NOEs, and the number of NOEs assigned has a profound effect on the precision of the structure. As shown by Clore & Gronenborn (1993) (see Figure 11), structure calculations involving few (5-10) NOEs per residue yield structures of low precision, in which an overall fold can be observed with a root mean squared deviation (RMSD) from an average structure of 1.5 \( \text{Å} \) between backbone atoms. Structure calculations involving more (10-15) NOEs (or other restraints) yield structures of moderate precision with RMSDs of 0.7 \( \text{Å} \) and 0.9 \( \text{Å} \) for backbone and all atoms, respectively. Finally, structure calculations involving >15 NOEs or other restraints give a high precision structure with RMSDs of 0.4 and 0.9 \( \text{Å} \) for backbone and all atoms, respectively, comparable to that of a 2.0-2.5 \( \text{Å} \) crystal structure. Thus, determination of a high-precision structure of a protein of 100 amino acids would require at least 1500 NOEs or other restraints (such as restraints on the dihedral angles \( \phi \) and \( \chi_1 \) through measurement of coupling constants).
Figure 10 -- NOEs and Secondary Structures in Proteins. A. NOEs which are observed in an anti-parallel β-sheet both in terms of short-range (Hα\textsubscript{i}/NH\textsubscript{i+1}) and long-range (across the β-strand) NOEs which define the β-sheet topology (taken from Sutcliffe, 1983). B. A table of short and medium-range distances found in helices, strands and turns. (Figure taken from Wüthrich, 1986.)
For the turns, the first of two numbers applies to the distance between residues 2 and 3, the second to that between residues 3 and 4 (Fig. 7.12). The range indicated for \( d_{\text{helix}(i,i+3)} \) corresponds to the distances adopted if \( \psi_i \) is varied between \(-180\) and \(180^\circ\).

The ranges given correspond to the distances adopted by a \( \beta \)-methine proton if \( \chi^1 \) is varied between \(-180\) and \(180^\circ\).
Figure 11 -- NOEs and Structure Precision in Proteins. As one increases the number of restraints (NOE and dihedral), the precision of the structure calculation increases, and the RMSD (root mean squared deviation) from the mean structure decreases. This is shown in the following figure taken from Clore & Gronenborn (1993).
1st Generation
- 7 restraints per residue
  rmsd: 1.5 Å for backbone atoms
  2.0 Å for all atoms
  example: purothionin

2nd Generation
- 10 restraints per residue
  rmsd: 0.9 Å for backbone atoms
  1.2 Å for all atoms
  example: BDS-I

3rd Generation
- 13 restraints per residue
  rmsd: 0.7 Å for backbone atoms
  0.9 Å for all atoms
  example: BDS-I

4th Generation
- 16 restraints per residue
  rmsd: 0.4 Å for backbone atoms
  0.9 Å for all atoms,
  ≤ 0.3 Å for ordered side chains
  example: Interleukin-8
B. Structure Calculations

Following assignment of NOEs and generation of dihedral restraints, the NOE and dihedral restraint information must be input into algorithms which calculate structures consistent with the input restraints. The method used in determination of the structure of the PLCC SH2 domain/pY1021 complex is that of distance geometry/simulated annealing, as described by Nilges et al. (1988) using the program XPLOR (Brunger, 1992). The first step of this procedure involves generating starting structures using distance geometry sub-embedding. This technique creates initial structures consistent with the input distance bounds using the metric matrix method of Crippen & Havel (1988), using a subset of atoms including the peptide backbone. Following sub-embedding, the remaining atoms are added in an extended conformation, and the resulting structure is subject to simulated annealing. Simulated annealing involves molecular dynamics at high temperature, followed by a slow cooling process and a final energy minimization. The force field, described in Nilges et al. (1988) includes energy terms for distance restraints which satisfy the input NOE-derived distance limits as follows:

\[
E_{\text{NOE}} = \begin{cases} 
  k_{\text{NOE}}(r_{\text{min}} - r)^2 & \text{if } r < r_{\text{min}} \\
  k_{\text{NOE}}(r - r_{\text{max}})^2 & \text{if } r > r_{\text{max}} \\
  0 & \text{if } r_{\text{min}} < r < r_{\text{max}} 
\end{cases}
\]

where \(r_{\text{min}}\) and \(r_{\text{max}}\) define the minimum and maximum distance between the two protons, \(r\) is the distance in the structure between these two protons, and \(k_{\text{NOE}}\) is an NOE force constant. Similar energy terms exist for van der Waals radii, bond lengths, angles and improper torsion angles to ensure proper geometry of atoms and ensure no atomic overlap, and the total energy is the sum of these energy terms:
In order to ensure that the total conformational space available as specified by the NOEs is sampled, these structure calculations are repeated several times, and the precision of the ensemble is measured by the root mean square deviation (RMSD) of the ensemble to the average structure of this ensemble.

C. NMR Spectra: From One to Four Dimensions

The most simple NMR spectrum is the one-dimensional NMR spectrum, shown schematically for the molecule in Figure 12A, which plots intensity as a function of frequency (in ppm). Every proton absorbs at a particular frequency and therefore gives rise to a resonance at this frequency which may or may not be further split into a doublet or multiplet due to scalar coupling. In order to perform structural determination, one must obtain as much NOE information as one can, and thus one must be able to unambiguously assign as many resonances as possible to particular protons on the molecule. However, even with small peptides, many of the resonances overlap, which makes unambiguous assignment of NOEs difficult.

To alleviate spectral overlap, two dimensional NMR (2D) experiments were developed. The principle of these experiments is that if two NMR-active nuclei interact either through their scalar (through-bond) or dipolar (through-space) coupling, a crosspeak is observed in the 2D experiment at frequency \(\omega_A,\omega_B\) and \(\omega_B,\omega_A\), where \(\omega_A\) and \(\omega_B\) are the frequencies of nuclei \(H_A\) and \(H_B\), respectively (see Figure 12b and c). If these two nuclei are scalar coupled or within a scalar coupled spin system, the 2D experiments used are the COSY (COrrelated Spectroscopy) and HOHAHA (HOMonuclear HArtnann HAhn), while if the nuclei are dipolar coupled, the experiment used is the NOESY (NOE Speroscopy). These experiments were used to solve the first three dimensional
Figure 12 -- NMR Spectra: One and Two Dimensions. A. Schematic of a molecule containing three protons (a, b and c), of which protons a and b are scalar coupled and a and b, and b and c are close in space. B. A 1D spectrum for this molecule, in which protons $H_a$, $H_b$ and $H_c$ resonate at frequencies $\omega_A$, $\omega_B$, and $\omega_C$, respectively. 2D spectra of this molecule can be performed (part C), and crosspeaks will be observed depending on if the protons are scalar coupled (in the COSY experiment) or if they are close in space (in the NOESY experiment).
structures of a number of small (< 100 amino acid) proteins (Wüthrich, 1986).

Though the 2D experiments used previously for structure determination are homonuclear experiments (i.e. measuring proton chemical shifts in both dimensions), heteronuclear 2D experiments are also quite useful. In particular one experiment which is quite important is the $^{15}$N/$^1$H-HSQC (Heteronuclear Single Quantum Coherence) experiment, which correlates the frequency of a proton to its attached (i.e. separated by one chemical bond) $^{15}$N nucleus. This experiment yields NMR information concerning protons and $^{15}$N nuclei of NH or NH$_2$ groups, of which the majority in protein samples will be from the peptide backbone. These experiments can be rapidly (< 30 minutes) recorded on $^{15}$N-labeled samples and can be used to assess global properties of the protein, such as whether it is folded or aggregated. If an HSQC spectrum of a protein sample shows a narrow range of frequencies, this indicates that the protein is unfolded or in a random coil state (Figure 13A), as opposed to the spectrum in Figure 13B which shows a range of frequencies covering several ppm, indicating that the protein is folded. In addition, if the resonances are broadened, this is indicative of higher-molecular weight oligomers, or aggregation.

Structure determination using homonuclear NMR techniques is very difficult for proteins with molecular weights greater than 10K for two reasons, large number of NOEs leading to signal overlap in 2D NOESY spectra and short $T_2$ values. Signal overlap can be alleviated by expansion of a 2D spectrum into a 3D or 4D spectrum, in which the third and fourth dimensions provide the chemical shifts of $^{13}$C or $^{15}$N nuclei to which the protons are attached (see Figure 14). Decreasing $T_2$ values leads to a loss of signal over the course of the pulse sequence; this can be minimized by performing magnetization transfers via heteronuclear scalar couplings, which are significantly larger than the homonuclear couplings (e.g. 140 Hz vs. 10 Hz), and thus shortening the time of the pulse sequence. These 3D experiments have found wide application to the determination of solution structures of macromolecules and macromolecular complexes with molecular weights
Figure 13 -- $^{15}\text{N}-^{1}\text{H}$ HSQC Spectra of Folded and Unfolded Proteins. Two HSQC spectra of the drk N-terminal SH3 domain (taken from Zhang & Forman-Kay, 1995), the first (A) of an unfolded sample in 2 M guanidinium chloride and the second (B) of a sample in high salt in order to stabilize the folded species.
Figure 14 -- NMR Spectra: From Two to Four Dimensions. Figure taken from Clore & Gronenborn (1993), showing how a 2D NMR spectrum can be spread into three or even four dimensions.
greater than 25K; beyond this molecular weight, $T_2$ relaxation is still too fast and deuteration of the protein must be employed to increase the $T_2$ of the macromolecule under study.

IV. Summary

In this chapter, I discussed the biology and structural biology of growth factor receptors, SH2 domains, and the enzyme PLC-γ. The SH2 domains of PLC-γ play an important role in signal transduction by linking the activation and autophosphorylation of growth factor receptors such as PDGFR and EGFR to the hydrolysis of the phospholipid PIP$_2$ to the second messengers DAG and IP$_3$. SH2 domains show sequence specificity in their binding to sites of Tyr phosphorylation, and it has been shown in vitro and in vivo that this sequence specificity manifests itself in PLC-γ recognizing sequences about Tyr 992 and Tyr 1021 of the EGFR and PDGFR, respectively. The peptide specificities of the two SH2 domains of PLC-γ have been investigated by degenerate phosphopeptide libraries, and differences between the two SH2 domains were observed. Based on these differences, PLC-γ is thought to bind to the high-affinity sites of EGFR and PDGFR through its C-terminal (PLCC) SH2 domain. Multinuclear NMR methods and spin filtering techniques allow us to determine the solution structure of SH2 domain/phosphopeptide complexes when the SH2 domain is uniformly labeled with $^{15}$N and $^{13}$C nuclei. In addition, we can learn about rates of peptide binding and pK$_a$'s of various titratable groups using solution NMR techniques. However, we must first produce the SH2 domain in high yield and purity, and the resulting protein must meet a number of conditions, i.e. soluble, folded, and not aggregated. In the following chapter, I discuss the expression and purification of the PLCC SH2 domain, as well as show its suitability for further NMR studies. In addition, some preliminary NMR experiments are described including a backbone assignment and demonstration of the kinetics of binding to two
phosphopeptides. In later chapters I employ solution NMR techniques to discern the molecular details of binding and specificity of this SH2 domain. SH2 domain specificity is a key event in the early stages of signal transduction, as binding of different SH2 domain-containing proteins dictates the extent of downstream pathways in signal transduction which will be activated following growth factor stimulation. SH2 domains contain similar tertiary folds, however they contain a wide variety of binding interfaces, and by understanding the binding interface of this SH2 domain, we can learn about recognition in SH2 domains specifically and in signal transduction process in general.
Chapter 2: Purification and Preliminary Characterization of the PLCC SH2 Domain

I. Introduction

When initiating structure determination of a biological macromolecule by NMR, there are a number of requirements which must be met. First, the biological macromolecule must be purified to homogeneity and the quantity of the molecule available must be sufficiently large, on the order of a μmole. Recent developments to facilitate the study of proteins from 10K to ~30K using 3D and 4D-heteronuclear NMR experiments (Clore & Gronenborn, 1991) involve incorporation of $^{13}$C and/or $^{15}$N nuclei into the proteins. At present, over-expression of the cloned gene of interest in *E. coli* and growth of these *E. coli* cells in minimal media containing $^{15}$N-labeled ammonium chloride as the sole nitrogen source and $^{13}$C-glucose as the sole carbon source is the most economical method to achieve uniform $^{13}$C- and $^{15}$N-labeling. $^{13}$C-labeled carbon sources are costly: $^{13}$C-glucose is purchased at approximately $400 \text{ US per gram}$ of $^{13}$C-labeled glucose, and often at least three grams of glucose are required per liter of minimal media. Thus the need for a very good expression system becomes quite clear. Since approximately $1/2 \mu$ mole of protein is needed per sample and often at least two samples are required, a guideline for cost-effective protein expression for multinuclear solution NMR studies is the ability to purify 1 μmole of protein to homogeneity per 2 liters of bacterial culture. In this section, the purification strategy and yield of the PLCC SH2 domain utilized for NMR studies is presented.

Two other requirements should be met in order to be able to perform a structural determination of a protein using heteronuclear NMR techniques. The macromolecule should be monomeric or form low molecular weight oligomers (< 50K) in solution at NMR concentrations. The linewidth of the NMR resonances increases with the tumbling time of the molecule in solution, and thus with the molecular weight of the protein in solution.
Increased linewidth leads to poor sensitivity. In the case of the PLCC SH2 domain (MW ~ 12 kDa), monomeric or dimeric species are optimal for NMR study and trimeric or tetrameric species would be difficult to study using present methods. In this section I describe experiments indicating that the PLCC SH2 domain does not aggregate to any significant extent.

The protein should also be well-structured in solution and not be disordered as a random-coil or "molten globule". This may seem somewhat intuitive, since it is more difficult using present methods to describe such an ensemble of structures. However, in addition, resonances of random coil peptides are often highly degenerate resulting in signal overlap. An NMR spectrum indicating that the PLCC SH2 domain used in this study adopts a well-ordered conformation in solution is also shown in this section.

In addition to results demonstrating that the PLCC SH2 domains has levels of expression and solution properties which make it tractable for NMR study, a few preliminary NMR experiments to characterize both the free PLCC SH2 domain and its complex with the pY1021 phosphopeptide are described. Initially, assignment of NMR resonances must be performed in NMR structure determination, and two experiments used for assignment of backbone resonances of the PLCC SH2 domain are described. I also describe experiments used to study the kinetics of binding of the PLCC SH2 domain to two high-affinity binding phosphopeptides using NMR methods, specifically lineshape analysis.
II. Results

A. Purification

1. Sub-cloning and Conditions for Protein Expression

Three constructs of the PLCC SH2 domain were made, and the amino acid sequences of the PLCC SH2 domain resulting from each of these constructs are shown in Figure 1. The first construct of the PLCC SH2 domain was made by Dr. Gerry Gish in the lab of Dr. Tony Pawson. In this construct, a DNA segment coding for residues 663-759 of bovine Phospholipase C-γ was amplified using PCR and subcloned into pGEX-KT (Hakes & Dixon, 1992) using the vector's Bam H1 and Hind III restriction sites. In this cloning vector, the gene is fused to the coding sequence for the *S. japonicum* Glutathione-S-Transferase (GST). In addition, a thrombin cleavage site is placed between the GST sequence and the cloned gene, and directly 3' from the thrombin site is a linker region coding for five glycine residues used to make the thrombin cleavage site more accessible. The promoter for the fusion protein contains the lac operon, so its production is inducible using 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). The first sequence shown in Figure 1 is from the protein that would result following cleavage by thrombin and removal of the GST fragment.

The other two constructs were made by myself and involved subcloning the PLCC SH2 domain into the pET-11d expression vector (Dubendorff & Studier, 1991). In this system, the subcloned gene is not expressed as a fusion protein. Instead, the DNA fragment of interest, following amplification using PCR, is cloned directly into the plasmid's Nco I and Bam H1 sites, the Nco I site containing the codon for the initiator Met, and the stop codon being 3' of the Bam H1 site. Upstream of the cloning site is a strong promoter for T7 polymerase-expressed genes. The plasmid is transformed into
Figure 1 -- Primary Sequence of PLCC SH2 Domain Constructs. Sequence of the protein products of the PLCC SH2 domain constructs used in this study. The first (seq 1) was generated from the pGEX-KT expression system, with the first two amino acids (GS) derived from the thrombin cleavage site. The second (seq 2) and third (seq 3) constructs result from cloning into the pET-11d expression vector. The figure was generated using the program SeqVu (Gardner et al., 1995). The numbering scheme for all three clones used is the numbering for seq 1.
BL21(DE3) cells, which contain a lysogen (DE3) containing the T7 RNA polymerase gene. The T7 RNA polymerase is in turn under the transcriptional control of the lac promoter so that its expression is induced by IPTG.

All three constructs express high levels of protein after a few hours of induction by IPTG in both rich media, LB, TB (Sambrook et al., 1989) or Celtone (Martek Ltd.), and minimal M9 media (Sambrook et al., 1989). In Figure 2a is shown the expression of the GST-fusion protein in LB, Celtone and M9 media after two hours induction. In Figure 2b is shown induction of one of the two pET-11d constructs just before and after four hours of induction in M9 minimal media. Generally, the level of expression of the cloned gene is not detectable above the level of bacterial proteins of similar molecular weight prior to induction. However, following induction, the level of expression is considerably higher than any other protein expressed in E. coli.

Figure 2c shows the fraction of expression of soluble protein for the longer of the two constructs (seq3, see Figure 1). We observed that < 50% of expressed protein was soluble at 37° C while nearly all expressed protein was soluble at 30° C. For all PLCC SH2 domain constructs, a large percentage of expressed protein was insoluble when the induction was performed at 37° C, but nearly 100% was soluble when induction was performed at 30° C or lower. Thus, bacteria were grown at 37° C until they reached an OD600 of approximately 0.7 and then were cooled either to 30° C and induced for four hours or to 20° C and induced overnight. Bacterial cultures were grown in a 2 L fermentor (BioFlo IIIC, New Brunswick Scientific).

In the following sections I describe the purification procedures which I developed with the assistance of Dr. Gerry Gish, a research associate in the lab of Tony Pawson, for isolation of the PLCC SH2 domain from both the pGEX and pET-11d expression systems. The identity of purified SH2 domain was confirmed in all three constructs by N-terminal sequencing of the purified protein for 5-15 cycles.
Figure 2 -- Induction of Expression of PLCC SH2 Domain Constructs

a. Induction of expression of the GST-SH2 domain fusion protein in three different media, two rich media, LB and the algal lysate Celtone (Martek), and minimal (M9) media. 50 ml of media were inoculated with GST-SH2 domain-expressing bacteria, incubated at 37° C until an OD of 1.0, 1 ml sample of culture was taken (t=0), cultures were induced with 100 μl 125 mg/ml IPTG for 2 hours at 30° C, and another 1 ml sample was taken from each culture (t=2). 1 ml samples were spun down, resuspended in 200 μl loading dye containing (Tris, 1% SDS, β-mercaptoethanol) and loaded on a 15% SDS-PAGE gel.

b. Induction of expression of PLCC SH2 domain (seq 3) before and after four hours of induction at 30° C of a culture containing 3 ml M9 minimal media. Samples were run on a 10% Tricine gel.

c. Expression of the PLCC SH2 domain (seq 3) at two different temperatures, demonstrating considerable improvement in the fraction of soluble protein when induction was performed at 30° C versus 37° C. 50 ml LB cultures were inoculated with an overnight culture of E. coli expressing the PLCC SH2 domain (seq 3), and incubated at 37° C to an OD600 of 1.5, cooled to the appropriated temperature and induced with 100 μl 125 mg/ml IPTG, and induced for four hours. Following induction, cells were pelleted, resuspended in 2 ml lysis buffer (as described previously), sonicated for 2 minutes, centrifuged at 6 K for 15 minutes, the pellet was resuspended in 2 ml of lysis buffer, sonicated for another 2 minutes and centrifuged at 6 K for 15 minutes. Samples were run on a 15% Tricine gel. Son 1 and son 2 are the soluble fraction following the first and second two minute sonication periods of the E. coli cells and pellet is the insoluble fraction following this treatment.
2. GST-PLCC SH2 Domain Purification

The procedure used for large-scale purification of the PLCC SH2 domain using the GST system is shown in Figure 3. Lysis of cells was tested using enzymatic approaches with lysozyme and DNase. However, sonication was found to be more efficient and therefore was used for cell lysis in all expression systems. The buffer used for cell lysis in all cases contained 20 mM sodium phosphate, pH 7.3, 150 mM NaCl, 1% Triton X-100, 5 mM dithiothreitol, 1 mM EDTA and 1.0 mM benzamidine. Typically, the cell pellet following centrifugation of 2 L of bacterial culture was resuspended in 20 ml of this lysis buffer. The usual procedure for cell lysis was two minutes of pulsed sonication, followed by centrifugation, collection of the supernatant and resuspension of the pellet in the same volume of lysis buffer. This procedure was repeated an additional two or three times, and the three or four supernatants were pooled prior to further steps in chromatography. Cell lysis and all other steps in purification were performed on ice or at 4°C unless otherwise stated (e.g. thrombin digestion).

Following cell harvesting and lysis, the supernatant from sonication of E. coli cells expressing the GST-fusion protein was mixed with glutathione sepharose beads (500 mg/2L bacterial culture) for at least 30 minutes at 4°C. The beads were then rinsed four times with 10 ml 100 mM ammonium bicarbonate, pH 8.0, to remove other proteins in the cell extract, followed by 40 ml of buffer containing 100 mM ammonium bicarbonate and 10 mM glutathione, pH 8.0, to elute the GST-SH2 domain fusion protein from the beads. The above was done in batch mode. At this point, we obtained GST-SH2 domain fusion protein purified to near homogeneity. However, removal of the GST-SH2 domain from the beads was not efficient and typically there was a considerable amount of fusion protein still bound to the glutathione sepharose beads. I estimated this to be 1/3 to 1/4 of the total quantity of soluble fusion protein by densitometry.
Figure 3 -- Purification Scheme Used for the PLCC SH2 Domain Cloned into the pGEX-KT Vector.
Sonication

Spin 8 rpm, 20 minutes.
Collect Supernatant.
Repeat 2-3 times.

Glutathione Sepharose

Wash with 10 mM Glutathione

Thrombin digestion

DE-52 (anion exchange) column

SH2 domain in the flow-through

CM-52 (cation exchange) column

Elute with a gradient of 0 to 0.4 M NaCl

> 99% pure SH2 domain
The GST-fusion protein was then dialyzed against 50 mM sodium phosphate, pH 8.0, to remove glutathione, and then digested with thrombin. For 1 mg of GST-SH2 domain fusion protein, a four hour digestion at 37° C using 1 ug of thrombin was sufficient for digestion to >95% completion. Cleavage conditions varied taking into account that lowering of the temperature by 10° C doubled the time for digestion and doubling the amount of thrombin reduced the digestion time two-fold. Additional purification steps to remove GST (approximate MW of 29 kDa) and thrombin (~ 36 kDa) are summarized in Figure 3. Initially, the thrombin digest was added to a column containing 5 ml of swelled DE-52 (Whatman) beads. As shown in Figure 4a, because of the positive charge of the resin and the similar positive charge of the SH2 domain, the PLCC SH2 domain was bound weakly to the column while the GST bound quite tightly and could be eluted with 250 mM NaCl. The flow-through from the DE-52 column containing the SH2 domain was dialyzed into 50 mM HEPES, pH 7.5, and then loaded on to a CM-52 (Whatman) column. As shown in Figure 4b, the SH2 domain bound reasonably tightly to the column and could be eluted using a NaCl gradient. Initially, I used a step gradient with increments of 50 mM NaCl, however in later isolations, a smooth gradient of between 0-400 mM NaCl was used. As shown in Figure 4b, the SH2 domain eluted from the column over a number of fractions. These fractions correspond to elution at 145-175 mM NaCl, a few fractions later than thrombin. As can be seen from Figure 4b, purity of the PLCC SH2 domain following this column was very high. Fractions from this column were then reduced using 5 to 10-fold molar excess of dithiothreitol, dialyzed into 100 mM sodium phosphate, pH 6.0-6.3, concentrated to a volume of 500 µl and placed into an argon-purged 5 mm NMR tube.

While the purity of samples obtained by this technique was high and the protein remained soluble at NMR concentrations, yields of the SH2 domain were quite low. Typically, only one NMR sample per 2 L of E. coli. culture was produced. Since induction was high, the low yield must have resulted from the purification procedure. One step which led to losses was the inefficient elution from the glutathione sepharose beads, as
Figure 4 -- Purification of PLCC SH2 Domain from the GST-PLCC SH2 Domain Fusion Protein. 

a. Fractions from a DE-52 column loaded with a thrombin digest of the GST-PLCC SH2 domain fusion protein, containing two major bands of approximately 12 kDa (SH2) and 29 kDa (GST) respectively. The SH2 domain binds poorly to the column, while the GST binds tightly and can be eluted with 250 ml NaCl.

b. Fractions from a CM-52 column using a NaCl gradient from 0-400 mM. Two peaks on the UV trace monitored at OD280 were observed, the first (fractions 71-72) from a contaminating protein of similar molecular weight (contam. a), the second the PLCC SH2 domain (fractions 78-85). The eluted SH2 domain is of high purity in these fractions, although there is a weak-intensity contaminating band in fractions 83-84 (contam. b).
described earlier. Since low yields were a concern using this approach, we decided to clone PLCC SH2 domain DNA fragments into pET expression plasmids. pET expression vectors are optimized for transcription by T7 polymerase; chain elongation by T7 RNA polymerase has been found to be about five-fold faster than E. coli RNA polymerase (Studier et al., 1990).

3. Purification of pET-derived PLCC SH2 Domain

The PCR products made using the previous (pGEX-KT) construct as template were cloned into the pET-11d expression vector. This plasmid has a Nco I and a Bam H1 cloning site: the Nco I site (sequence CCATGG) can be used for the initiator Met (codon ATG) when the codon for the next amino acid begins with the base G. In the first clone, residues 662-759 of bovine phospholipase C-\(\gamma\) were used with an Ala residue (codon GCG) added between the initiator Met and the PLC-\(\gamma\) sequences. As stated previously, this particular plasmid contains no coding sequences for Ni\(^{2+}\)-binding (the so-called "His tag") or any other sequences which may be useful for affinity purification, so that proteins produced using this expression system must be purified based on the properties of the expressed protein itself, such as its hydrophobicity or charge characteristics. The amino acid sequence of the SH2 domain produced from this construct is the second sequence in the alignment shown in Figure 1. After expression and purification of this protein as described below, it was concentrated for NMR experiments. However, at these higher protein concentrations this PLCC SH2 domain protein was found to precipitate. Therefore a second construct was made which contained the identical protein sequence to the GST clone following thrombin cleavage (found previously to be more soluble), except for the addition of an initiator Met. This was possible because the first amino acid in the GST clone (a Gly) can use a codon whose first base is a "G". The sequence of this clone is the third sequence in Figure 1. These two pET constructs yield protein with very similar
chromatographic properties, and the purification scheme described below was used for both forms of the PLCC SH2 domain. This method of purification is outlined in Figure 5.

Following growth of *E. coli* cells, harvesting and cell lysis (as described previously), the bacterial supernatant was loaded onto a phosphocellulose (Whatman P10) column which had been prepared according to the instructions of the supplier. The phosphocellulose column (using 5-6 ml dry resin) was equilibrated into the lysis buffer described previously (20 mM sodium phosphate, pH 7.3, 150 mM NaCl, 1% Triton X-100, 5 mM dithiothreitol), except lacking 1% Triton (this buffer will be referred to as TMPBS). The bacterial supernatants following sonication were pooled and passed through the column, followed by washing of the column with an additional 5-10 column volumes of TMPBS. In Figure 6a, one can observe that the vast majority of the SH2 domain was bound to the phosphocellulose column and was not removed by washing the column with TMPBS. However, many bacterial proteins were not bound by the phosphocellulose resin. Following washing of the column with TMPBS, protein was eluted from the column by use of a 200 ml gradient of 150 mM to 1 M NaCl gradient. Protein elution was monitored by UV absorbance at 280 nm, and peaks were analyzed by SDS-PAGE. As can be seen in Figure 6b, the SH2 domain eluted over a large range of salt concentrations, thus the NaCl gradient was critical and most of the purification achieved with this step was due to the tight binding of the PLCC SH2 domain to the phosphocellulose resin while most *E. coli* proteins did not bind.

All fractions containing the SH2 domain eluted from the phosphocellulose column were pooled, the NaCl concentration was adjusted to 1.3 M, and the protein was concentrated to a volume of 20-25 ml and loaded onto a 5-ml phenyl sepharose column. In Figure 7a is shown the sample loaded on to the column, the flow-through, and fractions
Figure 5 -- Purification Scheme Used for PLCC SH2 Domains Expressed from pET-11d Vectors.
Sonication
Spin 8000 rpm, 20 Minutes.
Collect Supernatant.
Repeat 2-3 times.

Phosphocellulose Column (Cation Exchange/Affinity)
Elute with Gradient of 0.15 to 1.0 M NaCl

Phenyl Sepharose Column (Hydrophobic Interaction)
Elute with Gradient of 1.3 to 0.15 M NaCl

Mono S Column (Cation Exchange Chromatography)
Elute with Gradient of 0 to 1 M NaCl

> 99% pure SH2 Domain
Figure 6 -- Purification of the PLCC SH2 Domain Expressed from the pET Expression Vector 1. Phosphocellulose Chromatography. Results of phosphocellulose chromatography with the pET-expressed SH2 domain (seq 2). Similar results were obtained for seq 3. a. A crude bacterial extract (start) depleted of SH2 domain by passage through phosphocellulose resin (flow-through). The SH2 domain and a smaller subset of bacterial proteins binds to the phosphocellulose resin (bound).

b. Fractions eluted from the phosphocellulose column, using a NaCl gradient from 150 mM to 1 M. The SH2 domain starts to elute at ~300 mM. The eluted SH2 domain is considerably purified, but fractions still contain a number of contaminating proteins.
Figure 7 -- Purification of the PLCC SH2 Domain Expressed From the pET Expression Vector 2. Phenyl Sepharose and Mono S Chromatography. 

a. Results of phenyl sepharose chromatography. Fractions containing the PLCC SH2 domain (seq 3) from the phosphocellulose column (see Figure 6) were concentrated to ~40 ml, the NaCl was adjusted to 1.3 M, and the sample was loaded onto the column. One column volume was collected (see the lane marked 5 ml) and then the flow-through was collected (flow-through). Following this, 10 ml fractions were collected containing 20 mM sodium phosphate and 1.3 M (wash), 500 mM, 250 mM and 150 mM NaCl. Samples of all fractions as well as the starting bacterial extract and beads from the phenyl sepharose column at the end of the 150 mM NaCl wash are shown. b. Fractions around the major peak at OD280 from a MonoS (Pharmacia) column. Approximately 3-4 fractions are observed to contain the PLCC SH2 domain at high concentration (concentration of fractions 18-19 is > 10 mg/ml) purified to homogeneity.
eluted with 10 ml washes with TMPBS containing 1.3, 1.0, 0.5, 0.25 and 0.15 M NaCl, as well as the resin following elution. The PLCC SH2 domain bound very poorly to this column, the vast majority of the PLCC SH2 domain either coming off the column in the flow-through or the 1.3 M NaCl wash. However, a number of bacterial proteins were more tightly bound to the column, confirming the utility of this step in purification.

Finally, the flow-through and the 1.3 M NaCl washes were pooled and dialyzed overnight against a buffer containing 100 mM sodium phosphate, pH 7.3, 5 mM dithiothreitol, and 1 mM benzamidine. The sample was loaded onto a Mono-S HR 16/10 (Pharmacia) column and eluted with a 200 ml salt gradient containing 0-1 M sodium chloride. The progress of the column was monitored via UV absorption, and peaks on the UV trace were analyzed by SDS-PAGE analysis. Typically, the majority of the SH2 domain was eluted in a few fractions around 350 mM NaCl. In Figure 7b is shown fractions from a Mono S column eluate, showing the high concentration and degree of purity of the SH2 domain from this column: it was not uncommon to have fractions containing >1 mM protein in a highly pure state. Fractions from this column were reduced using 5 to 10-fold molar excess of dithiothreitol, followed by buffer exchange into 100 mM sodium phosphate, pH 6.0-6.3, concentrated to a volume of 500 μl and placed in an argon-purged 5 mm NMR tube.

Sample concentration could be measured in one of two ways. Firstly, UV absorption at 280 nm was used, with an extinction coefficient measured in the laboratory of Dr. Cyril Kay (ε= 0.85 OD_{280}/mg/ml for the PLCC SH2 domain, seq 3). Second, the Bradford dye-binding assay (BioRad) was used, with 20 μl of protein solution added to 1 ml of Bradford dye and monitored by absorbance at 595 nm (using 20 μl of buffer alone and 1 ml of Bradford dye as a reference). It was found that the OD_{595} from a Bradford assay containing 20 μl of protein sample could be converted to an OD_{280} by dividing by 1.15, so the following formula could be used to obtain the concentration (in mg/ml) from the Bradford assay,
where \( A \) is the OD595 from the Bradford assay from 20 ul of protein solution, \( B \) is the protein concentration in mg/ml and \( \varepsilon \) is the extinction coefficient for the PLCC SH2 domain (0.85 OD280/mg/ml). This could be divided by the molar mass of the PLCC SH2 domain (12.2 kDa) to obtain the concentration in mM.

**B. Aggregation State of the Protein**

In Figure 8 are results from a BioRad P30 gel filtration column. In part a, column calibration using a set of four protein standards of molecular weight 45 kDa (ovalbumin), 29 kDa (carbonic anhydrase), 18.5 kDa (myoglobin) and 12.5 kDa (cytochrome c) is shown. In part b, a thrombin digest of the GST-SH2 domain fusion protein to which 1 mg of myoglobin (molecular weight 18.5 kDa) was added, was loaded onto the same column. Results are shown in Figure 8b. The addition of myoglobin is useful for monitoring the progress of the column due to its brown color, as well as comparing the migration of the myoglobin between the two column runs. In the first column (Figure 8a) myoglobin eluted between fractions 16-18, while in the second column (Figure 8b) myoglobin eluted between fractions 15-17. GST and the PLCC SH2 domain, which have very similar molecular weights to carbonic anhydrase and cytochrome c, respectively, would be expected in column 2 to elute between fractions 14-16 and 21-23, respectively. GST, in fact eluted much earlier, with an apparent molecular weight greater than ovalbumin (45 kDa), supporting previous experimental evidence that GST occurs as a dimer in solution (Lim et al., 1994). The PLCC SH2 domain migrated more slowly than myoglobin, indicating that it was > 50% monomeric. However, the PLCC SH2 domain also migrated at least three fractions earlier than expected, indicative either of the SH2 domain adopting a
Figure 8 -- Aggregation State of the PLCC SH2 Domain. Results from a BioRad P30 gel filtration column. a. Column standardized using a mixture of the following protein standards: 5 mg/ml ovalbumin (oval, 45 kDa), 0.8 mg/ml carbonic anhydrase (CA, 29 kDa), 2 mg/ml myoglobin (myo, 18.5 kDa) and 2 mg/ml cytochrome c (cyt c, 12.5 kDa). 0.3 ml of this mixture was loaded on this column, and 2.7 ml fractions were collected. b. 1 mg of myoglobin was added to a GST-SH2 domain thrombin digest and loaded onto the same column. The SH2 domain (MW 12.2 kDa) elutes after myoglobin but migrates significantly faster than cytochrome c, as evident by comparing the migration of myoglobin in the two columns.
non-spherical shape or being in equilibrium between monomeric and dimeric states, with the monomeric state predominating. Light scattering experiments (data not shown) and subsequent relaxation studies and ultracentrifugation and equilibrium dialysis experiments (Farrow et al., 1994) established that the free PLCC SH2 domain occurs as a mixture of monomer and dimer in solution but strictly as a monomer when in complex with the high-affinity binding pY1021 phosphopeptide. In the free PLCC SH2 domain the proportion of monomer to dimer is concentration-dependent.

C. Initial NMR Spectra

In Figure 9 is shown an $^{15}$N,$^{1}$H HSQC (Zhang, 1994) spectrum of the free PLCC SH2 domain in 100 mM sodium phosphate, pH 6.0 with assignments for all NH resonances listed. Assignments were made by Dr. Julie Forman-Kay and myself using the 3D $^{15}$N-NOESY (Kay, 1989), $^{15}$N-TOCSY (Marion, 1989), HNCACB (Wittekind & Mueller, 1993) and CBCA(CO)NNH (Grzesiek & Bax, 1992) experiments as described below. Of the 99 expected NH resonances (excluding four Pro residues and the N-terminus), 98 resonances are observed yielding sharp crosspeaks indicative of rotational correlation times consistent with a monomeric molecular weight. In addition, most peaks are distinct and cover a wide $^{1}$H chemical shift range of nearly 4 ppm, indicative of a folded protein, as opposed to a much narrower chemical shift range indicative of a disordered protein or a molten globule (see Chapter 1).

The linewidth of the NMR resonance also contains important information. The linewidth is related to the rate of relaxation of magnetization in the direction orthogonal to the static magnetic field. This rate is referred to as T2 relaxation, and is related to $\nu$, or the linewidth at half-height by the relation

$$\nu = \frac{1}{\pi \times T_2}.$$
Figure 9 --$^1$H-$^{15}$N HSQC of the PLCC SH2 Domain. $^{15}$N/$^1$H HSQC spectrum of a 0.3 mM sample of the PLCC SH2 domain (seq 3) in 100 mM sodium phosphate, pH 6.0. Backbone NH-$^{15}$N crosspeaks are labeled according to assignments.
T2 is inversely affected by the tumbling time of the molecule, in that the slower the rate of rotation in solution, the smaller the T2 value and the larger the linewidth. The rate of tumbling is affected by many factors such as viscosity, temperature, and the molecular weight in solution. It is this dependence on molecular weight which places a severe restriction on the size of molecules that can be studied by NMR, because as the size increases, the lineshape becomes too broad to be useful. For this reason, monomeric species are more tractable for study than oligomeric ones.

D. Backbone Assignment

Using an 15N-labeled sample and 3D 15N-HOHAHA and 15N-NOESY experiments, Dr. Julie Forman-Kay performed an initial assignment of most of the protein resonances with the exception of loop regions from Gly 70 to Asn 71, from Asp 75 to Ser 76, and from Leu 89 to Tyr 97. To complete and confirm assignments, double-labeled samples were prepared and 3D HNCACB and CBCA(CO)NNH experiments were performed. In Figure 10 is shown strips from these two experiments in the region between amino acids Gly 32 to Asn 40 (containing the highly conserved FLVR sequence necessary for pTyr binding), demonstrating the quality of the data. Assignments in the region between amino acids Leu 89 and Tyr 97 were carried out with little difficulty despite problems of signal overlap. In fact, complete backbone assignment of the PLCC SH2 domain from amino acids Pro 3 to Ser 105 were made based on this data (see Table 1), confirming previous assignments made using the 3D 15N-NOESY and 15N-TOCSY experiments. Assignment was performed independently for both the free and the pY1021-complexed SH2 domain. The HNCACB and HBCBCA(CO)NNH experiments have proved very useful in backbone assignment of many proteins for the following reasons.
**Figure 10 -- Backbone Assignment of the PLCC SH2 Domain.** Strips from a. a CBCA(CO)NNH experiment and b. an HNCACB experiment, taken at backbone NH and $^{15}$N chemical shifts of residues Gly 32 to Asn 40 of the PLCC SH2 domain/pY1021 complex in 100 mM sodium phosphate, pH 6.0. Peaks representing correlations of the $^{13}$C$_{\alpha}$ and $^{13}$C$_{\beta}$ resonances of residue i-1 are labeled in a. and of residue i are labeled in b. Unrelated peaks from close NH and $^{15}$N shifts are marked with an X.
Table 1: NH and $^{15}$N Chemical Shifts for the Free and pY1021-Bound PLCC SH2 Domain

<table>
<thead>
<tr>
<th></th>
<th>Gly</th>
<th>δNH(f)</th>
<th>δ15N(f)</th>
<th>δNH(b)</th>
<th>δ15N(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ile</td>
<td>4</td>
<td>8.55</td>
<td>109.2</td>
<td>8.54</td>
<td>109.2</td>
</tr>
<tr>
<td>Glu</td>
<td>5</td>
<td>8.17</td>
<td>120.8</td>
<td>8.19</td>
<td>120.9</td>
</tr>
<tr>
<td>Ser</td>
<td>7</td>
<td>7.49</td>
<td>119.5</td>
<td>7.46</td>
<td>119.4</td>
</tr>
<tr>
<td>Lys</td>
<td>8</td>
<td>7.52</td>
<td>112.8</td>
<td>7.50</td>
<td>112.8</td>
</tr>
<tr>
<td>Glu</td>
<td>9</td>
<td>7.55</td>
<td>122.0</td>
<td>7.53</td>
<td>122.0</td>
</tr>
<tr>
<td>Glu</td>
<td>10</td>
<td>8.91</td>
<td>120.8</td>
<td>8.98</td>
<td>121.0</td>
</tr>
<tr>
<td>Trp</td>
<td>11</td>
<td>6.26</td>
<td>109.5</td>
<td>6.31</td>
<td>109.4</td>
</tr>
<tr>
<td>Tyr</td>
<td>12</td>
<td>7.67</td>
<td>123.4</td>
<td>7.65</td>
<td>123.5</td>
</tr>
<tr>
<td>His</td>
<td>13</td>
<td>8.60</td>
<td>128.2</td>
<td>8.58</td>
<td>128.2</td>
</tr>
<tr>
<td>Ala</td>
<td>14</td>
<td>8.25</td>
<td>125.2</td>
<td>8.28</td>
<td>125.2</td>
</tr>
<tr>
<td>Ser</td>
<td>15</td>
<td>8.30</td>
<td>111.8</td>
<td>8.29</td>
<td>111.8</td>
</tr>
<tr>
<td>Leu</td>
<td>16</td>
<td>8.05</td>
<td>126.8</td>
<td>8.07</td>
<td>128.9</td>
</tr>
<tr>
<td>Thr</td>
<td>17</td>
<td>8.75</td>
<td>117.2</td>
<td>8.74</td>
<td>117.1</td>
</tr>
<tr>
<td>Arg</td>
<td>18</td>
<td>9.15</td>
<td>122.6</td>
<td>9.15</td>
<td>122.6</td>
</tr>
<tr>
<td>Ala</td>
<td>19</td>
<td>8.36</td>
<td>119.5</td>
<td>7.76</td>
<td>119.6</td>
</tr>
<tr>
<td>Glu</td>
<td>20</td>
<td>7.78</td>
<td>118.2</td>
<td>7.79</td>
<td>118.0</td>
</tr>
<tr>
<td>Ala</td>
<td>21</td>
<td>8.64</td>
<td>122.7</td>
<td>8.63</td>
<td>122.7</td>
</tr>
<tr>
<td>Glu</td>
<td>22</td>
<td>8.49</td>
<td>116.5</td>
<td>8.48</td>
<td>116.5</td>
</tr>
<tr>
<td>His</td>
<td>23</td>
<td>7.96</td>
<td>117.3</td>
<td>7.97</td>
<td>117.4</td>
</tr>
<tr>
<td>Met</td>
<td>24</td>
<td>7.93</td>
<td>117.7</td>
<td>7.92</td>
<td>117.9</td>
</tr>
<tr>
<td>Leu</td>
<td>25</td>
<td>8.05</td>
<td>117.7</td>
<td>8.05</td>
<td>117.6</td>
</tr>
<tr>
<td>Met</td>
<td>26</td>
<td>8.06</td>
<td>118.7</td>
<td>8.08</td>
<td>118.8</td>
</tr>
<tr>
<td>Arg</td>
<td>27</td>
<td>7.11</td>
<td>116.4</td>
<td>7.10</td>
<td>116.5</td>
</tr>
<tr>
<td>Val</td>
<td>28</td>
<td>7.08</td>
<td>119.1</td>
<td>7.08</td>
<td>119.1</td>
</tr>
<tr>
<td>Arg</td>
<td>29</td>
<td>7.92</td>
<td>121.8</td>
<td>7.88</td>
<td>121.6</td>
</tr>
<tr>
<td>Glu</td>
<td>30</td>
<td>8.74</td>
<td>124.6</td>
<td>8.74</td>
<td>124.6</td>
</tr>
<tr>
<td>Gly</td>
<td>31</td>
<td>9.52</td>
<td>109.8</td>
<td>9.52</td>
<td>109.9</td>
</tr>
<tr>
<td>Ala</td>
<td>32</td>
<td>7.90</td>
<td>126.4</td>
<td>7.89</td>
<td>126.4</td>
</tr>
<tr>
<td>Phe</td>
<td>33</td>
<td>8.69</td>
<td>117.8</td>
<td>8.68</td>
<td>117.9</td>
</tr>
<tr>
<td>Phe</td>
<td>34</td>
<td>9.29</td>
<td>114.8</td>
<td>9.30</td>
<td>114.9</td>
</tr>
<tr>
<td>Val</td>
<td>35</td>
<td>9.58</td>
<td>120.8</td>
<td>9.58</td>
<td>120.8</td>
</tr>
<tr>
<td>Arg</td>
<td>36</td>
<td>9.54</td>
<td>122.8</td>
<td>7.99</td>
<td>122.9</td>
</tr>
<tr>
<td>Lys</td>
<td>37</td>
<td>8.05</td>
<td>124.9</td>
<td>7.99</td>
<td>125.1</td>
</tr>
<tr>
<td>Arg</td>
<td>38</td>
<td>9.07</td>
<td>127.3</td>
<td>8.08</td>
<td>126.2</td>
</tr>
<tr>
<td>Asn</td>
<td>39</td>
<td>8.54</td>
<td>118.3</td>
<td>8.53</td>
<td>118.5</td>
</tr>
<tr>
<td>Glu</td>
<td>40</td>
<td>7.37</td>
<td>119.8</td>
<td>7.38</td>
<td>119.9</td>
</tr>
<tr>
<td>Asn</td>
<td>41</td>
<td>7.66</td>
<td>109.3</td>
<td>8.80</td>
<td>116.0</td>
</tr>
<tr>
<td>Ser</td>
<td>42</td>
<td>7.66</td>
<td>109.3</td>
<td>7.64</td>
<td>109.4</td>
</tr>
<tr>
<td>Tyr</td>
<td>43</td>
<td>9.12</td>
<td>124.5</td>
<td>9.11</td>
<td>124.5</td>
</tr>
<tr>
<td>Ala</td>
<td>44</td>
<td>9.49</td>
<td>121.2</td>
<td>9.49</td>
<td>121.2</td>
</tr>
<tr>
<td>Ile</td>
<td>45</td>
<td>9.18</td>
<td>121.5</td>
<td>9.17</td>
<td>121.5</td>
</tr>
<tr>
<td>Ser</td>
<td>46</td>
<td>9.15</td>
<td>126.7</td>
<td>9.18</td>
<td>126.7</td>
</tr>
<tr>
<td>Phe</td>
<td>47</td>
<td>8.82</td>
<td>121.3</td>
<td>8.89</td>
<td>121.5</td>
</tr>
<tr>
<td>Arg</td>
<td>48</td>
<td>8.31</td>
<td>119.3</td>
<td>8.37</td>
<td>119.5</td>
</tr>
<tr>
<td>Ala</td>
<td>49</td>
<td>8.94</td>
<td>127.0</td>
<td>8.93</td>
<td>127.1</td>
</tr>
<tr>
<td>Glu</td>
<td>50</td>
<td>9.30</td>
<td>121.7</td>
<td>9.31</td>
<td>121.8</td>
</tr>
<tr>
<td>Gly</td>
<td>51</td>
<td>9.38</td>
<td>105.8</td>
<td>9.39</td>
<td>105.8</td>
</tr>
<tr>
<td>Lys</td>
<td>52</td>
<td>7.89</td>
<td>120.8</td>
<td>7.88</td>
<td>120.7</td>
</tr>
<tr>
<td>Ile</td>
<td>53</td>
<td>8.27</td>
<td>119.8</td>
<td>8.26</td>
<td>119.9</td>
</tr>
<tr>
<td>Residue</td>
<td>δNH</td>
<td>δ15N</td>
<td>δNH</td>
<td>δ15N</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Lys 56</td>
<td>8.53</td>
<td>129.1</td>
<td>8.55</td>
<td>127.2</td>
<td></td>
</tr>
<tr>
<td>His 57</td>
<td>7.95</td>
<td>115.6</td>
<td>7.96</td>
<td>115.9</td>
<td></td>
</tr>
<tr>
<td>Cys 58</td>
<td>9.57</td>
<td>121.7</td>
<td>9.51</td>
<td>121.5</td>
<td></td>
</tr>
<tr>
<td>Arg 59</td>
<td>9.06</td>
<td>126.5</td>
<td>8.90</td>
<td>126.0</td>
<td></td>
</tr>
<tr>
<td>Val 60</td>
<td>8.93</td>
<td>123.9</td>
<td>8.94</td>
<td>124.0</td>
<td></td>
</tr>
<tr>
<td>Gln 61</td>
<td>8.59</td>
<td>125.8</td>
<td>8.57</td>
<td>125.8</td>
<td></td>
</tr>
<tr>
<td>Gln 62</td>
<td>8.66</td>
<td>125.3</td>
<td>8.74</td>
<td>125.2</td>
<td></td>
</tr>
<tr>
<td>Glu 63</td>
<td>8.75</td>
<td>129.7</td>
<td>8.75</td>
<td>129.7</td>
<td></td>
</tr>
<tr>
<td>Gly 64</td>
<td>8.93</td>
<td>116.2</td>
<td>8.90</td>
<td>116.0</td>
<td></td>
</tr>
<tr>
<td>Gln 65</td>
<td>9.05</td>
<td>126.2</td>
<td>9.06</td>
<td>126.4</td>
<td></td>
</tr>
<tr>
<td>Thr 66</td>
<td>7.76</td>
<td>113.4</td>
<td>7.81</td>
<td>113.3</td>
<td></td>
</tr>
<tr>
<td>Val 67</td>
<td>8.74</td>
<td>117.4</td>
<td>8.74</td>
<td>117.5</td>
<td></td>
</tr>
<tr>
<td>Met 68</td>
<td>8.02</td>
<td>121.6</td>
<td>9.03</td>
<td>121.6</td>
<td></td>
</tr>
<tr>
<td>Leu 69</td>
<td>8.42</td>
<td>125.2</td>
<td>8.43</td>
<td>125.3</td>
<td></td>
</tr>
<tr>
<td>Gly 70</td>
<td>9.32</td>
<td>118.3</td>
<td>9.38</td>
<td>118.2</td>
<td></td>
</tr>
<tr>
<td>Asn 71</td>
<td>8.87</td>
<td>124.8</td>
<td>8.89</td>
<td>124.9</td>
<td></td>
</tr>
<tr>
<td>Ser 72</td>
<td>8.28</td>
<td>118.2</td>
<td>8.25</td>
<td>118.0</td>
<td></td>
</tr>
<tr>
<td>Glu 73</td>
<td>7.96</td>
<td>124.1</td>
<td>7.97</td>
<td>124.1</td>
<td></td>
</tr>
<tr>
<td>Phe 74</td>
<td>9.08</td>
<td>120.2</td>
<td>9.10</td>
<td>120.1</td>
<td></td>
</tr>
<tr>
<td>Asp 75</td>
<td>9.41</td>
<td>119.8</td>
<td>9.42</td>
<td>119.7</td>
<td></td>
</tr>
<tr>
<td>Ser 76</td>
<td>7.39</td>
<td>105.5</td>
<td>7.39</td>
<td>105.5</td>
<td></td>
</tr>
<tr>
<td>Leu 77</td>
<td>9.28</td>
<td>122.7</td>
<td>9.27</td>
<td>122.7</td>
<td></td>
</tr>
<tr>
<td>Val 78</td>
<td>7.88</td>
<td>116.5</td>
<td>7.87</td>
<td>116.4</td>
<td></td>
</tr>
<tr>
<td>Asp 79</td>
<td>7.74</td>
<td>121.6</td>
<td>7.76</td>
<td>121.6</td>
<td></td>
</tr>
<tr>
<td>Leu 80</td>
<td>7.34</td>
<td>123.3</td>
<td>7.34</td>
<td>123.1</td>
<td></td>
</tr>
<tr>
<td>Ile 81</td>
<td>7.80</td>
<td>119.4</td>
<td>7.80</td>
<td>119.5</td>
<td></td>
</tr>
<tr>
<td>Ser 82</td>
<td>8.03</td>
<td>112.6</td>
<td>8.02</td>
<td>112.5</td>
<td></td>
</tr>
<tr>
<td>Tyr 83</td>
<td>7.90</td>
<td>121.9</td>
<td>7.90</td>
<td>122.1</td>
<td></td>
</tr>
<tr>
<td>Tyr 84</td>
<td>7.70</td>
<td>118.9</td>
<td>7.69</td>
<td>119.0</td>
<td></td>
</tr>
<tr>
<td>Glu 85</td>
<td>7.46</td>
<td>117.2</td>
<td>7.51</td>
<td>117.3</td>
<td></td>
</tr>
<tr>
<td>Lys 86</td>
<td>7.12</td>
<td>116.4</td>
<td>7.13</td>
<td>116.9</td>
<td></td>
</tr>
<tr>
<td>His 87</td>
<td>7.79</td>
<td>120.2</td>
<td>7.78</td>
<td>120.1</td>
<td></td>
</tr>
<tr>
<td>Leu 89</td>
<td>8.43</td>
<td>124.0</td>
<td>8.43</td>
<td>124.1</td>
<td></td>
</tr>
<tr>
<td>Tyr 90</td>
<td>6.82</td>
<td>117.8</td>
<td>6.82</td>
<td>117.7</td>
<td></td>
</tr>
<tr>
<td>Arg 91</td>
<td>8.97</td>
<td>126.1</td>
<td>9.03</td>
<td>126.2</td>
<td></td>
</tr>
<tr>
<td>Lys 92</td>
<td>0.00</td>
<td>100.0</td>
<td>8.41</td>
<td>119.7</td>
<td></td>
</tr>
<tr>
<td>Met 93</td>
<td>8.83</td>
<td>124.4</td>
<td>8.84</td>
<td>124.5</td>
<td></td>
</tr>
<tr>
<td>Lys 94</td>
<td>7.98</td>
<td>127.4</td>
<td>7.98</td>
<td>127.5</td>
<td></td>
</tr>
<tr>
<td>Leu 95</td>
<td>8.30</td>
<td>119.5</td>
<td>8.29</td>
<td>119.5</td>
<td></td>
</tr>
<tr>
<td>Tyr 96</td>
<td>8.44</td>
<td>121.7</td>
<td>8.44</td>
<td>121.4</td>
<td></td>
</tr>
<tr>
<td>Arg 97</td>
<td>7.77</td>
<td>118.0</td>
<td>7.83</td>
<td>118.0</td>
<td></td>
</tr>
<tr>
<td>Ile 99</td>
<td>8.34</td>
<td>124.3</td>
<td>8.35</td>
<td>124.3</td>
<td></td>
</tr>
<tr>
<td>Asn 100</td>
<td>8.48</td>
<td>124.5</td>
<td>8.47</td>
<td>124.5</td>
<td></td>
</tr>
<tr>
<td>Glu 101</td>
<td>8.64</td>
<td>120.2</td>
<td>8.64</td>
<td>120.2</td>
<td></td>
</tr>
<tr>
<td>Glu 102</td>
<td>8.45</td>
<td>121.1</td>
<td>8.45</td>
<td>121.1</td>
<td></td>
</tr>
<tr>
<td>Asn 103</td>
<td>8.39</td>
<td>119.0</td>
<td>8.40</td>
<td>120.0</td>
<td></td>
</tr>
<tr>
<td>Ser 104</td>
<td>8.09</td>
<td>115.7</td>
<td>8.08</td>
<td>115.6</td>
<td></td>
</tr>
<tr>
<td>Ser 105</td>
<td>7.86</td>
<td>123.2</td>
<td>7.86</td>
<td>123.2</td>
<td></td>
</tr>
</tbody>
</table>

Note: δNH(f), δ15N(f) refer to the chemical shifts of the NH and 15N resonances of the free PLCC SH2 domain, while δNH(b), δ15N(b) refer to the chemical shifts of the NH and 15N resonances of the pY1021-bound PLCC SH2 domain.
First, assignment depends on through-bond and not through-space correlations, which may be ambiguous depending on local structure. Secondly, few correlations are observed in these experiments, making it easier to assign in regions of spectral overlap. Finally, $^{13}\text{C}_\beta$ chemical shifts are diagnostic of amino acid type (Clore et al., 1990). As examples, Ser and Thr have diagnostic downfield chemical shifts, with $^{13}\text{C}_\alpha$ resonances at 55-60 and 59-61 ppm, respectively, while Ser and Thr $^{13}\text{C}_\beta$ resonances lie at 61-67 and 70-73 ppm, respectively. Ala $^{13}\text{C}_\beta$ resonances are uniquely upfield at 19 ± 2 ppm.

E. Binding Kinetics

Following backbone assignment, peptide titrations of $^{15}\text{N}$-labeled samples were used to study the binding of the PLCC SH2 domain to the high-affinity binding phosphopeptide derived from sequences about Tyr 1021 of the PDGFR. A 12-residue phosphopeptide containing the sequence DNDpYIPLPDPK was synthesized in the lab of Dr. Steve Shoelson, and later by Dr. Gerry Gish in the lab of Dr. Tony Pawson. These experiments involved titration of small quantities of this peptide until a 1:1 complex was formed; after each addition of phosphopeptide a $^{15}\text{N}^1\text{H}$-HSQC experiment was recorded. Between six and eight additions of phosphopeptide were typically recorded per titration experiment. Once the 1:1 complex was formed, the $^{15}\text{N}$ and $^1\text{H}$ chemical shifts were compared to those in the free state and used to define the binding site. This is described in more detail in Chapter 6. In addition, lineshape analysis of those resonances involved in phosphopeptide binding during intermediate stages of the titration allows one to obtain information about the kinetics of binding. The subject has been treated mathematically by McConnell (1958) and in a more qualitative manner by Sykes (1982) and has been summarized in the introductory chapter. Briefly, the lineshape depends on the difference between the off-rate and the chemical shift difference (in Hz) between the free and the complexed states. If the
off-rate is large relative to the chemical shift difference (fast exchange), a single sharp peak is observed at the chemical shift ($\delta$) given by the equation below,

$$\delta = \frac{p_f \delta_f + p_c \delta_c}{2}$$

where $p_f$ and $p_b$ represent the populations in the free and complexed states, and $\delta_f$ and $\delta_c$ represent the chemical shifts of the free and complexed states, respectively. In contrast, if the off-rate is small relative to the chemical shift difference between the two states (slow exchange), two sharp resonances are observed with the volume of the two peaks proportional to the ratio of populations in the two states. In situations where the off-rate is comparable to the chemical shift change, there either may be two peaks which are broad (intermediate slow exchange) or one broad peak (intermediate fast exchange). Different resonances may differ in their exchange behavior even if the off-rate is identical due to differences in chemical shift difference between the free and complexed states. Thus by fitting a number of resonances to the appropriate equations (McConnell, 1958), one can obtain the off-rate. This is described in more detail in Appendix 1.

In Figure 11 is shown a region of the HSQC spectrum of a sample of the $^{15}$N-labeled PLCC SH2 domain into which pY1021 phosphopeptide has been added to yield a ratio of phosphopeptide to SH2 domain of 0.46 to 1. The sample was in 100 mM sodium arsenate, pH 7.5 and the spectrum was taken at 10°C. These conditions are somewhat unusual relative to the typical conditions used for studying the PLCC SH2 domain/pY1021 complex i.e. pH 5.8-6.3 at 30°C. However, similar titration behavior was observed in these conditions as well. A number of protein resonances which do not change in chemical shift over the course of the titration, and thus are not involved in peptide binding, show a single sharp resonance throughout the titration. A number of resonances which are close to the peptide binding site show two resonances (for the free and the bound states), and these resonances are broadened. This broadening is particularly severe for Lys 56, His 57, Cys
Figure 11 -- Exchange Kinetics of the PLCC SH2 Domain/pY1021 Complex. $^{15}$N/$^1$H HSQC spectrum of the PLCC SH2 domain/pY1021 complex at a ratio of pY1021 to SH2 domain of 0.46:1. The protein concentration is 1.9 mM, the sample was dissolved in 60 mM sodium arsenate pH 7.5 and the experiment was run at 10° C. Intermediate to intermediate-slow exchange was observed between the free and pY1021-complexed states.
58 and Tyr 90 which have large chemical shift changes upon peptide binding both in the NH and $^{15}$N chemical shift. The effect of this broadening is the reduced intensity of the free and bound peaks, as shown here. Thus many of the resonances show evidence of intermediate slow exchange. In addition, a number of peaks showing smaller chemical shift changes (< 50 Hz) did not yield two separate peaks for the free and bound state over the course of the titration, but only one which was broadened. Thus over the chemical shift range of < 35 Hz, intermediate to intermediate fast exchange was observed.

Lineshape measurements were made on a number of peaks in slow exchange close to the midpoint of the titration. Some measurements were made at pY1021 to SH2 domain ratios of 0.31:1 rather than at pY1021 to SH2 domain ratios of 0.46:1 since signal-to-noise for Lys 56, His 57, Cys 58 and Tyr 90 for the free and complexed peaks at this ratio tended to be quite weak. In contrast, the signal to noise of the free peaks at the pY1021:SH2 domain ratio of 0.31 to 1 was better, although the extent of broadening was usually on the order of only ~5 Hz at this ratio (see Table 1). Lineshape measurements using Mathematica® (see Appendix) were also made on some peaks which showed intermediate fast and intermediate slow behavior near ratios of pY1021:SH2 domain of 1:2. The results are summarized in Table 1, and a lifetime of 20 ± 8 ms for the SH2 domain/pY1021 complex at 10°C has been calculated.

In Figure 12, a similar spectrum for the complex of the PLCC SH2 domain with an 11-residue phosphopeptide containing sequences about Tyr 992 of the EGFR (protein sequence ADEpYLIPQQGF, synthesized in the lab of Steve Shoelson) is shown. The ratio of phosphopeptide to PLCC SH2 domain is 0.66 to 1. A number of resonances give rise to a single broad resonance which, under similar conditions with the pY1021 peptide yields two broad peaks (e.g. Arg 46, Gln 61). In addition, a number of resonances which showed chemical shift changes of > 100 Hz with the pY1021 peptide (i.e. Lys 56, Cys 58, Arg 59) have been broadened very severely such that the intensity from these resonances is either very low or absent. Thus the exchange in this case ranges from intermediate to
Table 2: Linewidth Data for the pY1021 Complex

a. $p_{\text{free}} = 0.69$

<table>
<thead>
<tr>
<th>Residue</th>
<th>Nucleus</th>
<th>$T_2$(ms)</th>
<th>$\Delta \delta_{\text{free-bound}}$(Hz)</th>
<th>$\nu_1/2\text{free}$(Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyr 90</td>
<td>$^{15}\text{N}$</td>
<td>30.4</td>
<td>121.5</td>
<td>15.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {30.5, 20.5, 16.5, 15.0, 14.0, 13.5, 13.0}</td>
<td>tau = 5-35 ms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {0.5, 0.5, 32.0, 29.5, 29.0, 28.5, 27.5}</td>
<td>tau ~ 20 ms</td>
</tr>
<tr>
<td>Tyr 90</td>
<td>$^1\text{H}$</td>
<td>13.2</td>
<td>85</td>
<td>32.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {0.5, 0.5, 32.0, 29.5, 29.0, 28.5, 27.5}</td>
<td>tau = 5-35 ms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {0.5, 22.5, 17.5, 16.0, 14.5, 13.5, 13.0}</td>
<td>tau ~ 15 ms</td>
</tr>
<tr>
<td>His 57</td>
<td>$^{15}\text{N}$</td>
<td>32.3</td>
<td>40</td>
<td>14.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {0.5, 22.5, 17.5, 16.0, 14.5, 13.5, 13.0}</td>
<td>tau = 5-35 ms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {0.5, 37.0, 32.5, 30.0, 29.0, 28.0, 27.5}</td>
<td>tau ~ 25 ms</td>
</tr>
<tr>
<td>His 57</td>
<td>$^1\text{H}$</td>
<td>13.3</td>
<td>74.5</td>
<td>37.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {0.5, 37.0, 32.5, 30.0, 29.0, 28.0, 27.5}</td>
<td>tau = 5-35 ms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {0.5, 37.0, 32.5, 30.0, 29.0, 28.0, 27.5}</td>
<td>tau ~ 10 ms</td>
</tr>
<tr>
<td>Lys 56</td>
<td>$^{15}\text{N}$</td>
<td>32.7</td>
<td>83</td>
<td>13.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {31.5, 20.0, 17.0, 15.0, 14.0, 13.0, 12.5}</td>
<td>tau = 5-35 ms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {31.5, 20.0, 17.0, 15.0, 14.0, 13.0, 12.5}</td>
<td>tau ~ 27.5 ms</td>
</tr>
<tr>
<td>Lys 56</td>
<td>$^1\text{H}$</td>
<td>11.1</td>
<td>32</td>
<td>45.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {40.5, 45.5, 47.5, 48.5, 49.0, 49.5, 49.5}</td>
<td>tau = 5-35 ms</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>linewidths {40.5, 45.5, 47.5, 48.5, 49.0, 49.5, 49.5}</td>
<td>tau ~ 10 ms</td>
</tr>
</tbody>
</table>
b. $p_{\text{free}} = 0.54$

<table>
<thead>
<tr>
<th>Residue</th>
<th>Nucleus</th>
<th>$T_2$(ms)</th>
<th>$\Delta\delta_{\text{free-bound}}$(Hz)</th>
<th>$\nu_{1/2\text{free}}$(Hz)</th>
<th>$\nu_{1/2\text{bnd}}$(Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arg 18</td>
<td>$^1H$</td>
<td>13.6</td>
<td>70.3</td>
<td>31.25</td>
<td>35.15</td>
</tr>
</tbody>
</table>

free linewidths  
{0.5, 0.5, 38.5, 34.0, 31.5, 30.5, 29.0, 28.5, 28.0, 27.5} $\tau = 5-50$ ms
bound linewidths  
{0.5, 0.5, 45.5, 37.0, 34.0, 32.0, 31.0, 30.0, 290.0, 29.0} $\tau = 5-50$ ms
$\tau \sim 25$ ms

Tyr 90 $^1H$  
15.3 85 32.6 36.7

free linewidths  
{0.5, 0.5, 32.5, 29.5, 28.0, 27.0, 26.0, 25.5, 24.5, 24.5} $\tau = 5-50$ ms
bound linewidths  
{0.5, 0.5, 35.5, 31.5, 29.5, 27.5, 27.0, 26.0, 25.5, 25.0} $\tau = 5-50$ ms
$\tau \sim 15$ ms

Tyr 90 $^{15}N$  
31.2 121.5 14.5 15.9

free linewidths  
{0.5, 25.5, 20.5, 17.5, 16.0, 15.5, 14.5, 0.5, 13.5, 13.5} $\tau = 5-50$ ms
bound linewidths  
{0.5, 28.5, 22.0, 19.0, 17.0, 16.0, 15.0, 14.5, 14.5, 13.5} $\tau = 5-50$ ms
$\tau \sim 35,30$ ms

Lys 56 $^{15}N$  
31.8 83 -- 15.5

bound linewidths  
{22.0, 19.5, 17.0, 16.0, 15.0, 14.5, 14.0, 13.5} $\tau = 15-50$ ms
$\tau \sim 32$ ms

*Ser 48 $^{15}N$  
30.4 14.5 17.72

linewidths  
{17.0, 20.0, 21.5, 23.5, 23.5, 24.5} $\tau = 10-35$ ms
$\tau \sim 10$ ms

*Glu 52 $^1H$  
13.6 19.5 35.15

linewidths  
{33.0, 36.0, 37.5, 38.5, 0.5, 0.5} $\tau = 10-35$ ms
$\tau \sim 15$ ms

*Glu 52 $^{15}N$  
30.4 13.7 17.73

linewidths  
{16.5, 18.5, 20.5, 22.0, 23.0, 23.0} $\tau = 10-35$ ms
$\tau \sim 13$ ms

Note: the crosspeaks listed preceded by a * showed intermediate to slow exchange behavior

| Overall Statistics of Table 1 |
|-----|-----|-----|
| $\tau$ average | SD  |
| 20  | 8   |
Figure 12 -- Exchange Kinetics of the PLCC SH2 Domain/pY992 Complex. 

$^{15}$N/$^1$H HSQC spectrum of the PLCC SH2 domain (seq 3) in complex with the pY992 phosphopeptide at a ratio of pY992 to SH2 domain of 0.66 to 1.0. The protein concentration is 1.5 mM, the sample is dissolved in 100 mM sodium phosphate pH 6.3 and the experiment was run at 27°C. From this spectrum one observes intermediate to intermediate-fast exchange between the free and pY1021-complexed states. Resonances shifting significantly upon phosphopeptide binding are highlighted. In addition, since this experiment was performed at pH 6.3, a number of resonances are present which are absent or very weak in Figure 11 where the higher pH was 7.5; these resonances (Leu 77, Gln 65 and Arg 91) are starred.
Lineshape calculations using Mathematica® (see Appendix) were performed on a number of peaks at ratios of phosphopeptide to SH2 domain of 0.22:1, 0.44:1 and 0.66:1. At a ratio of 0.22:1, resonances with large chemical shift changes upon binding were only partially broadened, so that their linewidths could be measured. The results are shown in Table 2 and, from these measurements, a lifetime of $1.2 \pm 0.6$ ms was measured for the PLCC SH2 domain/pY992 complex.

Two methods have commonly been used to study the kinetics of binding of SH2 domains to high affinity phosphopeptides, specifically surface plasmon resonance (trade-name BIAcore) and NMR. In surface plasmon resonance, the peptide is bound to the surface of a gold chip, the protein of interest is flowed along the gold chip and binding is measured as a change in the index of refraction of monochromatic light reflected at the surface of the gold chip. This technique has been applied to measurement of binding constants and on- and off-rates of several SH2 domains to high affinity binding phosphopeptides (Panayotou et al., 1993; Bu et al., 1995; Ladbury et al., 1995; Morelock et al., 1995; Borrello et al., 1996). Typically, dissociation constants ($K_d$) were found of $<20$ nM, association rate constants ($k_a$) of $1.4 \times 10^4$ to $2.5 \times 10^6$ M$^{-1}$s$^{-1}$ and dissociation rate constants ($k_d$) of $3.6 \times 10^{-3}$ to $0.1$ s$^{-1}$. For the PLCC SH2 domain/pY1021 and PLCC SH2 domain/pY992 complex, $K_d$ values of $0.61 \pm 0.16$ μM and $0.82 \pm 0.06$ μM were measured, respectively, with $k_a$ values of $4.87 \times 10^{-3}$ s$^{-1}$ and $0.0157$ s$^{-1}$, respectively (Shun-Cheng Li, personal communication). The $k_d$ values from surface plasmon resonance experiments are several orders of magnitude larger than the values reported from this and another NMR study of SH2 domain binding (Hennsman et al., 1995), and the reasons may be two-fold. Firstly, BIAcore measures rates at surfaces while NMR measures rates of exchange at equilibrium (Hennsman et al., 1994). Secondly, the SH2 domain measurements were made with GST-SH2 domain fusion proteins, however GST is known to dimerize. The dimerization of GST is speculated to be the reason for the
Table 2: Linewidth Data for the pY992 Complex

<table>
<thead>
<tr>
<th>Residue</th>
<th>pfree</th>
<th>Nucleus</th>
<th>T2(ms)</th>
<th>$\Delta \delta_{\text{free-bound}}$(Hz)</th>
<th>$\tau$1/2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tyr 90</strong></td>
<td>0.78</td>
<td>$^{15}\text{N}$</td>
<td>31.8</td>
<td>81</td>
<td><strong>21.1</strong></td>
</tr>
<tr>
<td>linewidths</td>
<td>17.0, 23.0, 28.5, 31.0, 31.5, 30.0, 28.5, 0.5, 25.5</td>
<td>$\tau$ = 0.5-4.5 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linewidths</td>
<td>17.0, 18.5, 19.5, 21.0, 22.5, 23.0</td>
<td>$\tau$ = 0.5-1.0 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\sim 0.8$ ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lys 56</strong></td>
<td>0.78</td>
<td>$^{15}\text{N}$</td>
<td>30.9</td>
<td>77.5</td>
<td><strong>24.0</strong></td>
</tr>
<tr>
<td>linewidths</td>
<td>16.5, 23.0, 27.5, 30.0, 30.5, 30.0, 28.5, 27.0, 26.0</td>
<td>$\tau$ = 0.5-4.5 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linewidths</td>
<td>19.5, 20.0, 21.5, 23.0, 23.5, 24.5, 26.0</td>
<td>$\tau$ = 0.7-1.3 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\sim 1.15$ ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>His 57</strong></td>
<td>0.78</td>
<td>$^{1}\text{H}$</td>
<td>12.3</td>
<td>105</td>
<td><strong>38.6</strong></td>
</tr>
<tr>
<td>linewidths</td>
<td>37.5, 47.5, 53.5, 54.5, 52.5, 50.0, 47.5, 45.0, 43.0</td>
<td>$\tau$ = 0.5-4.5 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linewidths</td>
<td>37.5, 39.5, 41.5, 43.5, 45.5</td>
<td>$\tau$ = 0.55 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\sim 0.55$ ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>His 57</strong></td>
<td>0.34</td>
<td>$^{1}\text{H}$</td>
<td>12.3</td>
<td>105</td>
<td><strong>59.4</strong></td>
</tr>
<tr>
<td>linewidths</td>
<td>41.5, 57.5, 71.5, 0.5, 0.5, 0.5, 0.5, 0.5</td>
<td>$\tau$ = 0.5-4.5 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linewidths</td>
<td>47.5, 51.0, 54.0, 57.5, 60.5, 63.5, 66.5</td>
<td>$\tau$ = 0.7-1.3 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\sim 1.1$ ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ala 46</strong></td>
<td>0.34</td>
<td>$^{15}\text{N}$</td>
<td>10.8</td>
<td>95</td>
<td><strong>58.6</strong></td>
</tr>
<tr>
<td>linewidths</td>
<td>42.5, 55.5, 67.5, 77.0, 82.5, 0.5, 0.5, 0.5, 0.5</td>
<td>$\tau$ = 0.5-4.5 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linewidths</td>
<td>47.5, 50.0, 52.5, 55.5, 58.0, 60.5, 62.5</td>
<td>$\tau$ = 0.7-1.3 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\sim 1.1$ ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ile 47</strong></td>
<td>0.34</td>
<td>$^{15}\text{N}$</td>
<td>35.9</td>
<td>31</td>
<td><strong>13.7</strong></td>
</tr>
<tr>
<td>linewidths</td>
<td>10.0, 12.0, 13.0, 14.5, 16.0, 17.0, 18.5, 20.0, 21.0</td>
<td>$\tau$ = 0.5-4.5 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linewidths</td>
<td>13.0, 13.5, 13.5, 13.5, 14.5, 14.5</td>
<td>$\tau$ = 1.5-2.0 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\sim 1.7$ ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ala 51</strong></td>
<td>0.56</td>
<td>$^{1}\text{H}$</td>
<td>13.6</td>
<td>75</td>
<td><strong>39.06</strong></td>
</tr>
<tr>
<td>linewidths</td>
<td>32.0, 41.5, 51.5, 62.5, 73.0, 81.5, 0.5, 0.5, 0.5</td>
<td>$\tau$ = 0.5-4.5 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linewidths</td>
<td>37.5, 39.5, 41.5, 43.5, 45.5</td>
<td>$\tau$ = 0.8-1.2 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\sim 0.9$ ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ala 51</strong></td>
<td>0.34</td>
<td>$^{1}\text{H}$</td>
<td>13.6</td>
<td>75</td>
<td><strong>35.15</strong></td>
</tr>
<tr>
<td>linewidths</td>
<td>31.5, 39.0, 47.0, 55.0, 60.5, 64.5, 66.5, 67.5, 66.0</td>
<td>$\tau$ = 0.5-4.5 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>linewidths</td>
<td>31.5, 33.0, 35.0, 36.5, 38.0, 39.0</td>
<td>$\tau$ = 0.5-1.0 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\sim 0.7$ ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gln 61</strong></td>
<td>0.56</td>
<td>$^{1}\text{H}$</td>
<td>13.6</td>
<td>35</td>
<td><strong>31.25</strong></td>
</tr>
<tr>
<td>linewidths</td>
<td>25.5, 27.5, 29.5, 31.5, 33.0, 35.0, 37.0, 39.0, 40.5</td>
<td>$\tau$ = 0.5-4.5 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau$</td>
<td>$\sim 2$ ms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Gln 61 0.34  $^1$H 13.6  35  31.25
linewidths {25.0, 26.5, 28.5, 30.0, 32.0, 33.5, 35., 36.5, 38.0} \( \tau = 0.5\text{-}4.5 \text{ ms} \)
\( \tau \sim 2.3 \text{ ms} \)

\( \tau \) average SD
1.2 0.6
low $K_d$ values measured for SH2 domain interactions using BIACore versus other methods for measuring binding affinity (Ladbury et al., 1995). For example, typically BIACore methods obtained nanomolar $K_d$ values for SH2 domain-phosphopeptide interactions, while isothermal titration calorimetry typically yielded $K_d$ values on the order of hundreds of nM, both using SH2 domains alone and GST-SH2 domain fusion proteins (Ladbury et al., 1995). It was not discerned whether this $K_d$ overestimation using BIACore methods was due to over- or under-estimates of the true on- or off-rate.

As stated previously, another study of SH2 domain-peptide binding involving the N-terminal SH2 domain of p85 (Hensmann et al., 1994) also showed that peptide binding resulted in broadening of SH2 domain resonances i.e. peptide binding occurs in the region of intermediate exchange. However, they also presented evidence that SH2 domain binding is a two-step process, though rate constants were not explicitly given. A two-state binding process may be related to CD changes observed in the p85 N-terminal SH2 domain (Panayotou et al., 1992; Shoelson et al., 1993) related to conformational changes about the BG loop upon binding (Nolte et al., 1996). However, CD (Gerry Gish, personal communication) and preliminary analysis of 3D 15N NOESY experiments (data not shown) showed little change in the conformation of the PLCC SH2 domain upon phosphopeptide binding. Thus a one-step binding process may be sufficient. However, the rate at which the free PLCC SH2 domain is exchanging between monomer and dimer is not known (Farrow et al., 1994), and this could be valid in terms of treating peptide binding by this SH2 domain as a two-step process.

III. Summary:

In order to perform a complete structural determination by NMR, a number of separate conditions must be met. First of all, the protein must be cloned into a high expression vector due to the economics of $^{13}$C- and $^{15}$N-labeling. In addition, one must
be able to purify the expressed protein to homogeneity, and the resulting protein should be soluble to concentrations > 1 mM, must be folded in solution and must not aggregate. In this chapter, I show that the PLCC SH2 domain, when cloned into the pET expression system, met all of these conditions. I described in detail the purification scheme used for the PLCC SH2 domain when cloned into the pET expression system, as well as a purification for an earlier construct involving expression of the PLCC SH2 domain as a GST-fusion protein. From a sequence analysis of the PLCC SH2 domain, one observes a large excess of positively-charged amino acids. This was taken advantage of in purification schemes described: both the Whatman CM-52 resin and the Pharmacia MonoS pre-packed PHAST columns act as cation-exchangers binding positively-charged proteins (such as the PLCC SH2 domain), while the Whatman DE-52 resin acts as an anion-exchanger, binding negatively-charged proteins (GST, but not the PLCC SH2 domain). Phosphocellulose resins have been described both as cation-exchangers and affinity columns, in terms of their tight binding to proteins such as protein kinases which bind to phosphate groups such as ATP, or in the case of the SH2 domain, the pTyr phosphate group.

Evidence was also presented showing that the PLCC SH2 domain in solution is folded, based on the dispersion of NH chemical shifts observed in $^{15}$N,$^1$H HSQC spectra of the PLCC SH2 domain both alone in solution and in complex with high-affinity binding phosphopeptides. Further experiments showed that the PLCC SH2 domain in solution is a mixture of monomeric and dimeric species, but upon forming a complex with the pY1021 phosphopeptide, the resulting complex is exclusively monomeric.

The behavior of the PLCC SH2 domain allowed us to perform a preliminary characterization and assignment of the free SH2 domain and its complex with a pair of high-affinity phosphopeptides. We assigned the backbone resonances both free and in complex with the pY1021 phosphopeptide using the HNCACB and CBCA(CO)NNH experiments. In addition, I showed that the off-rate of binding of high-affinity phosphopeptides is quite rapid: the pY1021 phosphopeptide yields intermediate-slow
exchange behavior with an estimated off-rate of $20 \pm 8$ ms, while the pY992 phosphopeptide yields intermediate-fast exchange behavior with an estimated off-rate of $1.2 \pm 0.6$ ms, i.e. one order of magnitude faster than the pY1021 phosphopeptide.
Chapter 3: Structure of the pY1021 Phosphopeptide as it is Bound to the PLCC SH2 Domain

I. Introduction

2D and 3D solution NMR techniques have been used to solve the solution structure of both macromolecules and macromolecular complexes. Macromolecular complexes include those between protein and DNA, protein and peptide, and protein and drug. The Nuclear Overhauser Effect (NOE), in which two protons exchange magnetization due to their close proximity, has been crucial in solving the solution structures of macromolecules by NMR techniques (see Chapter 1). NOEs in macromolecular peptide complexes can be divided into three classes, namely protein-protein, protein-peptide and peptide-peptide NOEs. Homonuclear techniques cannot differentiate these three classes of NOEs. However, metabolic labeling of proteins expressed in E. coli with $^{13}$C-labeled glucose and $^{15}$NH$_4$Cl, $^{13}$C,$^{15}$N doubly-labeled protein can be synthesized at relatively low cost. Then by utilizing a double-labeled protein and a chemically-synthesized peptide (with natural abundance carbon and nitrogen nuclei), the three classes of NOEs can be distinguished using the three types of experiments as shown in Figure 1 (for a review, see Wand & Short, 1994). By selecting for protons to which $^{15}$N and $^{13}$C nuclei are attached, one can select for protein protons only, and perform assignment and NOE identification of the protein resonances in the presence of bound ligand. These are the F$_1$/F$_2$-edited experiments. Similarly, by selecting against protons to which $^{15}$N and $^{13}$C nuclei are attached, one can select for peptide protons only, and perform assignment and structure determination of the peptide as it is bound to the protein. These are the F$_1$/F$_2$- (or double-) filtered experiments shown in Figure 1, and will be discussed in this chapter. Finally, an experiment has been developed in which magnetization begins on the protein (i.e. peptide magnetization is selected.
Figure 1 -- NMR Spectra: Filtered and Edited Experiments. Strategy for determination of solution structures of macromolecular complexes, in which one component has been labeled by $^{13}$C and $^{15}$N nuclei, while the other has not. Filtering of magnetization from protons attached to $^{13}$C and $^{15}$N nuclei allows for observation of peptide resonances, while editing magnetization from protons attached to $^{13}$C and $^{15}$N nuclei allows for observation of protein resonances. In addition, an experiment which begins with $^{15}$N,$^{13}$C attached proton magnetization, followed by an NOE mixing period, and followed by filtration of $^{15}$N,$^{13}$C attached proton magnetization (the F$_1$-edited, F$_2$-filtered NOESY) allows observation of protein-peptide contacts exclusively.
1. SH2 domain resonances only -- F1/F2-edited experiments

2. Phosphopeptide resonances only -- F1/F2-filtered experiments

3. SH2-phosphopeptide interactions -- F1-edited/F2-filtered NOESY
against), and then is transferred via an NOE mixing time to the peptide, so that protein-peptide NOEs only are observed. This is the F₁-edited/F₂-filtered (or half-filtered) NOESY experiment, and data from this experiment obtained from the PLCC SH2 domain/pY1021 complex is discussed in detail in Chapter 4.

By this "divide and conquer" approach, peptide and protein assignment of the PLCC SH2 domain/pY1021 phosphopeptide complex could be performed separately, as well as determination of peptide/protein contacts. In the previous chapter dealing with protein assignment, I dealt with a few of the F₁/F₂-edited experiments used in SH2 domain backbone assignment. However, Dr. Steve Pascal was responsible for most work on protein resonance assignment and protein-protein NOE assignment, and so this aspect of the work will not be discussed in detail except where it deals with the overall fold of the protein (Chapter 4), or in special cases where I played a role in assignment of the protein-protein NOEs, such as those involving Arg guanidinium groups (see Chapter 6). In this chapter, I will discuss my use of the F₁/F₂-filtered experiments to obtain complete assignment of the peptide when bound to the SH2 domain, as well as determination of a number of peptide-peptide NOEs.

In Figure 2 is shown the pulse sequence for the double-filtered NOESY in H₂O (written by Lewis Kay) in order to explain the mechanism by which ¹⁵N- and ¹³C-attached proton magnetization is filtered. Since the mechanism of filtering unwanted magnetization is the same among the experiments discussed in this chapter (namely the double-filtered NOESY, HOHAHA, COSY and PE-COSY), the pulse sequences are identical up to point B and after point C, and in these regions filtering occurs, and between these two regions are contained many of the pulses and delays used in the unfiltered versions of these experiments. For example, in the double-filtered NOESY, spin filtering is followed by a t₁ delay, followed by a 90° proton pulse in order to bring the transverse magnetization to the z axis, followed by an NOE mixing period, and followed by a second 90° pulse which brings proton magnetization back to the transverse plane. This sequence
Figure 2 -- Pulse Sequence Employed For the 2D $^{13}$C,$^{15}$N $F_1/F_2$-Filtered NOESY. The mechanism of filtration of $^{13}$C- and $^{15}$N-attached proton magnetization is similar among all experiments discussed in this chapter. Three points are highlighted during this experiment (A, B and C) in order to demonstrate the mechanism of filtering $^{13}$C- and $^{15}$N-attached proton magnetization (A), and the NOESY pulse sequence (starting at point B and ending at point C).
was described in the original pulse sequence for the 2D NOESY (Kessler et al., 1988).

In Figure 2, the pulse sequence up to part A illustrates the basic mechanism of filtering, namely that an initial 90° pulse creates transverse proton magnetization, followed by a delay of $1/(2^* J_{HX})$, where $X$ represents the heteronucleus of interest (Ikura and Bax, 1992). During this delay, proton magnetization involving protons attached to heteronuclei is transformed from in-phase proton magnetization to antiphase magnetization; a subsequent 90° pulse on the heteronucleus converts the antiphase magnetization into unobservable zero- and double-quantum coherence, and this magnetization is lost. Proton magnetization arising from protons not attached to NMR-active heteronuclei do not give rise to this type of antiphase magnetization, so the subsequent 90° heteronuclear pulse has no effect.

While this basic philosophy is utilized, between the pulse in part A and the NOESY sequence beginning in part B are a number of additional pulses which allow for a more sophisticated filtering scheme (Lee et al., 1994). Filtering of magnetization of proton nuclei attached both to $^{13}$C nuclei and $^{15}$N nuclei is desired; thus the delay $\tau_A$ is set to $1/(2^* J_{HC})$, approximately 3 ms, while the delay $\tau_A + \tau_B + \tau_C$ is set to $1/(2^* J_{HN})$, approximately 5.5 ms. A range of $J_{HC}$ is often observed in proteins, from 120 Hz in aliphatic groups to 160 Hz in aromatic groups (Lee et al., 1994), and selecting a $J_{HC}$ value of 120 Hz leads to inefficient filtration of signals of protons attached to aromatic carbons. Thus the delay $\tau_A$ is set to $1/(2^* J_{HC})$ for the smaller coupling and the delay period ($-\tau_B + \tau_C + \tau_D$) is set to $1/(2^* J_{HC})$ for the larger coupling.

Residues of phosphopeptides in contact with SH2 domains generally adopt an extended conformation which lies perpendicular to the large $\beta$-sheet of the SH2 domain (Breeze et al., 1996; Eck et al., 1993; Eck et al., 1994; Gilmer et al., 1994; Gosser et al., 1995; Hatada et al., 1995; Lee et al., 1994; Narula et al., 1995; Nolte et al., 1996; Waksman et al., 1992; Waksman et al., 1993; Xu et al., 1995; Zhou et al., 1995). Residues not in contact with the SH2 domain are not expected to adopt a preferred
conformation and can be thought of as random coil. Recently, the peptide of the GRB-2 SH2 domain/YXNX complex has been observed to adopt a turn conformation (Rahuel et al., 1996), as does the region from -3 to the pTyr residue in a number of PTB-phosphopeptide complexes (Zhou et al., 1995; Zhou et al., 1996). Thus, the conformation of this peptide in this SH2 domain complex is of interest.

II. Materials and Methods

The majority of double-filtered experiments were performed on a 1.8 mM sample of the PLCC SH2 domain, pH 6.4, in 100 mM sodium phosphate, to which pY1021 phosphopeptide was added until a 1:1 SH2 domain/pY1021 complex was achieved. Double-filtered experiments included NOESY and HOHAHA experiments at 10° C in water. Following these experiments, the sample was lyophilized and resuspended in D2O. After D2O exchange, double-filtered experiments were recorded on the sample at 30° C including two NOESY (at mixing times of 50 and 150 ms), HOHAHA, COSY and PE-COSY experiments. Initial experiments at 10° C were performed in order to test for optimal conditions for protein-peptide NOEs (see Chapter 4): peptide resonances however were not observed to change over the temperature range studied. A double-filtered NOESY and TOCSY had also been performed on an earlier sample (1.1 mM PLCC SH2 domain in 100 mM sodium phosphate, pH 6.3) which contained an excess of peptide (1.3 mM). This situation was not optimal because the presence of excess peptide led to observation of exchange NOEs, where one of the protons in the NOE was at the chemical shift of the free peptide resonance, leading to additional NOEs and spectral overlap. HOHAHA experiments were used for spin system identification, while the COSY and PE-COSY experiment were used to identify vicinally coupled spins in the spin system and for measurement of coupling constants, respectively.
Parameters for the NOESY experiment in H₂O at 10° C include sweep widths of 8000 and 4750 Hz in t₂ and t₁, respectively, 128 transients, 1024 points in t₂, 128 t₁ points and a mixing time of 150 ms, with the transmitter set to the center of the H₂O resonance. A mixing time of 42.4 ms was used for the HOHAHA experiment in H₂O. Other parameters were identical between the NOESY and HOHAHA.

With respect to the NOESY experiments in D₂O, the parameters for the first experiment include a sweep width in t₂ of 8000 Hz, a sweep width in t₁ of 5250 Hz, a mixing time of 150 ms, 96 transients, 256 t₁ points, 1024 t₂ points and the transmitter set to the center of the H₂O resonance. Another NOESY experiment was run with a mixing time of 50 ms in order to obtain stereoassignments and information about the χ₁ dihedral angle. For this experiment, sweep widths in t₁ and t₂ of 4000 and 6000 Hz, respectively, were used centered at 4.10 ppm, and 1536 t₂ and 220 t₁ points, respectively, were collected. The HOHAHA experiment had similar parameters except for a mixing time of 28.7 ms and 128 t₁ points. The COSY and PE-COSY also had similar parameters, except that 64 transients and 512 t₁ points were collected, respectively.

The pulse sequences for all double-filtered experiments were written by Lewis Kay. NMR experiments were run on Varian UNITY 500 spectrometers equipped with a triple-resonance pulsed-field gradient probe with an actively shielded z-gradient and a gradient amplifier unit. NMR data were processed and displayed using the NMRPipe suite of programs (Delaglio et al., 1995) and analyzed using the program PIPP (Garrett et al., 1991). HOHAHA and NOESY experiments were apodized using 60-90° phase-shifted sine-bell squared functions, while the COSY and PE-COSY experiments were apodized using a 36° phase-shifted sine-bell squared function.
III. Results

A. Assignment

In this chapter and subsequent chapters, the pY1021 peptide (sequence DNDYIIPLPDPK) will be numbered with reference to the pTyr residue according to the convention used first by Eck et al. (1993) and Waxman et al. (1992), i.e. Leu +4, Asn -2.

I performed spin system identification using an F1/F2 filtered HOHAHA, and these spin systems were linked with an F1/F2-filtered NOESY by observation of sequential NOEs. The strongest sequential NOEs were of the type CαH_i/NH_{i+1}, indicative of an extended strand. For sequential NOEs involving proline residues, strong sequential CαH_i/CγH_{i+1} NOEs involving both Hδ protons were observed, indicative of an extended strand with a trans X-Pro peptide bond (Wüthrich, 1986). No amide signal was observed for the N-terminal residue or Asn -2, likely because the N-terminal NH3- and Asn -2 amide is in rapid exchange with water.

Following spin system identification, side chain resonances were assigned. For AMX systems (five in this peptide), side chain assignment is trivial, since there are only two or three resonances, one from the H_α, and another one or two from the H_β protons. H_α protons can be identified by being downfield of the H_β resonances. Among the AMX spin systems, both pTyr and Asn -2 have additional side-chain resonances which are not connected to backbone resonances through scalar coupling networks: the H_β protons are not scalar coupled to the aromatic protons of pTyr or NH2 (Hδ) protons of Asn. However, intra-residue H_β/NH2 and H_β/Hδ NOEs were observed for Asn -2 and pTyr, respectively, allowing for complete resonance assignment of these amino acids. For three residues with non-AMX spin systems (Leu +4, Pro +7, Lys +8), chemical shifts of aliphatic resonances were practically identical to random coil chemical shifts for these.
amino acids; therefore resonances could be assigned by using the table of random coil chemical shifts given in Wüthrich (1986).

However, assignment was more difficult for four other residues, namely Ile +1, Ile +2, Pro +3 and Pro +5. In Figure 3 is shown an upfield region of a double-filtered TOCSY spectrum (mixing time = 26.7 ms) displaying correlations for Ile +1, Ile +2, Leu +4 and Lys +8. For the two Ile residues, correlations from the methyl group protons to vicinal protons, i.e. from the Hγ2 methyl group to the Hβ proton and from the Hδ methyl protons to both Hγ1 protons, were observed. The presence of these crosspeaks in the double-filtered COSY and HOHAHA spectra are expected, since methyl groups have relatively large coupling constants with vicinal protons due to their rapid rotation. Longer correlations (such as from the Hδ to the Hβ) are of weaker intensity in these spectra.

Of note in this TOCSY spectrum are the correlations involving the Ile +2 methyl protons, including the Hβ proton correlations to the Hγ2 protons at 1.01 ppm and the two Hγ1 correlations to the Hδ protons at 0.97 ppm. These methyl groups have very similar chemical shifts to each other, as well as to one of the Hγ1 methylene protons of Ile +1 and both methyl groups of Leu +4 (which resonate at 0.99 ppm). Thus a great deal of chemical shift overlap is observed at or near 0.99 ppm, involving four methyl groups in Ile +2 and Leu +4 and a methylene proton of Ile +1. This overlap becomes problematic in unambiguous assignment of both peptide-peptide NOEs and protein-peptide NOEs as will be discussed in the following chapter.

The assignment of the Hβ protons in the two Ile residues is in part based on their resonance frequencies being more downfield of the methylene and methyl protons (Wüthrich, 1986), as well as through-bond and through-space couplings. For Ile +2, an Hα/Hβ crosspeak is observed in the COSY and HOHAHA spectra. In Ile +1, however, no Hα/Hβ crosspeak is observed in the COSY and HOHAHA spectra, indicating that the coupling constant between these two protons is small. However, a large NOE is observed between these two protons. These observations will be discussed later with reference to the
Figure 3 -- Assignment of the pY1021 Resonances When Bound to the PLCC SH2 Domain. The upfield region of an F1/F2-filtered TOCSY of the PLCC SH2 domain in D2O, 30°C, with a mixing time of 28.7 ms, showing correlations involving Ile +1, Ile +2, Leu +4 and Lys +8.
conformation about the dihedral angle $\chi_1$ for these Ile residues.

For Pro residues, assignment of $H_\alpha$ and $H_\delta$ protons was in part based on chemical shift arguments with $H_\alpha$ protons of trans Pro in random coil peptides resonating at $\sim$4.4 ppm, and $H_\delta$ protons of trans Pro in random coil peptides resonating between 3.65-3.70 ppm (Wüthrich, 1986). Their assignment was also based on the sequential NOEs mentioned previously. For assignment of Pro $H_\beta$ and $H_\gamma$ protons, NOEs in both 50 ms and 150 ms double-filtered NOESY experiments were very useful. In these spectra, the largest NOEs from the $H_\alpha$ proton are to the $H_\beta$, typically the $H_\beta_1$ proton, while the largest NOEs from the $H_\delta$ protons are to the $H_\gamma$ protons. In Pro +3, the largest NOE and scalar coupling correlations of the $H_\alpha$ and $H_\delta$ protons were to a resonance at 2.13 ppm, leading to the conclusion that the $H_\beta_1$, $H_\gamma_1$ and $H_\gamma_2$ protons have degenerate chemical shifts.

A table of complete chemical shifts for pY1021 is shown in Table 1. Note that some $H_\beta$ protons were stereoassigned; this stereoassignment will be discussed in a later section. The resonances of the free pY1021 peptide were also assigned; the chemical shifts of both the free and bound peptides are given as well as the difference in chemical shift between the bound and free peptides. Assignment of the free phosphopeptide was accomplished by a HOHAHA of the free phosphopeptide, and by use of a sample in which excess phosphopeptide was added to the PLCC SH2 domain, in which NOEs were observed to both the free and bound resonances due to chemical exchange of the peptide between the free and the bound forms. Large chemical shifts changes are observed for the pTyr, Ile +1 and Ile +2 resonances upon binding to the SH2 domain, as well as the $H_\alpha$ resonance of Leu +4. The significance of these observations will be discussed later with respect to the structure and binding of the phosphopeptide to the PLCC SH2 domain.

It is of note that the pTyr $H_\alpha$ and Ile +1 NH resonances were much weaker in intensity and somewhat broader than the other pY1021 resonances. The reason for this is not immediately clear, however it is of note that these two protons are close in space.

121
Table 1: Comparison of Chemical Shifts of the Free and PLCC SH2 Domain-complexed pY1021 Phosphopeptide

<table>
<thead>
<tr>
<th></th>
<th>δf (ppm)</th>
<th>δb (ppm)</th>
<th>δΔ (ppm)</th>
<th></th>
<th>δf (ppm)</th>
<th>δb (ppm)</th>
<th>δΔ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp -3</td>
<td></td>
<td></td>
<td></td>
<td>Ile +2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hα</td>
<td>4.23</td>
<td>4.24</td>
<td>0.01</td>
<td>NH</td>
<td>8.18</td>
<td>8.54</td>
<td>0.36</td>
</tr>
<tr>
<td>Hβ</td>
<td>2.84</td>
<td>2.84</td>
<td>0.00</td>
<td>Hα</td>
<td>4.45</td>
<td>4.73</td>
<td>0.28</td>
</tr>
<tr>
<td>Hβ</td>
<td>2.71</td>
<td>2.73</td>
<td>0.02</td>
<td>Hβ</td>
<td>1.87</td>
<td>1.99</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hγ1</td>
<td>1.56</td>
<td>1.62</td>
<td>0.06</td>
</tr>
<tr>
<td>Asn -2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hα</td>
<td>4.64</td>
<td>4.64</td>
<td>0.00</td>
<td>Hγ1</td>
<td>1.23</td>
<td>1.27</td>
<td>0.04</td>
</tr>
<tr>
<td>Hβ</td>
<td>2.70</td>
<td>2.67</td>
<td>-0.03</td>
<td>Hγ2</td>
<td>1.16</td>
<td>0.91</td>
<td>-0.25</td>
</tr>
<tr>
<td>Hβ</td>
<td>2.79</td>
<td>2.74</td>
<td>-0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hδ</td>
<td>6.89</td>
<td>6.92</td>
<td>0.03</td>
<td>Pro +3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hδ</td>
<td>7.66</td>
<td>8.04</td>
<td>0.38</td>
<td>Hα</td>
<td>4.42</td>
<td>4.52</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hβ1</td>
<td>2.27</td>
<td>2.14</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>8.30</td>
<td>8.42</td>
<td>0.12</td>
<td>Hγ2</td>
<td>1.90</td>
<td>1.85</td>
<td>-0.05</td>
</tr>
<tr>
<td>Hα</td>
<td>4.56</td>
<td>4.54</td>
<td>-0.02</td>
<td>Hγ</td>
<td>2.03</td>
<td>2.14</td>
<td>0.11</td>
</tr>
<tr>
<td>Hβ</td>
<td>2.51</td>
<td>2.50</td>
<td>-0.01</td>
<td>Hγ1</td>
<td>1.23</td>
<td>1.27</td>
<td>0.04</td>
</tr>
<tr>
<td>Hβ</td>
<td>2.60</td>
<td>2.57</td>
<td>-0.03</td>
<td>Hδ</td>
<td>3.85-3.93</td>
<td>3.89</td>
<td>-0.04-0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.65-3.73</td>
<td>3.82</td>
<td>0.09-0.17</td>
</tr>
<tr>
<td>pTyr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>8.10</td>
<td>8.58</td>
<td>0.48</td>
<td>Hα</td>
<td>4.42</td>
<td>4.52</td>
<td>0.10</td>
</tr>
<tr>
<td>Hα</td>
<td>4.60</td>
<td>5.07</td>
<td>0.47</td>
<td>Hβ1</td>
<td>2.27</td>
<td>2.14</td>
<td>-0.13</td>
</tr>
<tr>
<td>Hβ</td>
<td>3.12</td>
<td>3.19</td>
<td>0.07</td>
<td>Hβ2</td>
<td>1.90</td>
<td>1.85</td>
<td>-0.05</td>
</tr>
<tr>
<td>Hβ</td>
<td>2.98</td>
<td>2.83</td>
<td>-0.15</td>
<td>Hγ</td>
<td>2.03</td>
<td>2.14</td>
<td>0.11</td>
</tr>
<tr>
<td>Hδ</td>
<td>7.14</td>
<td>7.01</td>
<td>-0.13</td>
<td>Hδ</td>
<td>3.85-3.93</td>
<td>3.89</td>
<td>-0.04-0.04</td>
</tr>
<tr>
<td>Hε</td>
<td>7.19</td>
<td>6.91</td>
<td>-0.28</td>
<td>Hδ</td>
<td>3.65-3.73</td>
<td>3.82</td>
<td>0.09-0.17</td>
</tr>
<tr>
<td>Ile +1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH</td>
<td>8.04</td>
<td>8.82</td>
<td>0.78</td>
<td>NH</td>
<td>8.23</td>
<td>8.17</td>
<td>-0.06</td>
</tr>
<tr>
<td>Hα</td>
<td>4.14</td>
<td>4.67</td>
<td>0.53</td>
<td>Hα</td>
<td>4.62</td>
<td>4.88</td>
<td>0.26</td>
</tr>
<tr>
<td>Hβ</td>
<td>1.80</td>
<td>2.02</td>
<td>0.22</td>
<td>Hβ</td>
<td>1.59</td>
<td>1.61</td>
<td>0.02</td>
</tr>
<tr>
<td>Hγ1</td>
<td>1.46</td>
<td>1.37</td>
<td>-0.09</td>
<td>Hγ</td>
<td>1.75</td>
<td>1.75</td>
<td>0.00</td>
</tr>
<tr>
<td>Hγ1</td>
<td>1.16</td>
<td>0.99</td>
<td></td>
<td>Hδ</td>
<td>0.97</td>
<td>0.99</td>
<td>0.02</td>
</tr>
<tr>
<td>Hγ2</td>
<td>1.16</td>
<td>0.91</td>
<td>-0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hδ</td>
<td>0.84</td>
<td>0.72</td>
<td>-0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hα</td>
<td>Hβ1</td>
<td>Hβ2</td>
<td>Hγ</td>
<td>δ</td>
<td></td>
<td>Hδ</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>Pro +5</td>
<td>4.42</td>
<td>2.27</td>
<td>1.90</td>
<td>2.03</td>
<td>3.85-3.93</td>
<td>3.65-3.73</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>4.48</td>
<td>2.26</td>
<td>1.93</td>
<td>2.09</td>
<td>3.88</td>
<td>3.72</td>
<td>0.06</td>
</tr>
<tr>
<td>Pro +7</td>
<td>4.42</td>
<td>2.27</td>
<td>1.90</td>
<td>2.03</td>
<td>3.85-3.93</td>
<td>3.65-3.73</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>4.42</td>
<td>2.26</td>
<td>2.02</td>
<td>2.02</td>
<td>3.84</td>
<td>3.75</td>
<td>-0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>NH</th>
<th>Hα</th>
<th>Hβ</th>
<th>Hγ</th>
<th>δ</th>
<th></th>
<th>Hδ</th>
<th>δ</th>
<th></th>
<th>Hε</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp +6</td>
<td>8.30</td>
<td>4.84</td>
<td>2.49</td>
<td>2.72</td>
<td>7.83</td>
<td>6.30</td>
<td>0.03</td>
<td>-0.01</td>
<td>-0.03</td>
<td>3.03</td>
</tr>
<tr>
<td></td>
<td>8.33</td>
<td>4.83</td>
<td>2.44</td>
<td>2.69</td>
<td>7.87</td>
<td>6.33</td>
<td>-0.01</td>
<td>-0.03</td>
<td>2.99</td>
<td>0.04</td>
</tr>
<tr>
<td>Lys +8</td>
<td>NH</td>
<td>Hα</td>
<td>Hβ</td>
<td>Hγ</td>
<td>δ</td>
<td></td>
<td>Hδ</td>
<td>δ</td>
<td></td>
<td>Hε</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
<td>-----</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>8.14</td>
<td>4.15</td>
<td>1.84</td>
<td>1.74</td>
<td>7.91</td>
<td>6.40</td>
<td>-0.05</td>
<td>-0.03</td>
<td>1.67</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>7.83</td>
<td>4.10</td>
<td>1.81</td>
<td>1.71</td>
<td>7.87</td>
<td>6.40</td>
<td>-0.23</td>
<td>-0.03</td>
<td>1.67</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>7.87</td>
<td>4.00</td>
<td>1.81</td>
<td>1.71</td>
<td>7.87</td>
<td>6.40</td>
<td>-0.23</td>
<td>-0.03</td>
<td>1.67</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

Legend:

δ_f -- Chemical shift of the free phosphopeptide
δ_b -- Chemical shift of the phosphopeptide in the PLCC SH2 domain/pY1021 complex
δΔ -- Difference in chemical shift such that δΔ = δ_f - δ_b

Sample Conditions:

pY1021 Complex -- 1.8 mM sample of the PLCC SH2 domain, pH 6.4, in 100 mM sodium phosphate, to which pY1021 phosphopeptide was added until a 1:1 SH2 domain/pY1021 complex was achieved. Assignment experiments were run at 10°C, the sample was lyophilized and exchanged with D2O, then experiments were run at 30°C. Resonances were not observed to change over the temperature change. As well, assignments were based on a sample of 1.1 mM PLCC SH2 domain to which excess pY1021 peptide (1.3 mM) had been added.

Free Phosphopeptide -- Assignments are based on a TOCSY (60 ms mixing time) of 1 mM pY1021 peptide in 100 mM Sodium Phosphate, pH 6.0. Spin systems were connected by a 150 ms NOESY of a sample of 1.1 mM PLCC SH2 domain to which excess pY1021 peptide (1.3 mM) had been added.
B. Peptide Structure

As stated previously Hαi/NHi+1 NOEs were observed throughout the peptide (see Figure 4), and these NOEs were of high intensity with the exception of that between the Cα of pTyr and the NH of Ile +1, which were of medium intensity relative to other NOEs. However, since both of these resonances are broadened so that all correlations involving these resonances were weaker than those of the other resonances, the sequential NOE involving these two protons was treated as a strong interaction. In contrast, no NHi/NHi+1 NOEs were observed. In the case of proline residues, strong Cαi/Cδi+1 NOEs were observed to both Hδ protons. This pattern of sequential NOEs is diagnostic of the peptide backbone adopting an extended β conformation, rather than a helical one (Wüthrich, 1986). In addition, the presence of Cαi/Cδi+1 NOEs when residue i+1 is a Pro residue is indicative of a trans, rather than cis X-Pro peptide bond, in which Cαi,i+1 NOEs are observed (Wüthrich, 1986).

The pattern of sequential NOEs within the pY1021 peptide is shown in Figure 5, characteristic of an extended conformation throughout the peptide. In addition, chemical shift differences for the NH and Hα protons between the free and bound pY1021 phosphopeptide are shown: large downfield chemical shifts (> 0.4 ppm) are observed for the NH and Hα protons of the pTyr, Ile +1 and Ile +2 residues, and the Hα proton of Leu +4, while others are < 0.1 ppm of the free peptide values. Large downfield chemical shift changes for the NH and Hα protons for at least three sequential residues are indicative of the presence of an extended conformation for those residues (Wishart & Sykes, 1994). Thus, the chemical shift evidence in the pY1021 peptide confirms NOE data demonstrating that residues pTyr through Leu +4 adopt an extended conformation. However, outside this region, minimal chemical shift differences suggest that the phosphopeptide is flexible.

No long-range NOEs were observed among the peptide-peptide NOEs. This point is
Figure 4 -- Extended Structure for the pY1021 Peptide in Complex with the PLCC SH2 Domain. Regions of an F1/F2-filtered NOESY of the PLCC SH2 domain/pY1021 complex in H2O, 10° C, 150 ms mixing time. In (a) is shown the Hα-amide region while in (b) is the amide-amide region. The presence of numerous crosspeaks in (a) and the absence of crosspeaks in (b) is indicative of the peptide adopting an extended conformation.
Figure 5 -- Summary of Sequential NOEs in the pY1021 Peptide. Differences in the NH and H\textsubscript{\alpha} chemical shifts between the pY1021 peptide upon binding to the PLCC SH2 domain and in the free state ($\Delta \delta = \delta_{\text{bound}} - \delta_{\text{free}}$) are also shown.
\[ d_{\alpha N(i,i+1)} \]
\[ d_{\alpha \delta(i,i+1)} \]
\[ d_{\beta,\gamma 2N(i,i+1)} \]
\[ d_{\beta,\gamma 2\delta(i,i+1)} \]
\[ \Delta \delta_{HN} \]
\[ \Delta \delta_{Ho} \]
emphasized in Figure 6, in which the peptide-peptide NOE data are summarized in matrix form, with numbers of inter-residue NOEs along the diagonal and intra-residue NOEs as off-diagonal elements. The longest range NOE observed is a weak NOE involving the Hδ methyl group of Ile +1 and the NH of Leu +4 (i,i+3); all other inter-residue NOEs are either sequential or i,i+2, also indicative of an extended conformation.

NOEs involving residues i,i+2 were observed in two regions of the peptide. Those of moderate to weak intensity were observed between the Hβ of Asn -2 (and Asp -1) and the pTyr aromatic ring protons, indicating that Asn-2 and Asp -1 may interact with the pTyr ring. In addition, a strong intensity NOE between the Ile +1 Hδ methyl group and the Pro +3 Hα was observed, indicating that the +1 and +3 residues come within < 3 Å. A summary of the peptide/peptide NOEs is shown in Figure 7.

C. Stereospecific Assignment, χ1 Restraints and Proline Rings

Stereassignment of methylene protons can lead to greater precision in NMR structures for a number of reasons. NOEs involving methylene protons which are not stereassigned have an ambiguity in assignment, in that the NOE observed involves one of two protons. Typically the ambiguity in assignment is partially overcome by the use of pseudoatom corrections to the distances derived from NOEs; in the case of methylene protons, this pseudoatom correction is given a value of 1.0 Å (Brünger, 1992). However, by stereassignment of the methylene protons, the NOEs can be uniquely assigned and the pseudoatom correction does not need to be added, leading to greater precision in the calculated structures. In addition, stereassignment protocols often result in determination of restraints about side chain dihedral angles, leading to additional precision in the calculated structures. In particular, stereassignment of β-methylene protons leads to determination of restraints about the dihedral angle χ1. Ile residues only contain one Hβ proton, however when methods used to stereassign β-methylene protons are applied to
**Figure 6 -- pY1021 NOEs I.** Matrix showing the number of NOEs between residues, with diagonal elements representing the number of intra-residue NOEs. Matrix elements represented by an "-" are indicative of no NOEs being observed between these residues.
<table>
<thead>
<tr>
<th></th>
<th>D</th>
<th>N</th>
<th>D</th>
<th>Y</th>
<th>I</th>
<th>I</th>
<th>P</th>
<th>L</th>
<th>P</th>
<th>D</th>
<th>P</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>4</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>12</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>9</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td>9</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>5</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 7 -- pY1021 NOEs II. Strong and medium intensity peptide-peptide NOEs observed in F1/F2-filtered NOESY spectra of the PLCC SH2 domain/pY1021 complex. NOEs are represented by dashed lines, with thicker lines for the strong NOEs. Protons are not indicated and NOEs are shown between the heavy atoms to which the two protons involved in the NOE are attached.
Ile, one can also get information about the dihedral angle $\chi_1$ for this residue.

Stereoassignment can be performed by using a combination of distance information for the $\beta$-methylene protons and measurement of coupling constants. The most accurate distance information can be obtained by analysis of a NOESY with a short mixing time (e.g. 50 ms) to limit the effect of spin diffusion. Homonuclear coupling constants can be most easily measured using the PE-COSY experiment (Griesinger et al., 1987; Mueller, 1987). The PE-COSY experiment can be used to obtain values for coupling constants by measurement of the peak-to-peak separation between passive-coupled spins; in cases where measurement of coupling constants in this manner is not possible, coupling constants can be measured by analysis of peak to peak separations in active-coupled spins (Kim & Prestegard, 1989; Kim & Prestegard, 1990), as discussed in Appendix 2.

Stereospecific assignment of the phosphopeptide Hβ protons was not possible for Asp -3, Asn -2, Asp -1 and Leu +4 because of the degeneracy or near-degeneracy of the Hβ protons. From the PE-COSY experiment, $^3J_{H\alpha H\beta}$ coupling constants for Asp +6 and Lys +8 were measured (see Table 3) and found to be indicative of conformational averaging about $\chi_1$, so these residues were not considered further.

With respect to the pTyr residue, there are two Hβ protons, resonating at 3.19 and 2.83 ppm. As shown in Figure 8b, the $\alpha/\beta$ NOE was stronger to the Hβ proton at 3.19 ppm and the NH/β NOE was stronger to the Hβ proton at 2.83 ppm. In addition, in the COSY and the PE-COSY, the coupling constant between the $\alpha$ and $\beta$ proton at 3.19 was vanishingly small, such that the $\alpha/\beta$ crosspeak was not present. Small coupling constants lead to cancellation of antiphase absorptive signals when linewidths are large (Wüthrich, 1986): it was observed that two protons with a $^3J_{H\alpha H\beta}$ of $> 4.4$ Hz (the coupling constant measured for the Pro +5 coupling between the Hα and Hβ2 protons, see Table 2) gave rise to a weak crosspeak in PE-COSY spectra; thus the absence of a crosspeak for two scalar coupled spins is indicative of the $^3J$ value to be small (i.e. $< 4$ Hz). Using the method of Kim & Prestegard (1989), the active coupling at the $H\alpha/H\beta$ crosspeak involving
Table 2: Coupling Constants Measured from the Double-Filtered PE-COSY of the PLCC SH2 domain/pY1021 Complex (Hz)

<table>
<thead>
<tr>
<th></th>
<th>Pro +3</th>
<th>Pro +5</th>
<th>Asp +6</th>
<th>Pro +7</th>
<th>Lys +8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{3}JH\alpha\beta_a$</td>
<td>12.3</td>
<td>7.3</td>
<td>5.9</td>
<td>8.8</td>
<td>6.3</td>
</tr>
<tr>
<td>$^{3}JH\alpha\beta_b$</td>
<td>&lt; 3</td>
<td>4.4</td>
<td>10.2</td>
<td>5.9</td>
<td>8.8</td>
</tr>
</tbody>
</table>

Note -- $\beta_a$ and $\beta_b$ refer to the more downfield and upfield resonances, respectively.
Figure 8 -- Stereoassignment of Hβ Protons of the pTyr Residue.

a. Newmann projections of the three staggered rotamers about the Cα-Cβ bond in amino acids containing two Hβ protons, with the values of the $^3J_{H\alpha H\beta}$ coupling constants expected among the three rotamers.  
b. Table summarizing data concerning the pTyr residue and its $\chi_1$ dihedral angle.
a. 

\( \chi_1 = 60^\circ \) 

\( \chi_1 = 180^\circ \) 

\( \chi_1 = -60^\circ \)

\( ^3J_{H\alpha H\beta 1} < 4 \text{ Hz} \) 

\( ^3J_{H\alpha H\beta 1} > 10 \text{ Hz} \) 

\( ^3J_{H\alpha H\beta 2} < 4 \text{ Hz} \) 

\( ^3J_{H\alpha H\beta 2} > 10 \text{ Hz} \)

b. 

\( \delta_{H\beta} \) 

\( \beta_a \)

3.19

\( \beta_b \)

2.83

H\( \alpha / H\beta \) NOE 

50 ms NOESY 

\( H\beta_a \gg H\beta_b \)

150 ms NOESY 

\( H\beta_a > H\beta_b \)

NH/H\( \beta \) NOE 

150 ms NOESY 

\( H\beta_b > H\beta_a \)

Coupling Constants 

H\( \alpha / H\beta \) COSY 

crosspeak 

J\( \alpha \beta \) (Hz) 

< 4 

12.5* 

assignment 

H\( \beta_1 \) 

H\( \beta_2 \)

\( \chi_1 \) rotamer 

-60.0

the Hβ resonance at 2.83 ppm was measured to be 12.5 Hz. (The broadening of the Hα proton yielded poor signal intensity for this Hα/Hβ crosspeak in the PE-COSY, so the passive 3JHαHβ could not be measured.) This information indicates that the 3.19 ppm resonance can be assigned as β3, that the 2.83 ppm resonance can be assigned as β2 and that χ1 is near the staggered rotamer χ1 = -60. This was further confirmed using the program Stereosearch (Nilges et al., 1990) (see Appendix 2). This χ1 conformation is consistent with the crystal structures of phosphopeptide complexes of the src (Gilmer et al., 1994; Waksman et al., 1992; Waksman et al., 1993), lck (Eck et al., 1994, Eck et al., 1993) and syp N-terminal (Lee et al., 1994) SH2 domains, in which the χ1 of pTyr was found to be -75° +/- 20°.

For Ile residues, stereospecific assignment of Hβ protons is not necessary since there is only one β proton. However, by comparing relative strengths of NOEs involving the NH, Hα, Hβ and Hγ2 methyl resonances and measuring 3JHαHβ, one may obtain information about the dihedral angle χ1. Newman projections about an Ile Cα-Cβ bond are shown in Figure 9 as well as the distance relationships and the 3JHαHβ coupling constant expected for each rotamer. Note that Newman projections of Figure 9 are identical to those of Figure 8 if one substitutes the β2 and β3 protons in Figure 8 with the β proton and the γ methyl group in Figure 9, respectively. For Ile +1, the 3JHαHβ was assumed to be < 4 Hz based on the lack of an Hα/Hβ crosspeak in any of the F1/F2-filtered COSY, PE-COSY and HOHAHA experiments. Based on the small 3JHαHβ and the NOE intensities given in Figure 9, χ1 was restrained to 60° ± 30°. For Ile +2, the presence of an α/β crosspeak in the F1/F2-filtered PE-COSY and COSY experiments provided evidence that 3JHαHβ2 was > 4 Hz, but its measurement via a passive coupling in the PE-COSY was not possible because there is only one Hβ proton so the α/β crosspeak contains only the active coupling. However, the coupling constant could be measured as described in Kim & Prestegard (1989), yielding a value of 14.5 Hz. This value is larger than one
Figure 9 -- Conformations About $\chi_1$ for Ile Residues.  

a. Newmann projections for the three staggered rotamers of Ile residues about the $C_\alpha$-$C_\beta$ bond, with values of the $^{3}J_{H\alpha H\beta}$ coupling constants expected for the three rotamers.  

b. Table summarizing data for the Ile +1 and Ile +2 residues and their $\chi_1$ angles.
\( \chi_1 = 60^\circ \)  
\( \chi_1 = 180^\circ \)  
\( \chi_1 = -60^\circ \)

\( 3J_{\text{H} \alpha \text{H} \beta} < 4 \text{ Hz} \)

\( 3J_{\text{H} \alpha \text{H} \beta} < 4 \text{ Hz} \)

\( 3J_{\text{H} \alpha \text{H} \beta} > 10 \text{ Hz} \)

b.

<table>
<thead>
<tr>
<th></th>
<th>Ile +1</th>
<th></th>
<th>Ile +2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H( \beta )</td>
<td>H( \gamma_2 )</td>
<td>H( \beta )</td>
</tr>
<tr>
<td>H( \alpha )-NOE</td>
<td>S</td>
<td>S</td>
<td>M</td>
</tr>
<tr>
<td>50 ms NOESY</td>
<td>S</td>
<td>S</td>
<td>M</td>
</tr>
<tr>
<td>150 ms NOESY</td>
<td>M</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>NH-NOE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150 ms NOESY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 3J_{\text{H} \alpha \text{H} \beta} ) (Hz)</td>
<td>(&lt; 4 )</td>
<td></td>
<td>14.6*</td>
</tr>
<tr>
<td>( \chi_1 ) rotamer</td>
<td>60</td>
<td></td>
<td>-60</td>
</tr>
</tbody>
</table>

* Measured by the method of Kim and Prestegard (1989)
would expect. However, measurement of peak-to-peak separations for absorptive and dispersive lineshapes for this residue is hampered by the closeness of the H\text{\textalpha} chemical shift to the water resonance with its dispersive signal. The dispersive lineshape of the water resonance interferes with measurement of the peak to peak separation in the H\text{\textalpha}/H\text{\textbeta} crosspeak, likely leading to an overestimate of the true coupling constant. The presence of this large coupling constant for $^{3}J_{H\text{\textalpha}H\text{\textbeta}}$ is consistent with NOE information given in Figure 9, leading to assignment of the $\chi_{1}$ to the -60° rotamer.

Proline residues can adopt one of two ring conformations in crystal and solution structures, or average between the two forms (Madi et al., 1990; Thomasson & Applequist, 1990). In solution, transitions between the two ring conformations are generally fast on the NMR time scale. A number of calculations have shown that the energy barrier between the two ring conformers is small (Thomasson & Applequist, 1990), and molecular dynamics simulations demonstrated transitions between the two ring geometries occurring on the picosecond time scale (Brunne et al., 1993). These ring forms consist of nearly coplanar N, C\text{\textalpha}, C\text{\textbeta} and C\text{\delta} atoms with the C\text{\gamma} atom either being above (C\text{\gamma} or N-type) or below (C\text{\gamma} or S-type) the plane of the ring. The geometry of these ring conformers is shown in Figure 10. Distance relationships are difficult to use in distinguishing between these two ring forms, in which the H\text{\textalpha}/H\text{\textbeta} distance is 2.7 Å and 3.0 Å for the N and S conformer, respectively. This may be significant in NOESY spectra with short mixing times (50 ms) since NOEs involving protons separated by a distance of > 3 Å become very weak in these spectra. Coupling constant information is more diagnostic in distinguishing ring conformers. The dihedral angle $\chi_{1}$ (involving the atoms N, C\text{\textalpha}, C\text{\beta} and C\text{\gamma}) has a value of approximately -30° or 30° for the N- and S-type, respectively, giving rise to values for $^{3}J_{H\text{\textalpha}H\text{\beta}3}$ values of approximately 1 Hz in the S-type and > 10 Hz in the N-type conformers, respectively (Cai et al., 1995; Madi, et al., 1990).

For Pro +3 and Pro +5, larger NOEs were observed in both the 50 and 150 ms NOESY spectra between the H\text{\textalpha} proton and the more downfield H\text{\beta} proton (at 2.13 and
Figure 10 -- Proline Ring Conformations in the pY1021 Peptide. The proline ring in (a) N- and (b) S-type conformations. (c) Hα/Hβ distances and coupling constants useful in distinguishing between the two ring conformers. (d) $^{3}J_{H\alpha H\beta}$ values measured for the three Pro residues in the pY1021 peptide in complex with the PLCC SH2 domain.
Values for Pro Residues in the pY1021 Peptide in Complex with the PLCC SH2 Domain

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>H(\alpha)/H(\beta) distance (Å)</td>
<td>2.41</td>
<td>2.33</td>
</tr>
<tr>
<td>H(\alpha)/H(\beta)2 distance (Å)</td>
<td>3.04</td>
<td>2.71</td>
</tr>
<tr>
<td>(\chi_1) dihedral angle (°)</td>
<td>-30</td>
<td>+30</td>
</tr>
<tr>
<td>(N-C(\alpha)-C(\beta)-C(\gamma))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(^3)J(\alpha)(\beta)1 (Hz)</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>(^3)J(\alpha)(\beta)2 (Hz)</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

2.26 ppm for Pro +3 and Pro +5, respectively). Therefore these resonances were assigned to the Hβ3 proton of these residues and the more upfield protons to Hβ2. The NOE involving the Hα and Hβ2 protons was always weaker for Pro +3 than Pro +5, and this Pro +3 NOE was at the level of the noise in the 50 ms NOESY. In addition, Pro +3 yielded no observable crosspeak between the Hα and the Hβ2 proton in the double-filtered HOHAHA, COSY and PE-COSY spectra, and in the PE-COSY the \(^3J_{HαHβ2}\) measured from the passive coupling in the Hα/Hβ3 crosspeak appeared to be quite small (< 3 Hz). Using the double-filtered COSY experiment and the method of Kim & Prestegard (1989), a value for \(^3J_{HαHβ3}\) of 12.3 Hz was measured. Of note was the unusual appearance of this crosspeak in the PE-COSY, probably due to the tight coupling which exists between Hβ1 and both Hγ proton resonances. Thus \(^3J_{HαHβ2}\) is small and that this residue adopts exclusively an S-type conformation. For Pro +5, a \(^3J_{HαHβ2}\) of 4.5 Hz was measured, indicative of conformational averaging weighted towards the S conformer. In fact, using a value of 0.8 Hz for the \(^3J_{HαHβ2}\) in the S conformation and a value of 10.8 Hz for the \(^3J_{HαHβ2}\) in the N conformation (calculated from data in Brunne et al., 1993), Pro +5 is estimated to be in the S conformation roughly 2/3 of the time.

For the three N-terminal and three C-terminal residues in this phosphopeptide, stereospecific assignment and NOE determination was difficult due to the presence of residual COSY-type peaks in the NOESY spectra, particularly in the 50 ms data set. This may be indicative of a more rapid correlation time in these positions, thus decreasing the intensity of the resulting NOEs. However, for Pro +7, the coupling constants of 5.9 and 8.8 Hz are virtually identical to those observed for Pro 7 of the peptide antamanide (Brunne et al., 1993; Madi et al., 1990), in which the two conformers are estimated to be in a nearly 1:1 ratio. In such a situation, the \(^3J_{HαHβ2}\) coupling constant is 5.9 Hz. Based on this correspondence, I have assigned the proton at 2.02 ppm to the Hβ2 proton and the proton at 2.26 ppm to the Hβ3 proton, respectively.
IV. Summary:

Using $^{13}$C, $^{15}$N labeling of the PLC<sub>C</sub> SH2 domain and chemical synthesis of the phosphopeptide, we were able to assign resonances for and study the conformation of the pY1021 phosphopeptide when bound to the SH2 domain. The pattern of sequential NOEs indicates that the phosphopeptide adopts an extended $\beta$-conformation from residues Asp -1 to Lys +8, however chemical shifts indicate that the $\beta$-conformation is restricted from the pTyr residue to Leu +4, with other residues adopting a random coil conformation. No long-range peptide-peptide NOEs were observed, which is also diagnostic of the pY1021 phosphopeptide adopting an extended conformation. NOE and coupling constant data indicate restriction of $\chi_1$ for residues pTyr, Ile +1 and Ile +2. The conformation of pTyr about $\chi_1$ is observed to be similar to that found for other SH2 domains. In addition, the Pro +3 ring was found to be fixed in the S-type (or C$\gamma$) conformer. Finally, a strong NOE was observed between the Ile +1 H$\delta$ methyl protons and the Pro +3 ring. This indicates that the +1 and +3 binding sites are close in space.
Chapter 4: SH2 Domain-Phosphopeptide Contacts and Structure Calculations

I. Introduction

SH2 domains bind to specific sites of tyrosine phosphorylation in a sequence-specific manner. In the introductory chapter, I discussed the sequence specificity of the PLCN and PLCC SH2 domains, in particular with regards to the degenerate phosphopeptide libraries of Sonyang et al. (1992). Sequence specificity of SH2 domains has usually been studied only within the +1 to +3 region. This is in part because preliminary experiments with degenerate phosphopeptide libraries seemed to indicate no further sequence specificity beyond the +3 position. However, mutation of Leu +4 in the pY1021 phosphopeptide to a Ser abolished in vitro binding of the pY1021 phosphopeptide to PLC-γ (Larose et al., 1993). Binding studies have indicated the importance of the +5 residue in the syp N-terminal SH2 domain (Kay et al., 1997) and the +4 residue in the PLCN SH2 domain (personal communication). Smaller effects on binding of the PLCC SH2 to the pY1021 phosphopeptide by removal of the +4 and +5 residues have also been shown from these studies. In addition, the structure of a couple of SH2 domains in complex with phosphopeptides (Hatada et al., 1995; Lee et al., 1994) have shown specific contacts involving residues C-terminal to the +3 residue.

Thus the PLCC SH2 domain has a number of very interesting features concerning its binding to high affinity phosphopeptides. In order to understand the mechanism for this specificity, we must be able to identify contacts between the SH2 domain and the pY1021 phosphopeptide, as well as other phosphopeptides at a future date. In solution NMR, observation of NOEs allows one to show that two protons are close in space (< 5 Å) (see Chapter 1). As stated previously, NOEs observed in the PLCC SH2/pY1021 complex can be of three types: protein-protein, peptide-peptide, and protein-peptide. These NOEs can
be separated by differential labeling of the protein and peptide, with the SH2 $^{15}$N, $^{13}$C-labeled and the peptide unlabeled. A number of experiments have been developed which begin with magnetization from protons attached to $^{13}$C or $^{15}$N nuclei, followed by an NOE mixing period, followed by purging magnetization from $^{13}$C or $^{15}$N-attached protons (Burgering et al., 1993; Wider et al., 1990; Wider et al., 1991). These are the so-called F$_1$-edited/F$_2$(or F$_3$)-filtered (or half-filtered) experiments used to select exclusively for intermolecular (i.e. protein-peptide) NOEs. A number of modifications have been made to these experiments, including use of pulse field gradients to remove artifacts and aid in coherence pathway selection, minimization of the time which the magnetization stays in the transverse plane, and selection of a range of J$_{13CH}$ values (Lee et al., 1994).

This chapter will discuss data from these F$_1$-edited/ F$_2$-filtered experiments on the PLCC SH2/pY1021 complex, as well as results of structure calculations performed with the program X-PLOR (Brünger, 1992). The data will be presented in a residue by residue fashion with respect to the phosphopeptide. The results of these calculations will be compared to structures of other SH2 domain/high affinity phosphopeptide complexes solved to date. Most of the data presented in this chapter has been published previously (Pascal et al., 1994), however more recent structure calculations are reported here, and these structure calculations will be the basis for a report of the refined structure of the PLCC SH2/pY1021 complex (Pascal et al., 1997). In addition, the reported solution structure will be rationalized in terms of binding data on the PLCC SH2/pY1021 complex (Kay et al., 1997) plus additional binding data on the PLCN SH2 domain/pY1021 complex (Steve Shoelson, personal communication), both shown in Table 2.

II. Experimental Methods

NMR experiments were collected on a Varian UNITY + 500 MHz spectrometer equipped with z gradients. Experiments were processed using in-house and commercially
available routines, including the NMRPipe system (Delaglio et al., 1995), and analyzed using the program PIPP and STAPP (Garrett et al., 1991).

The majority of side chain assignments of the PLCC SH2 resonances was performed by Dr. Steven Pascal, with the exception of assignment of aromatic amino acids, which was performed by Dr. Toshio Yamazaki (Yamazaki et al., 1993a) and procedures used for assignment have been described elsewhere (Pascal et al., 1994). Of note in the assignment was that stereoassignment of Hα protons of Gly residues and Hβ protons of several residues was accomplished using a combination of measurement of coupling constants, NOE intensity from a 50 ms CN-NOESY-HMQC experiment (Pascal et al., 1994) and input of this information to the grid-search program Stereosearch (Nilges et al., 1990). 3JHNHa couplings were estimated by lineshape fitting of 1D traces through a 15N-1H HMQC-J spectrum (Kay & Bax, 1990). 3JHBN coupling constants were measured by crosspeak intensities of a HNHB spectrum (Archer et al., 1991), and 3JHBC and 3JHαHβ coupling constants were measured from cross-peak separations due to passive couplings in SOFT-HCCH-COSY and SOFT-HCCH-ECOSY spectra, respectively (Eggenberger et al., 1992). In addition, stereoassignment of methyl groups of Val and Leu residues was accomplished by the method outlined in Neri et al. (1989). Measurement of χ1 of Val and χ2 of Leu and Ile is described in Kay et al. (1996). The entire SH2 domain was completely assigned with the exception of the N-terminal Gly 1 and Ser 2 residues, as NH correlations for these residues were not observed in 15N-1H HSQC spectra. SH2 domain/SH2 domain NOEs were predominantly assigned by Steve Pascal, although I assigned a number of NOEs involving His imidazole groups and Arg guanidinium groups described in Chapters 5 and 6.

Assignment of pY1021 resonances bound to the protein was performed by me as described in Chapter 3. Assignment included stereoassignment of Hβ protons of pTyr, Pro +3, Pro +5 and Pro +7 as well as determination of χ1 restraints for pTyr, Ile +1 and Ile +2.
Half-filtered NOESY experiments were used to obtain protein-peptide NOEs using a 1.8 mM sample of $^{13}$C$^{15}$N-labelled PLCC SH2/pY1021 complex dissolved in D$_2$O, pH 6.4, in 100 mM sodium phosphate, to which pY1021 phosphopeptide was added until a 1:1 SH2/pY1021 complex was achieved (also used for peptide assignment, see Chapter 3). The experiment used for obtaining $^{13}$CH/$^{12}$CH NOEs is described in Lee et al. (1994). This experiment was modified so that it could be used for samples dissolved in H$_2$O to obtain protein-peptide NOEs involving NH and/or $^{15}$NH protons; modifications included selection for $^{15}$N-attached and $^{13}$C-attached proton magnetization and recording $^{15}$N and $^1$H chemical shift concurrently (Pascal et al., 1994). 2D versions of this experiment were performed on $^{15}$N and $^{13}$C,$^{15}$N-labelled samples in H$_2$O, however no definite protein-peptide NOEs involving NH protons (either on the protein or peptide) could be identified. Parameters for the 3D half-filtered experiment include 62 and 32 experiments in $t_1$ and $t_2$ respectively, 1024 complex points in $t_3$, 32 transients, purging pulses selected for $J_{HC}$ values of 125-140 Hz, and the $^{13}$C channel transmitter set to 43.0 ppm. Initially, a 350 ms mixing time was used, however for use of these NOEs in a semi-quantitative manner in structure calculations, another spectrum was recorded using a mixing time of 150 ms. 2D versions of these experiments were initially run to test for conditions, and were recorded with proton or $^{13}$C-detection in the indirect dimension. Parameters used for the 2D experiments include 96-192 scans, a sweep width of 8000 Hz, purging pulses selected for $J_{HC}$ values of 125-140 Hz or 160-200 ms, and a 350 ms mixing time; for $^1$H measurement in the indirect dimension, a sweep width of 4750 Hz was used in the indirect dimension, and 196-256 increments were collected, and for $^{13}$C measurement in the indirect dimension, a sweep width of 3000 Hz was used and 32 increments were collected.

Calibration of NOE intensity using NOEs involving fixed known interproton distances is often used to more accurately assign distance ranges. This cannot be done for the half-filtered NOESY experiments, in that there is no protein-peptide distance upon which to base these calibrations. For this reason it was important that many of the protein-
peptide NOEs observed in the half-filtered experiment were observed in a 3D $^{13}$C-NOESY experiment. Parameters for this experiment include 128 and 32 complex points in $t_1$ and $t_2$ respectively, sweep widths of 5849.7, 1650 and 8000 Hz in $t_1$, $t_2$ and $t_3$ respectively in which the $^{13}$C offset was set to 42 ppm, 1024 complex points in $t_3$ and a mixing time of 150 ms.

Structure calculations were performed using the program X-PLOR (Brünger, 1992) using a distance geometry/simulated annealing protocol very similar to that described in Nilges et al. (1988). In these calculations, starting structures are obtained by an embedding procedure using a subset of atoms (Crippen & Havel, 1988). The remaining atoms are added in an extended conformation, followed by molecular dynamics at high temperature (1000° K) with gradual cooling to a final temperature of 100° K, and a final energy minimization. All of our X-PLOR calculations use energy terms for van der Waals repulsion and covalent geometry as well as experimental restraints, but do not include a full Leonard-Jones potential or terms for the electrostatic energy. A few changes were introduced to the standard protocol, including the creation of parameters of pTyr by combination and modification of parameters for tyrosine and those of phosphate derived from nucleic acid parameters. In these calculations, a total of 1874 distance restraints derived either from NOEs or detection of hydrogen bonding were used, including 94 protein-peptide and 151 peptide-peptide NOE restraints. Included in the hydrogen-bonding restraints were six involving the Arg 37 ε12, Arg 37 ε22 and Arg 39 ε protons with any one of the four pTyr phosphate oxygens; the rationale and parameters for these hydrogen bonds will be explained in detail in Chapter 6. $\chi_1$ dihedral restraints for pTyr, Ile +1 and Ile +2 were obtained for the PLCC SH2 as described in the previous chapter, as well as 194 dihedral restraints for $\phi$, $\psi$, $\chi_1$ and $\chi_2$ for the SH2 domain. Centre-averaging was employed for non-stereoassigned methylene protons, aromatic rings and methyl groups, with the appropriate corrections made to the NOE values as described in the X-PLOR manual (Brünger, 1992).
Structures calculated were compared to the crystal structures or solution structures of SH2 domains either free or in complex with phosphopeptides whose coordinates are now publicly available. These include the src SH2, free (Waksman et al., 1993; PDB filecode 1SPR), in complex with low-affinity phosphopeptides (Waksman et al., 1992; filecodes 1SHA and 1SHB) and high-affinity phosphopeptides (Waksman et al., 1993; filecode 1SPS; Xu et al., 1995; filecode 1SHD), the crystal structure of the lck SH2 in complex with a high-affinity phosphopeptide (Eck et al., 1993; filecode 1LCJ), the crystal structure of the lck SH2 and SH3 domain expressed as a single polypeptide in complex with a low-affinity phosphopeptide (Eck et al., 1994; filecode 1LCK), the crystal structure of the N-terminal SH2 of syp, both free (filecode 1AYD) and in complex with three different high-affinity phosphopeptides (Lee et al., 1994; filecode 1AYA, 1AYB and 1AYC), the crystal structure of the free SHC SH2 domain (Mikol et al., 1995; filecode 1MIL), the crystal structure of the whole GRB2 molecule (Maignan et al., 1995; filecode 1GRI), the solution structures of the free abl SH2 domain (Overduin et al., 1992a; Overduin et al., 1992b; filecode 1AB2) and N-terminal SH2 of p85 (Booker et al., 1992; filecode 2PNB) SH2 domain, and the solution structures of a src-pYEEI (Gilmer et al., 1994; filecode 1HCT) and syk C-terminal SH2/pYEEI complex (Narula et al., 1995; filecode 1CSZ). These structures will be referred to in this chapter and subsequent chapters as well.

III. Results and Discussion

A. Initial Experiments

Two 3D half-filtered experiments were collected, the first with a mixing time of 350 ms, from which we observed a large number of protein-peptide NOEs. However, since NOE intensity could be affected significantly by spin diffusion at this mixing time, a second 3D was collected with a mixing time of 150 ms. These NOEs then could be
assigned based on their intensities as strong, medium or weak and used in structure calculations. NOEs from the 150 ms experiment were assigned based on the chemical shift tables of the peptide (see Chapter 3) and the protein (assigned principally by Steve Pascal, with assignment of aromatic resonances by Toshio Yamazaki (Yamazaki et al., 1993a). The results are discussed with respect to protein residues in contact with each residue of the phosphopeptide, beginning with the pTyr residue, then the +1 Ile and +3 Pro residues, to which numerous NOEs were also found, followed by the remaining residues to which much less NOE information was obtained. The discussion will focus both on the results from the half-filtered NOESY experiments as well as from structure calculations. Structure calculations were performed both by myself and by Steve Pascal, though Steve Pascal calculated the final structures reported here.

As stated previously, use of NOEs in half-filtered experiments in a semi-quantitative manner is more difficult because there are no known distances in which to scale NOE intensity to. Thus I employed the following procedure; the largest inter-residue NOE not involving a methyl group (Cys 58 Hα/pTyr Hδ with intensity of 5.25 X 10^5) was speculated to involve two protons that are in van der Waals contact (2.2 Å). Note that this NOE in principle involves contributions from two pTyr Hδ protons, however structures of other SH2 domains indicate that the pTyr ring interacts with the backbone of the large β-sheet (see Figure 1) in an edge-on manner, such that only one of the two Hδ protons would be in van der Waals contact. Since the weakest NOEs are approximately 10-fold less intense, I scaled the NOE intensity by a factor of r^-3 such that NOE intensities of > 2.0 X 10^5 defined distances of < 3.0 Å, NOE intensities between 7.0 X 10^4 and 2.0 X 10^5 defined distances of < 4.3 Å, and weaker intensity NOEs were defined as weak NOEs and were given an upper distance bounds of 6.0 Å. Lower distance bounds for NOEs were set to 1.8 Å, or 0.8 X the sum of the VDW radius of two protons, in all cases. This conservative approach with semi-quantitative use of the NOE information led to calibration values for the scaling of the protein-peptide NOEs which were compatible with the NOE
intensities of protein-protein NOEs of known distance and NOE intensities of protein-peptide NOEs observed in the 3D $^{13}$C-NOESY.

B. The Structure of the PLCC SH2 Domain/pY1021 Complex

1. Overall Features

In Table 1 is a summary of the number of NOE restraints and the goodness of fit. In general, the structures showed few deviations from ideal geometry and low NOE energies indicative of a reasonable fit of the data. Figure 1a shows a superposition of the 25 lowest energy structures calculated for the PLCC SH2/pY1021 complex and Figure 1b shows the overall fold of the PLCC SH2/pY1021 complex based on an energy-minimized average structure. In well-ordered regions of the protein, calculated structures vary little from one another, i.e. in the 25 lowest energy structures, an rmsd of $0.62 \pm 0.07$ Å for backbone heavy atoms and an rmsd of $1.13 \pm 0.09$ Å rmsd for all heavy atoms are found, excluding residues 1-9 and 100-105 of the SH2 domain and 1-2 and 11-12 of the phosphopeptide which are disordered in solution. Furthermore, if one considers only regions of secondary structure, the rmsd falls to $0.43 \pm 0.05$ for the backbone and $0.91 \pm 0.08$ for heavy atoms. The overall backbone fold of the PLCC SH2 domain consists of a large three-stranded and small three-stranded β-sheet flanked by two helices. N- and C-terminal residues in the ordered region of the SH2 domain hydrogen-bond to the first complete strand of the large β-sheet, forming a fourth strand for the large β-sheet. This fold is identical to that observed for other SH2 domains (Breeze et al., 1996; Hatada et al., 1995; Maignan et al., 1995; Narula et al., 1995; Nolte et al., 1996; Rahuel et al., 1996; Schaffhausen, 1995; Zhou et al., 1995). In describing the elements of secondary structure in this SH2 domain, we will henceforth refer to the nomenclature of Eck et al. (1993). Sequence correspondence between the residue number from the construct and this nomenclature is as
Table 1: Structural Statistics for the PLCC SH2/pY1021 Solution Structure Determination

<table>
<thead>
<tr>
<th>Restraints</th>
<th>no.</th>
<th>&lt;SA&gt;</th>
<th>&lt;SAF&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>all distance restraints</td>
<td>1874</td>
<td>0.020 ± 0.002</td>
<td>0.018</td>
</tr>
<tr>
<td>protein-protein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>inter-residue NOEs</td>
<td>336</td>
<td>0.012 ± 0.003</td>
<td>0.012</td>
</tr>
<tr>
<td>sequential NOEs</td>
<td>450</td>
<td>0.020 ± 0.002</td>
<td>0.019</td>
</tr>
<tr>
<td>short-range NOEs</td>
<td>232</td>
<td>0.019 ± 0.005</td>
<td>0.012</td>
</tr>
<tr>
<td>long-range NOEs</td>
<td>527</td>
<td>0.022 ± 0.003</td>
<td>0.018</td>
</tr>
<tr>
<td>hydrogen bonds</td>
<td>75</td>
<td>0.020 ± 0.004</td>
<td>0.017</td>
</tr>
<tr>
<td>protein-peptide NOEs</td>
<td>151</td>
<td>0.025 ± 0.005</td>
<td>0.023</td>
</tr>
<tr>
<td>*protein-peptide hydrogen bonds</td>
<td>6</td>
<td>0.045 ± 0.025</td>
<td>0.045</td>
</tr>
<tr>
<td>peptide-peptide NOEs</td>
<td>94</td>
<td>0.016 ± 0.006</td>
<td>0.012</td>
</tr>
<tr>
<td>Dihedral restraints (degrees)</td>
<td>197</td>
<td>0.0455 ± 0.076</td>
<td>0.433</td>
</tr>
</tbody>
</table>

Deviations from ideal geometry

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>bonds (Å)</td>
<td>1937</td>
<td>0.009 ± 0.00004</td>
<td>0.009</td>
</tr>
<tr>
<td>angles (degrees)</td>
<td>3486</td>
<td>2.42 ± 0.004</td>
<td>2.41</td>
</tr>
<tr>
<td>impropers (degrees)</td>
<td>1026</td>
<td>1.01 ± 0.014</td>
<td>1.00</td>
</tr>
</tbody>
</table>

\frac{F_{NOE}}{kcal/mol} \quad \frac{F_{dihed}}{kcal/mol} \quad \frac{F_{repul}}{kcal/mol} \quad \frac{F_{LJ}}{kcal/mol}

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|                                |        |         |         |
|                                |        |         |         |

|                                |        |         |         |
|                                |        |         |         |

\text{rmsd's} \quad \text{backbone} \quad \text{all-atom}

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>residues 11-99</td>
<td>0.6205</td>
<td>1.1290</td>
<td></td>
</tr>
<tr>
<td>secondary structure</td>
<td>0.4309</td>
<td>0.9067</td>
<td></td>
</tr>
<tr>
<td>strands (\beta)B, (\beta)C, (\beta)D \quad \beta D'</td>
<td>0.2856</td>
<td>0.8078</td>
<td></td>
</tr>
<tr>
<td>(p)Tyr binding site</td>
<td>0.3863</td>
<td>1.0510</td>
<td></td>
</tr>
<tr>
<td>(+1) to Pro +5 binding site</td>
<td>0.3515</td>
<td>0.5076</td>
<td></td>
</tr>
</tbody>
</table>

1 Superposition of the amide nitrogen, the C\(\alpha\) and the carbonyl carbon atoms.


3 Residues 33-39, 44-51 and 55-62.

4 Residues 18, 37, 39, 46, 59 and \(p\)Tyr.

5 Residues 49, 56, 58, 69, 70, 89, 90, Ile +1 to Pro +5.
Figure 1 -- Tertiary Structure of the PLCC SH2 Domain/pY1021 Complex.

A. Superposition of Cα atoms of the 25 lowest energy structures of the PLCC SH2/pY1021 complex. Cα atoms included residues 11-99 of the SH2 domain and residues 3-10 (from -1 to the +6 position relative to the pTyr) of the phosphopeptide. The backbone of the average structure of these 25 is shown in (B) using the program Molscript. The molecule adopts a conformation consisting of two α-helices separated by large and small β-sheets. The phosphopeptide adopts an extended β-conformation at a direction perpendicular to the large β-sheet.
follows; β-strands βA, βB, βC, βD, βD', βE, βF, and βG correspond to amino acids 11-13, 33-39, 44-49, 54-59, 60-62, 66-69, 72-74 and 97-99, respectively, and helices αA and αB correspond to residues 17-27 and 77-87, respectively (Pascal et al., 1994). In Figure 2, an alignment of the sequence with secondary structure for the PLCC SH2 domain (seq 3, amino acids 1-102) and with other SH2 domains whose structures have been solved is shown. In addition, the probable correspondence of the sequence of the PLCN SH2 domain with secondary structure is shown.

Little structure is observed for residues 7-10, however NOEs involving protons in residues His 6 and Ser 8 are observed to both Trp 11 (βA1) and residues in the N-terminus of helix αB. Though the residue defining the N-terminus of the SH2 domain was originally believed to be the highly-conserved N-terminal Trp (Trp 11 in the PLCC SH2, or Trp βA1), Nolte et al. (1996) have shown that in some SH2 domains there is a conserved hydrophobic residue five residues N-terminal to this conserved Trp residue. Crystal structures of the src SH2 (Gilmer et al., 1994), the lck SH2/SH3 (Eck et al., 1994) and the p85 C-terminal SH2 domain (Nolte et al., 1996) demonstrate that this conserved residue packs into a similar region as observed for His 6 in the PLCC SH2 (this residue is starred in Figure 2). However, this residue does not play a role in hydrophobic packing for all SH2 domains; in GRB2, the linker separating the N-terminal SH3 and SH2 domains is very small, such that the amino acid five residues N-terminal to the conserved Trp (a Met) is a part of the C-terminal β-strand of the SH3 domain (Maignan et al., 1995).

As observed for most SH2/phosphopeptide complexes (Kuriyan & Cowburn, 1993, Schaffhausen, 1995), the phosphopeptide forms a β-strand conformation in a direction orthogonal to the large β-sheet (see Figure 1b). The position at which the phosphopeptide crosses the backbone of the large β-sheet (along the βD strand) corresponds to the peptide bond between the pTyr and the +1 residue of the phosphopeptide, with residues C-terminal to the pTyr binding on one side of the β-sheet, while residues N-terminal to and including the pTyr lie on the other side.
Figure 2 -- Sequence Alignment and Secondary Structure of SH2 Domains. Structures of all SH2 domains have been solved with the exception of the PLCN SH2 Domain. Nomenclature of the secondary structural elements is the same as that used by Eck et al. (1993). In addition, the position five residues N-terminal to Trp βA1 is starred: this residue is hydrophobic and has been shown by Nolte et al. (1996) to interact with the N-terminus of helix αB. Regions of sequence similarity are shaded.
Figure 3 -- Surface and Electrostatic Potential of the PLCC SH2 Domain in the PLCC SH2 Domain/pY1021 Complex. Surface rendering of the energy-minimized rms-average structure of the PLCC SH2 domain in complex with the pY1021 phosphopeptide. Electrostatic contours of the SH2 domain are displayed, showing the presence of a large positively-charged pocket into which the pTyr residue (and possibly the Asp -1 residue) fit, as well as a hydrophobic channel into which residues +1 to +6 bind. The peptide is shown as a licorice model. The figure was generated by Dr. Steve Pascal using the program GRASP (Nicholls et al., 1991).
The surface and electrostatic potential of the PLCC SH2 relative to the position of the pY1021 phosphopeptide is shown in Figure 3. The pTyr residue binds in a large positively-charged pocket containing four Arg residues. This positively-charged phosphate-binding pocket has been observed in the crystal structures of all of the SH2/phosphopeptide complexes solved to date. Further interactions are observed between the SH2 domain and residues C-terminal to the pTyr, which bind in a large groove beginning near the large β-sheet and continuing between the EF and BG loops. This binding groove is large enough to at least partially bury residues +1 to +6 of the phosphopeptide, and nearly completely bury residues Ile +1 and Pro +3. Structure calculations showed greater variability in the pTyr-binding pocket than the hydrophobic binding pocket; an rmsd of 1.05 Å was observed for the pTyr binding site as opposed to an rmsd of 0.51 Å for the +1 to +3 binding site. Lack of NOEs for the phosphate group and mobility of the Arg residues accounts for much of the imprecision in the pTyr-binding site, and this will be discussed in greater detail in Chapter 6. The shape of the hydrophobic binding pocket is much larger than that observed in Group I SH2 domains (such as the src and lck SH2 domains), which is only large enough to accommodate the side chain of the +3 residue, with +1 and +2 residues making electrostatic interactions with charged groups on the surface of the SH2 domain. Thus the overall folds of the src and PLCC SH2 domains are very similar but the surface of the binding site is very different. Finally, in the BG loop are two positively-charged residues (Arg 91 and Lys 92) which may be in position to interact with negatively-charged amino acids on the phosphopeptide, and thus may be interacting with Asp +6 of the pY1021 phosphopeptide.

Recently, the solution structure of a complex of the SHC SH2 domain with a high-affinity phosphopeptide has been reported (Zhou et al., 1995), and in this structure, NOEs are observed in the small β-sheet with the phosphopeptide consistent with a different localization of the peptide than described for the PLCC-pY1021 structure. These include
residues βD7, βE2, βE3 and βF3 and the phosphopeptide; these interactions are not observed in the PLCC SH2/pY1021 structure.

I have previously discussed the PTB domain, which also binds to pTyr-containing phosphopeptides with high specificity (see Chapter 1). Structures of PTB domains have been solved (Zhou et al., 1996; Zhou et al., 1995) and both the fold (which is very similar to PH domains) and specifics of phosphopeptide binding are very different from SH2 domains. For example, the PTB domain binds residues N-terminal to the pTyr, in which the "NPXY" motif (where Y represents the pTyr residue) forms a turn. Further recognition occurs via hydrophobic binding of residues up to position -8 of the phosphopeptide. Recognition of pTyr also involves binding by positively-charged amino acids in the PTB domain, however the positions of these residues are not conserved, unlike the conservation of Arg αA2 and the FLVR sequence in SH2 domains. Because recognition by PTB domains is completely different from that observed in SH2 domains, it will not be discussed any further.

In Figure 4, the number of protein-peptide NOEs observed for each residue of the phosphopeptide are plotted, except for Asp -3, Asn -2 and Lys +8, for which no protein-peptide NOEs were observed. From this, one observes that the largest number of protein-peptide NOEs were found for pTyr, Ile +1 and Pro +3. In the following sections, I will discuss the protein-peptide interactions in greater detail with respect both to the results from the half-filtered experiments and structure calculations, residue by residue.

2. Contacts to Asp -1 and pTyr

In the half-filtered NOE experiment, many NOEs were observed between the SH2 domain resonances and the pTyr residue. However, only one medium-intensity protein-peptide NOE was observed involving protons of residues N-terminal to the pTyr residue, namely involving the Asp -1 Hα and the Glu 22 Hγ. The presence of this NOE is some-
Figure 4 -- Protein-Peptide NOEs as a Function of Peptide Residue. NOEs involving the pTyr, Ile +1 or Pro +3 residues account for > 75% of the total protein-peptide NOEs.
what surprising in that it places two negatively charged residues somewhat close in space. However, Chapter 6 will deal with the position of the Arg residues, and will present evidence that the Asp -1 residue interacts with the Arg 18 (Arg αA2) guanidinium group. Because of the lack of protein-peptide NOEs and paucity of peptide-peptide NOEs involving residues Asp -3 and Asn -2, structure calculations have shown that these residues are disordered in solution.

Strips showing NOEs involving pTyr resonances are shown in Figure 5, and a summary of the residues of the PLCC SH2 domain that are observed to be close in space to protons on the Asp -1 and pTyr residues are shown diagrammatically in Figure 6. Protons involved in these NOEs are on one side of the large β-strand, including specific protons of residues Arg 39 (Arg βB7), Ala 46 (Ala βC3), His 57 (His βD4), Cys 58 (Cys βD5) and Arg 59 (βD6). As seen from Figure 5, most of the protein-peptide NOEs involving the pTyr residue are to protons in the aromatic ring (Hβ and Hε); in the 150 ms half-filtered NOESY approximately 80% of the protein NOEs to the pTyr involve the aromatic protons (see Figure 7). A clear exception to this is the Cys 58 (Cys βD5) Hα, which shows strong to medium intensity NOEs with the α, β, δ and ε protons of the pTyr (see Figure 5). Subsequent structure calculations have shown that the phosphopeptide crosses the large β-sheet of the SH2 domain at the His 57-Cys 58 peptide bond. The position of this crossing over with respect to the phosphopeptide is the pTyr-ne +1 peptide bond.

The most intense protein-peptide NOE involves the pTyr Hε and Ala 46 methyl group (see Figure 5). This strong NOE intensity allows us to conclude that the Ala 46 methyl group is in van der Waals contact with the pTyr methyl group. In addition, since the chemical shift of the Ala 46 NH, Hα and Hβ are shifted downfield upon addition of pY1021 by 0.11, 0.18 and 0.15 ppm respectively, the Ala 46 methyl group contacts the plane of the aromatic ring in an edge-on fashion.

The first structures of SH2/phosphopeptide complexes (Eck et al., 1993; Kuriyan & Cowburn, 1993) demonstrated that pTyr binding occurs in part through electrostatic
Figure 5 -- Half-Filtered NOESY: NOEs to Asp -1 and pTyr. Strips from the 3D half-filtered NOESY experiment (mixing time = 150 ms) demonstrating interactions observed between the -1 and pTyr residues with the PLCC SH2 domain. All interactions shown involve residues of the SH2 on the large β-sheet, and amino acids are displayed both by number and nomenclature of Eck et al. (1993).
Figure 6 -- Protein-Peptide NOEs Involving Asp -1 and pTyr. A summary of the NOE interactions involving the -1 and pTyr residues of the pY1021 peptide and the SH2 domain. The figure was generated using the programs ChemDraw (Rubenstein, 1988) and MacDraw.
Figure 7 -- Protein-Peptide NOEs Involving the pTyr Residue: Summary by Individual pTyr Resonances. Inter-residue NOEs involving the pTyr residue.
interactions involving the pTyr phosphate group and a positively charged binding pocket on
the SH2 domain rich in Arg residues. In the PLCC SH2 domain, there are four possible
Arg residues which may be involved in pTyr binding based on comparison to other SH2
domains, namely Arg 18 (Arg αA2), Arg 37 (Arg βB5), Arg 39 (Arg βB7) and Arg 59
(Arg βD6). Results from the half-filtered NOESY showed extensive interaction of Arg 59
(from the α to the δ protons of this residue) with the pTyr aromatic protons, and these
NOEs were of moderate to strong intensity (see Figure 6). In addition, the α and δ protons
of Arg 39 (Arg βB7) have weak to moderate intensity NOEs to the pTyr HE protons.
However, no NOEs were observed involving Arg 18 (Arg αA2) and Arg 37 (Arg βB5)
even though these Arg residues are highly conserved among SH2 domains. Crystal
structures of SH2 domain/phosphopeptide complexes have demonstrated a role for Arg
αA2 and Arg βB5 in pTyr binding, as has in vitro mutagenesis of other SH2 domains
(Marengere & Pawson, 1992; Mayer et al., 1992). This lack of NOEs may be due to the
fact that a D2O sample of the PLCC SH2/pY1021 complex was used for the half-filtered
NOESY, so that protons from the guanidinium group of the Arg residues could not be
observed. It is possible that an Arg residue makes a direct hydrogen bond with the
phosphate group of the pTyr residue with no non-exchangeable protons from the Arg
residues within <5 Å of any non-exchangeable protons of pTyr. This is the situation
observed for residues Arg αA2 and Arg βD5 in the crystal structure of the src SH2 domain
(Waksman et al., 1993). However, 2D half-filtered experiments of an 15N-labeled and a
15N,13C labeled sample in H2O failed to show NOEs involving 14N- or 15N-attached
protons, including those on Arg guanidinium groups. Another possible explanation for the
lack of protein-peptide NOEs is that the side chain resonances of Arg 18, 37 and 39 are all
very broad (Pascal et al., 1995), with those of Arg 37 the most significantly broadened.
This was originally observed by Dr. Steve Pascal upon assignment of these resonances,
leading to fewer and weaker protein-protein NOE restraints for these protons. This
broadening is most likely due to μs to ms motions in the phosphate binding pocket,
however the nature of this motion has not been determined. In Chapter 6, a combination of NOE and chemical shift evidence will be presented which conclusively place the side chains of Arg αA2 and Arg βD5 in the phosphate-binding pocket in the PLCC SH2/pY1021 complex (also see Pascal et al., 1995).

No protein-peptide contacts were observed involving the BC loop; this is particularly important because in the first structures of SH2 domains, the BC loop was observed to fold over the phosphate group and make hydrogen-bonding contacts with it (Eck et al., 1993; Waksman et al., 1992; Waksman et al., 1993). Since then a number of SH2 structures have been reported both in solution (Narula et al., 1995; Zhou et al., 1995) or as crystal structures (Gilmer et al., 1994; Hatada et al., 1995) which show the BC loop not involved in pTyr binding. Further support for the lack of involvement of the BC loop in pTyr binding in the PLCC SH2 domain exists; no change in chemical shift of any protons in the BC loop and no change in order parameter of the backbone of residues in the BC loop is observed upon binding phosphopeptide (Farrow et al., 1994). It is speculated that the PLCC SH2 does not hydrogen-bond to the BC loop unlike other SH2 domains due to its numerous potential hydrogen-bonding partners among the four Arg residues discussed.

3. Ile +1

One of the features of binding specificity of the PLCC SH2 domain is the importance of the +1 residue. In one study (see Table 2), a phosphopeptide containing the sequence Ac-Asp-pTyr-Ile-NH$_2$ bound only 15-fold more poorly than the full-length pY1021 phosphopeptide, however binding was abolished by removal of the Ile +1 residue. In addition, replacement of the Ile +1 residue by Asp resulted in 8.6-fold poorer binding. The importance of a hydrophobic aliphatic amino acid at position +1 was also demonstrated in two separate studies using degenerate phosphopeptide libraries (Songyang et al., 1995;
Table 2: ID50 and $\Delta(\Delta G)$ Values for Interactions Between Peptides and the PLCC and PLCN SH2 Domains (taken from Kay et al., 1997 and Steve Shoelson, personal communication)

<table>
<thead>
<tr>
<th>Peptide Sequence*</th>
<th>PLC-γ C-SH2</th>
<th>PLC-γ N-SH2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ID$_{50}$(μM)</td>
<td>$\Delta(\Delta G)$ (kcal/mol)</td>
</tr>
<tr>
<td></td>
<td>per residue</td>
<td>cumulative</td>
</tr>
<tr>
<td>Asp-Asn-Asp-pTyr-Ile-Ile-Pro-Leu-Pro-Asp-Pro-Lys</td>
<td>1.0</td>
<td>-</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-Pro-Leu-Pro-Asp-NH$_3$</td>
<td>0.6</td>
<td>-</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-Pro-Leu-Pro-NH$_3$</td>
<td>0.5</td>
<td>-0.12</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-Pro-Leu-NH$_3$</td>
<td>1.6</td>
<td>0.67</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-Pro-NH$_3$</td>
<td>2.6</td>
<td>0.28</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-NH$_3$</td>
<td>4.6</td>
<td>0.33</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-NH$_3$</td>
<td>15</td>
<td>0.75</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-NH$_3$</td>
<td>&gt;300</td>
<td>&gt;1.65*</td>
</tr>
<tr>
<td>Ac-pTyr-Ile-NH$_3$</td>
<td>&gt;300</td>
<td>&gt;1.65*</td>
</tr>
<tr>
<td>Ac-pTyr-NH$_3$</td>
<td>&gt;300</td>
<td>-</td>
</tr>
<tr>
<td>H-pTyr-OH</td>
<td>2.260</td>
<td>-</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-Pro-Leu-Pro-Asp-NH$_3$</td>
<td>0.6</td>
<td>-</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-Pro-Leu-Pro-Arg-NH$_3$</td>
<td>0.33</td>
<td>-0.34</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-Pro-Leu-Asp-NH$_3$</td>
<td>0.8</td>
<td>0.17</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-Pro-Arg-Pro-Leu-Pro-Asp-NH$_3$</td>
<td>0.3</td>
<td>-0.40</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Ile-Asp-Leu-Pro-Asp-NH$_3$</td>
<td>2.3</td>
<td>0.77</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Ile-Asp-Leu-Pro-Asp-NH$_3$</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Ac-Asp-pTyr-Asp-Leu-Pro-Asp-NH$_3$</td>
<td>8.6</td>
<td>1.53</td>
</tr>
<tr>
<td>Ac-Ala-pTyr-Asp-Ile-Pro-Leu-Pro-Asp-NH$_3$</td>
<td>5.6</td>
<td>1.28</td>
</tr>
</tbody>
</table>

*All peptides correspond to sequences surrounding PDGF receptor Tyr1021.
Figure 8:  Half Filtered NOESY -- NOEs to Ile +1 and Pro +3. Strips from the 3D half-filtered NOESY experiment (mixing time = 150 ms) highlighting interactions observed from the +1, +2 and +3 residues to the PLCC SH2 domain. Interactions shown involve residues from several structural elements on the SH2 domain, including the large β-sheet, the EF loop and the BG loop.
Songyang et al., 1993), in which the largest enrichment factors for specific amino acids (Val and Ile in both cases) were found at the +1 position.

Many NOEs were observed involving residues of the SH2 domain and the Ile +1 residue (49 in total). Some of these NOEs are shown in Figure 8. Of these, the highest intensity NOEs were from the aromatic groups of Phe 49 (Phe βC6) and Tyr 90 (Tyr BG3), the Hβ protons of Cys 58 (Cys βD5) and the non-exchangeable protons of Lys 56 (Lys βD3) (see Figure 9). These four residues in subsequent structure calculations were found to form a pocket into which the side chain of Ile +1 fits (see Figure 10). The base of the pocket is partially formed by the aromatic ring of Phe 49. The structure of the PLCC SH2 domain is the only structure of an SH2 domain in which a residue from the βC strand forms part of the +1 binding site; this can occur because the Cys residue at the βD5 position is small enough to allow penetration, while the surface is blocked in the src and lck SH2 domains by Tyr at the βD5 position. Based on the relative intensities of the protein-peptide NOEs, the aromatic ring of Tyr 90 interacts significantly with the Hδ methyl of Ile +1, the aliphatic side chain of Lys βD3 with the Hγ2 methyl, and the Hβ protons of Cys βD5 with the Hγ1 methylene. Structure calculations subsequently confirmed this description of the +1 binding pocket, as shown in Figure 10.

The identity of the βD5 residue is very important in +1 specificity. Previously, Songyang et al (1994) have shown that SH2 domains that contain an aromatic amino acid at this position and bind hydrophilic (often negatively-charged amino acids) at the +1 position. In addition, SH2 domains with intermediate-sized amino acids at the βD5 position, such as the syp N-terminal SH2 domain (containing an Ile residue), bind hydrophobic amino acids at the +1 positions with weaker selection (Songyang et al., 1993), and structures of these SH2 domains with high affinity binding phosphopeptide show that the +1 residue is not deeply buried. In contrast, the PLCN, PLCC and vav SH2 domains have less bulky amino acids at the βD5 position, Cys in the case of the PLCN and PLCC SH2, and Thr in the case of the vav SH2 domain, and have strong selection factors
Figure 9 -- Protein-Peptide NOE Involving Ile +1 to Pro +3. A summary of NOEs involving the +1 to +3 residues of the pY1021 peptide and the SH2 domain. The following figure was generated in the same manner as Figure 6.
Figure 10: The Ile +1 Binding Site. Selected atoms from the 25 lowest energy structures showing Ile +1 and the 5 residues of the SH2 domain which make up the binding pocket for this residue. The methyl groups of Leu BG2 and the aromatic ring of Phe βC6 form the base of this pocket. Figure generated using QUANTA (Molecular Simulations Inc., 1993).
at the +1 position for specific hydrophobic amino acids (Songyang et al., 1994; Songyang et al., 1993) (see Tables 1 and 3). Mutation of the Cys βD5 residue in the PLCC SH2 domain to a Tyr causes a preference for acidic amino acids at the +1 position (as observed in Group I SH2 domains) (Songyang et al., 1995). However, the C-terminal SH2 domain of p85 contains a Cys at position βD5 but does not have strong selection at position +1; this is because the large BG loop of this SH2 domain is in contact with the +1 binding site as well, altering its characteristics.

Differences are observed between the specificity of the N- and the C-terminal SH2 domains of PLC-γ in the +1 position, namely that in the PLCC SH2 domain, the preference is Val > Ile > Leu, as compared to the preference for the PLCN SH2 domain, which is Leu > Ile > Val (see Chapter 1). To understand this difference, I have compared the sequences of the two SH2 domains at the +1 binding site residues shown for the PLCC SH2 domain (see Table 3). From this, one can conclude that the βC6 and βD5 positions are not important as they are conserved among the two SH2 domains. However, the change at the βD3 position from a Lys to a Gln would involve a loss of one methylene group, which may be enough volume to change the order of preference of hydrophobic amino acids (from Val to Leu, adding 1 methyl group). The greatest change between the PLCC and PLCN SH2 domains for +1 binding is at the BG3 position. Alignment of the BG loops of SH2 domains can be problematic because BG loops vary considerably both in sequence and sequence length, and the BG loop of the PLCN SH2 domain contains three additional amino acids relative to that of the PLCC SH2 domain. In Appendix 4 I present an alignment of BG loops from SH2 domains based on conservation of residues present in the hydrophobic core. Based on this alignment (see Appendix, Figure 2), the equivalent position in BG3 for the PLCN SH2 domain be either Arg BG3 or Cys BG4. This difference in sequence may also be responsible for the change in specificity.
Table 3: Comparison of Sequence of the PLCN SH2 Domain with the PLCC SH2 Domain at Residues Involved in +1 Binding

<table>
<thead>
<tr>
<th></th>
<th>SH2</th>
<th>βC6</th>
<th>βD3</th>
<th>βD5</th>
<th>BG3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLCN</td>
<td>Phe</td>
<td>Phe</td>
<td>Gln</td>
<td>Cys</td>
<td>Arg/Cys</td>
</tr>
<tr>
<td>PLCC</td>
<td>Phe</td>
<td></td>
<td>Lys</td>
<td>Cys</td>
<td>Tyr</td>
</tr>
</tbody>
</table>

184
4. Pro +3

The most intense protein-peptide NOEs in the Pro +3 binding site involved the Hβ and/or Hγ protons of the Pro +3 residue and Hδ methyl groups of Leu 69 (βE4) and Leu 89 (Leu BG2) (see Figure 8). In addition, a number of NOEs were observed for the +3 residue involving the aromatic ring of Tyr 90 and the Hα protons of Gly 70. Leu 69 forms the base of the EF loop, while Pro 88 and Leu 89 form the base of the BG loop. Subsequent structure calculations have shown that Pro +3 is deeply buried in a hydrophobic canyon between the EF and the BG loops. NOEs from residues on the EF loop (methyl groups of Leu 69, Hα proton of Gly 70) show stronger intensity to the Hδ resonances of Pro +3 while NOEs from residues on the BG loop (methyl groups of Leu 89, aromatic ring of Tyr 90) show stronger intensity to the Hα and Hβ2 resonances of Pro +3. These differences in NOE intensity are manifested in the structure by the orientation of the Pro +3 ring; it lies in a direction perpendicular to the length of the canyon (see Figure 11). This observation is consistent with the preference of a Pro ring at this position, in that the width of the canyon between the EF and BG loop is an optimal fit for a Pro ring. In addition, this tight fit of the Pro ring would lead to a stabilization of one of the two Pro ring conformations; in the previous chapter, NOE and coupling constant data indicated that the Pro +3 ring adopted an S- or Cγ conformation when bound to the SH2.

Thus the structure of the PLCC SH2 domain/pY1021 complex shows a highly specific interaction at the +3 position of the pY1021 phosphopeptide involving a large number of peptide-protein NOEs and in subsequent structure calculations, complete burial of the +3 residue. However, other experimental evidence is less clear about the role of the +3 residue. Order parameters for the Leu 69 methyl groups are very low, both in the free PLCC SH2 domain and in this high-affinity complex. In addition, truncation of the Pro +3 residue in the pY1021 phosphopeptide results in a less than two-fold drop in phosphopeptide binding (Kay et al., 1997). However, replacement of the Pro +3 with an
**Figure 11: The Pro +3 Binding Site.** Selected residues from the energy-minimized average structure showing the Ile +2/Pro +3 binding site. Residues Leu 69 and Gly 70 from the EF loop, and Leu 89 and Tyr 90 from the BG loop, are shown both as licorice bonds for their covalent structure, and as dot surfaces for their Van der Waals surface. Note that the direction of the Pro ring is perpendicular to the direction of the two loops, and the tight fit of the Pro ring between the two loops. The Cβ and Cγ atoms are observed to be buried into the protein interior at the base of the EF and BG loops, where they make contact with the methyl groups of Leu βE4 (Leu 69) and Leu BG2 (Leu 89).
Asp residue decreases binding affinity by more than three-fold, indicating the importance of hydrophobic interactions at this site.

5. Ile +2

Of the 151 protein/peptide NOEs, the large majority (>75%) were found to involve either the pTyr residue, Ile +1 or Pro +3. However, a few contacts were found with other residues including those from Leu +4 to Pro +7, as well as Ile +2 (plus the single NOE involving the Asp -1 residue discussed previously). However, the problems of spectral overlap which hampered assignment of protein-peptide NOEs should be initially addressed. At 1 ppm (± 0.02 ppm) are the resonances of one of the Hγ protons of Ile +1, the Hγ and Hδ methyl protons of Ile +2, and both methyl groups of Leu +4 (see Chapter 3). Most resonances which showed NOEs to +1 to +6 residues of the phosphopeptide yielded NOEs to peptide resonances at 1.0 ppm so it was important to deal with this region of spectral overlap.

NOEs involving Gly 70 Hα resonances were centered at 1.01 and 0.97 ppm, the positions where the Ile +2 Hδ and Hγ methyl resonances, respectively (see Chapter 3). Of note is that half-filtered experiments have also been performed by Dr. Yves Aubin on the PLCC SH2/pY992 complex, in which the only peptide resonances at 1.0 ppm belong to the Hγ and Hδ methyl groups of Ile +2 (personal communication), and these resonances interact specifically with Leu 69 and Gly 70. Thus NOEs involving Leu 69 and Gly 70 and peptide resonances at 1.0 ppm were assigned to the methyl groups of Ile +2. The remaining protein-peptide NOEs involving resonances at 1.0 ppm involved one of two regions, the +1 and +4 binding sites; those in the +1 binding site were assigned to the Ile +1 Hγ1 protons, while those involving side chain resonances of residues far from the +1 binding site (i.e. Pro 88, Arg 91) were assigned to the Leu +4 methyl groups. Other NOEs involving peptide protons resonating at 1.0 ppm which were bordering the +1 and +4
binding sites, including the backbone of Tyr 90 and H8 protons of Leu 89, were constrained using logical "OR" statements in X-PLOR to be close to either the +1 Hγ1 protons or the Leu +4 methyl groups.

The only strong to moderate NOEs involving the Ile +2 residue were from the Hγ and Hδ methyl protons, and the problems in assignment of NOEs for these resonances were discussed previously. Of these NOEs, the strongest were from Gly 70 (Gly EF1), but also NOEs from Leu 69 methyl groups were observed. In Figure 11 the interaction of the BC loop with both the Ile +2 and Pro +3 residues are shown, and one observes that the Ile +2 methyl groups are in weak contact with the Gly 70 residue.

The importance of the +2 residue in the PLCC SH2 domain is quite confusing, varying from experiment to experiment, and study to study. In using the phosphopeptide library degenerate at positions +1 to +3 based on the sequence GDGpYXXXSPLL (Songyang et al., 1993), a preference of greater than 2-fold was observed for branched hydrophobic amino acids, of which Ile was the most strongly preferred (see Chapter 1). However, in using a phosphopeptide library degenerate at positions +1 to +3 based on the sequence of the pY1021 peptide (Songyang et al., 1995), a preference of only 1.3-1.5-fold was observed both for branched hydrophobic and acidic amino acids, i.e. no real selection at this position. In vitro studies with the pY1021 phosphopeptide indicate that changing the Ile +2 residue to an Asp has no effect on binding (see Table 2): however changing the Ile +2 residue to an Ala in the pY992 phosphopeptide from EGFR has a large effect on binding (Maclean et al., 1995). Finally, in the pY1021 phosphopeptide, truncation from a peptide of sequence Ac-Asp-pTyr-Ile-Ile-NH2 to Ac-Asp-pTyr-Ile-NH2 decreased binding affinity by > 3-fold (see Table 3), however these truncation studies do not distinguish weaker binding due to loss of side chain interactions from loss of interactions with the backbone. The structure of the PLCC/pY1021 complex demonstrates few interactions with the Ile +2 residue, involving weak contacts with the EF loop, and is in agreement with binding studies which indicate that the Ile +2 residue plays a weak role in binding strength.
and specificity of the pY1021 phosphopeptide. However, this does not preclude that the +2 position plays a larger role in the binding of other phosphopeptides to the PLCC SH2 domain, such as the pY992 phosphopeptide. One may envision situations in which poor fits in one site (e.g. the Ile +1 site) may result in better fits in another (the Ile +2 site) with a redistribution of binding energies from one site to another.

6. Residues Leu +4 to Pro +7

NOEs involving residues +4 to +7 were fewer and weaker in intensity than those of the pTyr, +1 and +3 residues. Of these, NOEs involving peptide backbone resonances showed higher intensity. For example, three NOEs were observed between residues in the BG loop and the Pro +5 Hα; these are 50-60% of the intensity of the strongest NOE involving the Pro +3 Hα (Tyr 90 Hδ). Side chain NOEs were quite weak however. For example, NOEs involving Pro 88 Hβ and Arg 91 resonances and the Leu +4 Hδ resonances are < 36% of the intensity of the previously mentioned Pro +3 Hα/Tyr 90 Hδ NOE. One would expect greater NOE intensities because these NOEs involve the contributions of both Leu +4 methyl groups (their chemical shift is degenerate), and the calculated structures demonstrate that the Leu +4 side chain is close to the side chain of Arg 91. It is possible that binding of the +4 to +7 residues is weak, and that in these residues experience more conformational flexibility than the pTyr to +3 positions. This is consistent with the chemical shifts of the side chains of residues +4 to +7 being nearly identical to random coil values (see Chapter 3). Thus structures calculated must be interpreted conservatively in this regions of the binding site, and it may be more reasonable to calculate structures using time-averaged restraints (Torda et al., 1990) in this region.

A number of group II SH2 domains have been observed to make contacts with residues C-terminal to the +3 residue. In the syp N-terminal SH2 domain, mutation or truncation of the +5 residue has been shown to strongly decrease binding activity (Kay et
al., 1997), and this is confirmed by hydrophobic burial of the +5 residue in crystal structures of this SH2 domain with high affinity phosphopeptides (Lee et al., 1994). In the PLCN SH2 domain, truncation studies and replacement of individual residues by Asp in the pY1021 phosphopeptide have shown a large effect at the Leu +4 position and smaller effects for the Pro +5 and Asp +6 residues (see Table 2). This data may explain studies which indicate that the +4 residue is important in specificity of PLC-γ SH2 domains (Larose et al., 1993; Larose et al., 1995; Maclean et al., 1995), since these studies used constructs containing both SH2 domains of PLC-γ. In particular they are important in explaining the importance of the +4 residue of the pY1021 binding to PLC-γ SH2 domains (Larose et al., 1993; Larose et al., 1995), since the pY1021 peptide binds equally well to the PLCN and PLCC SH2 domains.

Data with the PLCC SH2 domain alone indicate much smaller effects. For example, changing Pro +5 to Asp in the pY1021 phosphopeptide only slightly decreased binding, while changing Leu +4 to Asp actually increased binding. Furthermore, truncation of the pY1021 phosphopeptide to the +3 residue decreased binding only 2.6-fold, being responsible for less than 1 kcal/mol of binding energy. Similarly, small effects were observed for the PLCC SH2 for residues C-terminal to the +3 position in the pY992 residue (Maclean et al., 1995). The BG loop of PLCN contains an additional three residues to that of the PLCC SH2 domain, which may yield a larger binding interface. In addition, the BG loop in the syp N-terminal SH2 domain is identical in size to the BG loop of the PLCN SH2, and this has been shown to yield interactions with the +5 residue.

NOEs were observed between the EF and BG loops and residues Leu +4 to Pro +7, as summarized in Figure 12. A number of NOEs were observed involving the Hδ resonances of Leu +4 with resonances of Arg 91 and Pro 88, and the Hβ resonances of Leu +4 with the Hα proton of Tyr 90. These NOEs position the Leu +4 side chain in between the side chains of Tyr 90 and Arg 91, leading to hydrophobic burial of a portion of the Leu +4 sidechain.
Figure 12 -- Protein-Peptide NOEs Involving Leu +4 to Asp +6. Summary of NOE interactions involving the +4 to +6 residues of the pY1021 peptide and the SH2 domain. Figure generated in the same manner as Figures 4 and 9.
As shown in Table 2 binding of Leu +4 in the pY1021 phosphopeptide appears to be weak. In fact, stronger binding was observed upon changing the Leu +4 residue to an Asp, as the acidic residue may be in a better position to interact with the Arg BG4 guanidinium group at the +4 position than the +6 position. Thus the hydrophobic burial of the Leu +4 side chain observed in calculated structures between the side chains of Tyr BG2 and Arg BG3 may occur a fraction of the time. In the average structure from MD simulations of the PLCC SH2/pY1021 complex (Feng et al., 1996), the Leu +4 residue was not observed to bind between these two residues in the PLCC and was in fact free in solution.

The strongest protein-peptide NOEs involving the Pro +5 residue were to various protons in the BG loop, including those from Pro BG1 (Pro 88) and Arg BG4 (Arg 91). Pro BG1 to Arg BG4 form the first turn of the BG loop: the other two turns are not in contact with the phosphopeptide. Pro +5 is observed to pack weakly against the hydrophobic surface formed by the side chains of Pro BG1 and Arg BG4, however this interaction is probably weak in nature as it appears to involve interaction of smooth solvent-exposed surfaces rather than hydrophobic burial in a cavity. Truncation of the +5 residue in the pY1021 phosphopeptide has been shown to have a nearly three-fold decrease in binding, and a slight decrease in affinity is observed upon mutation of this residue to an Asp (see Table 2).

NOEs were observed between the α proton of Arg BG4 and both the α and β protons of Asp +6. Thus the Asp +6 side chain may point towards and make an electrostatic interaction with either Arg BG4 or Lys BG5, and this salt bridge may add further binding strength to an already extensive set of interactions between the SH2 domain and this phosphopeptide. However, removal of the +6 residue in pY1021 has little effect on phosphopeptide binding, and changing the Asp +6 residue to an Arg increases binding two-fold (see Table 2). Thus in vitro results indicate that this electrostatic interaction is of no importance, as does measurement of order parameters of Arg Ne-He bond vectors.
(Pascal et al., 1995) and measurements of the protection from solvent exchange of Arg 91 in the PLCC SH2 domain/pY1021 complex (see Chapter 5 and Pascal et al., 1995).

C. The PLCC SH2 Domain/pY992 Complex

Dr. Yves Aubin has observed that the pY992 phosphopeptide resonances yield extensive protein-peptide NOEs with the aromatic ring of Tyr 84 in the αB helix, in contrast to the PLCC SH2 domain in which no such NOEs were observed. Preliminary structural data on this complex indicates that the C-terminal residue of the phosphopeptide (Phe +7) interacts with the aromatic ring of Tyr 84. In the crystal and solution structures of a number of other SH2/phosphopeptide complexes, residues in the αB helix, specifically the aromatic ring of a conserved Tyr at positions αB9 (Tyr 84 in the PLCC SH2 domain) (Eck et al., 1993; Gilmer et al., 1994; Lee et al., 1994; Nolte et al., 1996; Waksman et al., 1993), or Ile αB5 in the SHC SH2 domain (Zhou et al., 1995) are involved in peptide recognition. In addition, mutation of Tyr αB9 in the tyrosine kinase btk in humans results in X-linked agammaglobulinemia (Saffran et al., 1994), indicating its probable importance in btk SH2 recognition processes.

In addition to the presence of protein-peptide NOE's involving Tyr 84, no protein-peptide NOEs are observed for Pro 88 or Arg 91 (Pro BG1 and Arg BG4 respectively) in the PLCC SH2/pY992 complex. Thus beyond the +3 binding site, the PLCC SH2 domain has a binding surface which can accommodate a variety of chemical groups: the BG loop, with its positively-charged residues, can interact with negatively-charged amino acids, while the αB helix may be suited for binding aromatic amino acids. It is also possible that the C-terminus of the small β-sheet may act as a binding surface for residues beyond +3: this region negatively charged, so it may interact with positively-charged residues on the phosphopeptide, and the increased binding energy upon changing the Asp +6 to an Arg (see Table 2) is noteworthy. Thus we may not completely describe the binding surface of
the PLCC SH2 domain with these two high-affinity binding phosphopeptides, and further studies with other phosphopeptides will be useful in this regard.

IV. Summary

Determination of protein-peptide NOEs and subsequent structure calculations have allowed us to understand phosphopeptide recognition in the PLCC SH2/pY1021 complex. The pTyr ring was observed to fit into a large positively-charged pocket containing the highly conserved Arg βB5 guanidinium group at the base of the pocket. The pTyr ring was bound by the hydrophobic side chain of Arg βD6 and by the Cα and Cβ of His βD4. Numerous interactions were observed between the pTyr ring and the SH2 domain, however the interactions between the pTyr phosphate group and the SH2 domain could not be elucidated using NOE data and will be discussed more thoroughly in another chapter. Residues +1 to +6 of the pY1021 phosphopeptide were found to bind in a large hydrophobic canyon formed by residues in the βD strand, the EF loop and the BG loop. Within this hydrophobic binding groove was a sub-pocket formed by Phe βC8, Lys βD3, Cys βD5 and Tyr BG3 into which the side-chain of Ile +1 was buried. This smaller pocket contributes significantly to the high degree of specificity and binding energy involved in binding of the +1 residue. Also highly buried in the hydrophobic binding pocket was the Pro +3 residue, placed between the EF and BG loops in a direction perpendicular to the direction of the hydrophobic binding pocket. Several inter-residue NOEs were also observed for Leu +4, Pro +5 and Asp +6, however based on their number and weaker intensity, the interactions observed in the NMR spectra may be present a fraction of the time. The Leu +4 side chain was observed to make contact with the side chains of Tyr BG3 and Arg BG4, the Pro +5 ring contacts the surface of Pro BG1 and Arg BG4, and Arg +6 may make an electrostatic interaction with Arg BG4.
Chapter 5: pH Titration Studies of an SH2 Domain-Phosphopeptide Complex

I. Introduction

Electrostatic interactions in proteins are very important in catalysis, ligand recognition processes, and protein stability (Gilson, 1995). We are interested in the electrostatic interactions of SH2 domains with their pTyr-containing targets. SH2 domain binding to phosphopeptides is characterized by a very rapid (diffusion-limited or faster) on-rate (Panayotou et al., 1993), most likely due to electrostatic attraction of the pTyr phosphate groups with the highly positively-charged binding pockets observed in the structures of SH2 domains and SH2 domain-phosphopeptide complexes solved to date (Eck et al., 1993; Gilmer et al., 1994; Lee et al., 1994; Narula et al., 1995; Nolte et al., 1996; Pascal et al., 1994; Pascal et al., 1997, Waksman et al., 1993; Xu et al., 1995). This basic pocket is particularly evident in the PLCC SH2 domain, which contains four arginine residues in the pTyr binding region (see Chapter 4). In order to better understand the role of electrostatic interactions in SH2 domain-mediated recognition, we have performed experimental pKa determinations of ionizable groups in the PLCC SH2 domain/pY1021 complex over the pH range of 5 to 8. Residues which may change charge state over this pH range, as well as the four arginines which make up the pTyr binding pocket are shown in Figure 1.

NMR has been shown to be very useful for studying electrostatic interactions since pKa values of ionizable groups in proteins can be determined by following chemical shifts as a function of pH. The chemical shift is dependent on through-space effects of titratable groups, which have a $1/r^2$ dependence (Asakura et al., 1992). Thus the effect of ionizable groups on chemical shift can be quite long-range, so that a particular chemical shift change may be due to contributions from multiple titratable groups. For this reason, it is preferable to probe chemical shift changes of the nucleus which is directly involved in the titration or
Figure 1 -- Potential Titratable Groups in the PLCC SH2 Domain/pY1021 Complex Over the pH Range of 5-8. Ribbon diagram of the PLCC SH2 domain/pY1021 complex highlighting particular residues described in this study with the secondary structural elements labeled using the nomenclature described in Eck et al. (1993).
of a nucleus which is only one-bond separated from the titratable group. Examples include changes in $^{31}\text{P}$ chemical shifts to study phosphate pK$_{a}$s (Vogel, 1989), $^{15}\text{N}_{\delta_{1}}$ and $^{15}\text{N}_{\epsilon_{2}}$ chemical shifts of histidine imidazole groups (Bachovchin, 1986; Lodi & Knowles, 1991; van Dijk et al., 1990, 1992; Annand et al., 1992; Tsilikonas et al., 1996; Pelton et al., 1993) and $^{13}\text{C}$ side chain carboxyl chemical shifts for aspartic and glutamic acid residues (Jeng & Dyson, 1996, Qin et al., 1996).

In this study (published in Singer & Forman-Kay, 1997), I have measured $^{31}\text{P}$ chemical shifts to investigate the electrostatic contribution of the phosphate group on phosphopeptide binding to the PLCC SH2 domain. I have also determined histidine imidazole pK$_{a}$ values over this pH range using imidazole $^{15}\text{N}$ chemical shift changes. Within the PLCC SH2 domain construct are five histidine residues, most of which are poorly conserved among SH2 domains. One highly conserved histidine in the PLCC SH2 domain is His $\beta$D4. Mutation of the conserved His $\beta$D4 in the GAP N-terminal SH2 domain to a Glu or Lys destroys in vitro binding to tyrosine-phosphorylated proteins (Marengere & Pawson, 1992), and mutation of the homologous residue in the adapter protein drk to an isoleucine disrupts its in vivo role in Drosophila eye development (Olivier et al., 1993). However, in the solution structure of the PLCC SH2 domain/pY1021 complex, NOEs were observed between the pTyr aromatic protons and the His $\beta$D4 $\alpha$ and $\beta$ protons but not the imidazole ring (Pascal et al., 1994, see Chapter 4). This lack of contact between the His $\beta$D4 ring and the pTyr has also been observed in other SH2 domain-phosphopeptide complexes, solved both by heteronuclear NMR techniques and X-ray crystallography (Eck et al., 1993, Gilmer et al., 1994, Lee et al., 1994, Pascal et al., 1994, Waksman et al., 1993, Xu et al., 1995). Thus, His $\beta$D4 is both highly conserved among SH2 domains and necessary for binding function, however, no direct interaction with the pTyr or other ligand residues is present. Residues without direct contact to a ligand can modify binding energy via electrostatic effects, as is well understood. Nonetheless, it is of interest to define the specific interactions which mediate these effects.
In this study, the observation of an unusual pK$_a$ for this residue is presented as evidence that His βD4 forms hydrogen bonds with the arginine of the highly conserved FLVR sequence, Arg 37 or Arg βB5, and positions this residue for pTyr recognition.

II. Materials and Methods

Cloning and purification of the C-terminal SH2 domain of bovine PLC-γ1 (seq 3) was performed as described in Chapter 2. The phosphopeptide, representing amino acids 1018-1029 of the β platelet-derived growth factor receptor, was synthesized in the lab of Steve Shoelson.

Titrations were performed over a pH range of 5 to 8, over which ionization states of histidine, cysteine and pTyr can be probed. Resonances involved in phosphopeptide recognition were somewhat broadened at pH 5.0, indicating weaker binding, so therefore spectra were not recorded below pH 5.0 to avoid complications of contributions from the free SH2 domain to the measured pK$_a$ values. All NMR experiments were recorded on a Varian UNITY 500 MHz spectrometer equipped with a triple-resonance, pulsed field gradient probe with an actively shielded z-gradient and a gradient amplifier unit, with the exception of $^{31}$P spectra, which were collected on a Varian UNITY 500 MHz spectrometer equipped with an inverse-detected heteronuclear broadband probe. Spectra were processed with the NMRPipe software system (Delaglio et al., 1995) and analyzed with the program PIPP (Garrett et al., 1991).

pK$_a$ values for histidine residues were determined using $^1$H-$^{15}$N HMBC spectra (Pelton et al., 1993). These spectra were recorded on a 1 mM $^{15}$N-labeled PLCC SH2 domain/pY1021 sample in buffer containing 50 mM sodium acetate, 50 mM sodium arsenate, 400 mM per-deuterated DTT and 40 mM EDTA. This sample was prepared by reducing the protein-peptide complex with 5 mM $^2$H-DTT at 30°C for 1 hour at pH 8.2 and then exchanging the sample into a buffer containing 50 mM sodium acetate, 50 mM sodium
arsenate pH 5.0. Small amounts of 100 mM NaOH and HCl were used to vary the pH from 5.0 to 8.0 in 0.2-0.3 pH increments. The pKₐ values of the histidines of the free PLCC SH2 domain in the absence of phosphopeptide and phosphate were determined using a 0.2 mM sample which had been prepared by reduction of the protein and exchange into a buffer containing argon-purged 100 mM sodium acetate, pH 5.0. The pulse sequence used for recording these HMBC experiments (Pelton et al., 1993) was modified by addition of z gradients but without sensitivity enhancement in a manner similar to that used for an experiment used to measure slow exchange processes in HSQC spectra (Farrow et al., 1994). Parameters for this experiment include an 15N sweep width of 98.8 ppm centered at 205 ppm, a delay of .011 ms to obtain antiphase 15N-1H magnetization (Pelton et al., 1993), 96-128 complex t₁ points and 64-160 scans.

Gradient-enhanced 1H-15N HSQC (Zhang et al., 1994) spectra were recorded on the same samples. These experiments were useful for monitoring binding at various pH values, for confirming the values of the pKₐs obtained from measurement of histidine 15Nδ₁ and 15Ne₂ chemical shifts and for investigation of cysteine and acidic residue pKₐ values. Parameters for these experiments include 16-48 scans, 128 complex t₁ points and a sweep width of 78.9 ppm centered at 100 ppm to include arginine Nₑ and Nη correlations.

31P experiments were recorded on the same samples to determine the pKₐ of the pTyr phosphate group in the PLCC SH2 domain/pY1021 complex. Experimental parameters include a sweep width of 19.8 ppm, 8000 complex points and 14000-20000 scans per experiment. For measurement of the pKₐ of the free phosphopeptide (MW 1.45 kDa), a 5 mM sample was prepared in 50 mM sodium acetate, 50 mM sodium arsenate, 0.1 mM EDTA, pH 5.0. Small amounts of 100 mM NaOH and HCl were used to vary the pH from 4.5 to 7.8 in 0.2-0.3 pH increments. Each 31P spectrum was recorded with 128 transients, an acquisition time of 1.0 second, and a dwell time of 2.0 seconds.

Chemical shift data was recorded as a function of pH and fit to the sigmoidal curve
\[ y = \left[ \delta_{xh} + \delta_x \times 10^{(pH - pKa)} \right] / \left[ 1 + 10^{(pH - pKa)} \right], \]

where \( y \) is the observed chemical shift and \( \delta_x \) and \( \delta_{xh} \) represent the chemical shifts of the unprotonated and protonated species, respectively. Curve fits were initially performed using the program CRVFIT (Boyko & Sykes, 1993). A Monte Carlo routine provided by Dr. Neil Farrow was then used for error analysis, incorporating the uncertainties in measured chemical shifts and pH values.

Analyses of a number of SH2 domain structures were performed in order to compare hydrogen-bonding patterns involving His \( \beta D4 \). These included a refined solution structure of the PLCC SH2 domain/pY1021 complex (Pascal et al., 1997) based on the previously published structure (PDB filecode 2PLD; Pascal et al., 1994), as well as crystal structures of the N-terminal SH2 domain of syp phosphatase (Lee et al., 1994), free (PDB filecode 1AYD) and in complex with phosphopeptides (PDB filecodes 1AYA, 1AYB & 1AYC), the SH2 domain of GRB2 (PDB filecode 1GRI; Maignan et al., 1995), the lck SH2 domain-phosphopeptide complex (PDB filecode 1LCJ; Eck et al., 1993) and the src SH2 domain (Waksman et al., 1993), free (PDB filecode 1SPR) and phosphopeptide-complexed (PDB filecode 1SPS).

III. Results and Discussion

A. Histidine pKa Values in the SH2 Domain-Peptide Complex

While \(^1\text{H}-^{15}\text{N}\) HMBC experiments and their use for histidine pKa determinations have been described previously (Pelton et al., 1993), we provide a brief summary of the features of these spectra due to the importance of the data for our results. The HMBC experiments used in this study differ significantly from 1-bond \(^1\text{H}-^{15}\text{N}\) HSQC experiments. Observation of the protons directly attached to \(^{15}\text{N}\) atoms on the imidazole ring is very difficult since they may be absent (for one of the two nitrogens in the neutral
state) or in very rapid exchange with solvent. Thus, carbon-bound protons which are two- or three-bond correlated with the $^{15}$N atoms are observed. The chemical shift of the $^{15}$N resonance is strongly correlated to the protonation state of the nitrogen atom with protonated nitrogens at $\sim 170$ ppm and deprotonated nitrogens at $\sim 250$ ppm. One can also differentiate between the two neutral tautomers using these experiments because coupling constants differ depending on the influence of the lone pair of electrons on the unprotonated $^{15}$N atom (Blomberg et al., 1977). Representative HMBC spectra for the imidazole ring in the different charged states and tautomers are shown in Figure 2 (taken from Pelton et al., 1993).

The complete assignment of the histidine imidazole side chain resonances has been performed previously (Yamazaki et al., 1993a). Regions of $^1$H-$^{15}$N HMBC spectra showing the two- and three-bond $^1$HC-$^{15}$N correlations for histidine residues at pH 5.4, 6.6, and 7.8 are shown in Figure 3, with a diagram of the histidine imidazole ring shown in the lower left corner. The five histidine residues of the SH2 domain have well-dispersed resonances. At pH 7.3, the resonances from His $\beta A3$ are broadened beyond detection, and at pH 8.0, resonances of His $\alpha A7$ and His $87/\alpha B12$ are significantly broadened, probably due to exchange with solvent. The $^{15}$N$_{\delta 1}$ and $^{15}$N$_{e2}$ resonances are clearly resolved with distinct chemical shifts, particularly at higher pH values where the imidazole ring approaches the neutral state. One exception to this is His $87/\alpha B12$, for which the two imidazole $^{15}$N chemical shifts are very similar throughout the titration, and identical between pH 6.6 to 7.0. Since the original assignment of the histidine residues was performed at pH 5.7, the identities of the $^{15}$N$_{\delta 1}$ and $^{15}$N$_{e2}$ resonances of His $87/\alpha B12$ were ambiguous above pH 7.0. This ambiguity can be resolved by requiring that the titration curves for both the $^{15}$N$_{\delta 1}$ and $^{15}$N$_{e2}$ resonances be sigmoidal and that both give identical pK$_a$ values (data not shown). As described above, imidazole $^{15}$N resonances cluster about 170 ppm or 250 ppm in the protonated or unprotonated states, respectively (Pelton et al., 1993). Intermediate values of these $^{15}$N resonances at higher pH are
Figure 2 -- Representative $^1$H-$^{15}$N HMBC Spectra of the Histidine Imidazole Ring. Spectra are shown for the two neutral tautomers (parts A and B) and the charged form (part C), as well as the chemical structure of the three forms. In addition, the nomenclature of the imidazole ring is also shown, which will be used throughout. Figure taken from Pelton et al. (1993).
Figure 3 -- $^{1}H^{15}N$ HMBC Spectra. The histidine $H_{\delta2}/H_{\varepsilon1}-N_{\delta1}/N_{\varepsilon2}$ region of $^{1}H^{15}N$ HMBC spectra of the PLCC SH2 domain/pY1021 complex at pH 5.4 (a), 6.6 (b) and 7.8 (c). Correlations for all five histidines are observed at lower pH. At higher pH, resonances of His 13 are broadened beyond detection due to exchange. A diagram of the histidine imidazole ring is shown in the lower left corner. Note that neither of the two labile protons which may be present ($H_{\delta1}$ or $H_{\varepsilon2}$) are included.
indicative of fast exchange between the two neutral tautomers (Van Dijk et al., 1992). Using a value of 250 ppm for the chemical shift of the non-protonated resonance for either of the two neutral tautomers, and 170 ppm as a value for the chemical shift of the protonated $^{15}$N resonance (Pelton et al., 1993), it would be expected that an imidazole ring with an equal population of the $^{15}$N$_{E}$- and $^{15}$N$_{B}$-protonated neutral tautomers would yield an average chemical shift of 210 ppm. Based on the fit of data for the chemical shift of the His 87 N$_{E}$ resonance, the chemical shift in the neutral state is $207.6 \pm 0.3$. Thus, His 87 appears to be approximately 50% populated by each tautomer under high pH conditions which stabilize the neutral state. By extension, at lower pH there could be an equilibrium between the protonated and both neutral tautomeric states. The $^{15}$N$_{E}$ resonance of His 23 was also upfield of 250 ppm at the end of the titration, and by the same procedure it was estimated to have approximately 75% of the N$_{E}$-protonated tautomer (see Figure 4) at high pH. However, these estimates of the percentage of the two tautomers are subject to error caused by secondary effects on the histidine $^{15}$N$_{E}$ chemical shift due to ring current effects and hydrogen bonds, particularly since His 87/αB12 contacts aromatic rings in the βF strand (Phe 74) and αB helix (Tyr 83 and Tyr 84)(Pascal et al., 1997). Nonetheless, the presence of conformational exchange between the two tautomeric states is clear from the general trends of the chemical shift values.

Figure 4 shows the chemical shift changes for the $^{15}$N$_{E}$ resonances of histidines as a function of pH with the pK$_{a}$ values calculated as described above (Materials and Methods). Identical pK$_{a}$ values were obtained by plotting $^{15}$N$_{E}2$, H$_{B2}$ and H$_{E1}$ chemical shifts as a function of pH (data not shown). Free histidine has a pK$_{a}$ of 6.0 (Tanokura, 1983), but modification of the histidine zwitterion by N-acetylation and carboxymethylation (Tanokura, 1983) or formation of peptide bonds results in the imidizole ring having a pK$_{a}$ value of 6.3-6.4 (Nozaki & Tanford, 1967). The pK$_{a}$ of His 23 is identical to this value, not surprising since this residue is on the solvent-exposed face of helix αA in calculated structures. The pK$_{a}$ of His 13, 5.9, is shifted slightly lower than this value, which may be
Figure 4 -- Histidine $^{15}\text{N}_\delta$ Chemical Shifts for the PLCC SH2 Domain/pY1021 Complex as a Function of pH with Calculated pK$_a$ Values. Solid lines represents the best fit to the data. Given the chemical shifts of the $N_\delta$1 and $N_e$2 resonances of His 57 and the fact that they are unchanged over the pH range studied, the His 57 pK$_a$ value must be less than 4.0.
due to its proximity to Arg 37 and Arg 39 (Arg βB5 and Arg βB7) in the highly positively-charged pTyr-binding site. The pKₐ of His 6 at 5.6 is shifted even lower. This residue is five residues N-terminal to Trp 11, the first residue in strand βA often thought to mark the beginning of the folded domain, but has been shown in the structure of this SH2-phosphopeptide complex and at least one other SH2-phosphopeptide complex (Nolte et al., 1996) to pack against hydrophobic residues in the N-terminus of helix αB. Thus the electrostatic field due to helix αB may contribute to the lower pKₐ of His 6. In contrast, the pKₐ of His 87 at 7.0 is significantly higher than the value for histidine in random coil peptides and proteins. His 87 is at the C-terminus of the αB helix βF3 and forms part of a hydrophobic cluster involving this residue and residues βF3 (Phe 74), βB8 (Tyr 83) and βB9 (Tyr 84); in the ¹³C-NOESY experiment discussed in the following chapter, a number of NOEs were observed between these aromatic rings. The electrostatic field due to the helix dipole could contribute to the higher pKₐ value for His 87 (Lodi & Knowles, 1993).

The most dramatic perturbation from the pKₐ of free histidine was exhibited for His 57 or His βD4, whose ¹⁵N₁₈ and ¹⁵Nₑ₂ chemical shifts remain unchanged over the course of the titration at values characteristic of the neutral Nₑ-protonated tautomer. Thus, the pKₐ of this residue must be less than 4.0. There are three explanations for this anomalous behavior. The presence of the strong positive electrostatic potential in the pTyr-binding site could lower the pKₐ of His βD4. Partial or complete burial of the side chain in a hydrophobic environment could also lower the pKₐ (Del Buono et al., 1994). Finally, specific hydrogen-bonding interactions could stabilize the neutral Nₑ-protonated tautomer. His βD4 is a very highly conserved residue among SH2 domains and mutation of this residue has been shown to affect SH2 domain function in vitro (Marengere & Pawson, 1992) and in vivo (Olivier et al., 1993). In the crystal structures of phosphopeptide-complexes of the src (Gilmer et al., 1994; Waksman et al., 1993), Ick (Eck et al., 1993) and N-terminal syp (Lee et al., 1994) SH2 domains, the Nₑ₂ atom is completely buried,
and this has also been observed in refined structures of the PLCC SH2 domain/pY1021 complex (Pascal et al., 1997). In these crystal structures, the His βD4 imidazole ring is involved in a network of hydrogen-bonding interactions involving the He2 atom with the Glu αA6 carboxylate group, and the Nδ1 atom with the η protons of Arg βB5 (see Figure 5). The Oγ of Ser βC5 also participates in this network, accepting a hydrogen bond from the He proton of Arg βB5. These four residues are all highly conserved among SH2 domains (Marengere & Pawson, 1992). The His βD4 residue may thus play a critical role in positioning the Arg βB5 residue for optimal binding interactions to pTyr through this hydrogen-bonding network which requires the Nδ1 atom to be deprotonated and the Ne2 atom to be protonated. NOEs from the He1 proton to protons of Ser βC5, Arg βB5 and Glu αA6 have been observed in the spectra of the PLCC SH2 domain/pY1021 complex (Pascal et al., 1997), consistent with preservation of this hydrogen-bonding network. Of note is that the pH dependence of the binding of phosphopeptides to the lck SH2 domain has also been interpreted in terms of compromising this hydrogen-bonding network which optimizes binding site geometry (Lemmon & Ladbury, 1994).

B. Histidine pKₐ Values in the Free PLCC SH2 Domain:

If the interactions involving His βD4 play a role in orienting Arg βB5 to facilitate pTyr binding, then these hydrogen bonds might also be expected in the free PLCC SH2 domain in order to facilitate the rapid binding of ligand. In this case, a lowered pKₐ for His βD4 in the free protein should also be observed. A 0.2 mM sample of the PLCC SH2 domain in 100 mM sodium acetate (phosphate-free) was used to study the pKₐ of this histidine in the free PLCC SH2 domain in order to remove any possible effects of phosphate complexation of the SH2 domain to the His βD4 pKₐ. Previous studies (Waksman et al., 1993; Pascal et al., 1995) have shown that free phosphate ions bind to SH2 domains. The phosphate binding site is very close to the His βD4 imidazole ring.
Figure 5 -- Hydrogen-Bonding Network Including His βD4, Arg βB5 and the pTyr Residue. Positions of the side chains of Glu αA6, Arg βB5, Ser βC5 and His βD4 of the syp N-terminal SH2 domain and pTyr residue of the phosphopeptide, as well as hydrogen bonds calculated for this structure (from the phosphopeptide complex structure with PDB filecode 1AYA; Lee et al., 1994). Structures were visualized and hydrogen bonds were calculated using Quanta v4.0 (Molecular Simulations Inc., 1993).
Regions of $^1$H-$^{15}$N HMBC spectra for the free PLCC SH2 domain in 100 mM sodium acetate, pH 5.0 and the $^{15}$N PLCC SH2 domain/pY1021 complex in 50 mM sodium acetate, 50 mM sodium arsenate, pH 5.0 are shown in Figure 6. These two spectra are very similar, indicating similarity in the electrostatic behavior of the two samples and likely similar pK$_a$ values. In particular, spectra show that His$\beta$D4 is present in the N$_e$ protonated tautomer in both the free and peptide-complexed states. Note that assignment of this residue in the free SH2 domain was made by inspection based on the similarity between the spectra of free and complexed SH2 domain.

The triad of hydrogen bonds involving Glu$\alpha$A6, Ser$\beta$C5 and His$\beta$D4 has been observed in the crystal structures of SH2 domain-phosphopeptide structures that we have analyzed, including the N-terminal SH2 domain of the syp phosphatase (Lee et al., 1994) and SH2 domains of lck (Eck et al., 1993) and src (Waksman et al., 1993). With respect to crystal structures of uncomplexed SH2 domains, the situation appears more variable. In the crystal structure of the N-terminal SH2 domain of syp (Lee et al., 1994), all three hydrogen bonds are observed. However, in the SH2 domain of src (Waksman et al., 1993), the presence of these hydrogen bonds differs among the four molecules that make up the asymmetric unit. In two of these molecules, all three hydrogen bonds are observed. In one of the others, the His$\beta$D4 H$_{e2}$/Glu$\alpha$A6 O$_5$ distance is 2.9 Å, a distance which would permit only a weak hydrogen bond. In the fourth one, the conformation of Arg$\beta$B5 is altered such that different Arg$\beta$B5 H$_{\eta}$ protons make contact with the Ser$\beta$C5 O$_{\gamma}$ and His$\beta$D4 N$_{\delta 1}$ than shown in Figure 5. Finally, these hydrogen bonds are poorly conserved in the two SH2 domains in the crystallographic dimer observed in the structure of GRB2 (Maignan et al., 1995). In one, the distance between His$\beta$D4 H$_{e2}$ and Glu$\alpha$A6 O$_5$ atoms is 3.4 Å, too far for hydrogen-bonding, but the other two hydrogen bonds are present. In the other, there is a dramatic change from the hydrogen-bonding geometry shown in Figure 5, with Arg$\alpha$A2 making a salt bridge to the Glu$\alpha$A6 carboxyl group and His$\beta$D4 no longer in contact with either Arg$\alpha$A2 or Glu$\alpha$A6. The reason for the altered
Figure 6 -- Demonstration of a Low pKₐ Value for His βD4 in Free and pY1021 Complexed SH2 Domain. The histidine Hβ2/He₁·Nδ₁/Nε₂ region of H-¹³N HMBC spectra of the (a) free PLCC SH2 domain in 100 mM sodium acetate, pH 5.0 and (b) the PLCC SH2 domain/pY1021 complex at pH 5.0 in buffer containing 50 mM sodium acetate and 50 mM sodium arsenate.
conformation in this SH2 domain of GRB2 in unclear, but interactions due to packing of
the asymmetric units in the crystal bring a pair of basic residues from the N-terminal SH3
domain in one molecule to within 5 Å of the phosphate-binding pocket in the other.

One possible explanation for the variability of these hydrogen bonds in free SH2
domains is the solvent accessibility of the Glu αA6 O§ and His βD4 Ne2. Peptide binding
partially buries the Glu αA6 O§ and almost completely buries the His βD4 Ne2. Thus, it is
not surprising that the His βD4 He2 to Glu αA6 O§ hydrogen bond is the least conserved
in free SH2 domains. Since the N§1 of the His βD4 imidazole ring is always buried in the
protein interior, the hydrogen bond to Arg βB5 is much less variable. The hydrogen bonds
involving the His βD4 N§1, Ser βC5 hydroxyl and Arg βB5 guanidinium appear to be
more important than the His βD4 He2 and Glu αA6 O§ hydrogen bond in stabilizing
the position of the Arg βB5 guanidinium group and maintaining the Ne-protonated tautomer of
His βD4. The fact that in the free PLCC SH2 domain the His βD4 exists in the neutral Ne-
protonated tautomer at physiologic pH values with a low pKa value suggests that
interactions stabilizing this state are maintained in the absence of pTyr ligands, most likely
including the His βD4 N§1 to Arg βB5 Hη hydrogen bond.

C. pTyr Phosphate pKa

Free pTyr has a pKa of 5.8 for the transition from the -2 to -1 state (Vogel, 1989). In
Figure 7, the 31P chemical shifts of the pTyr phosphate groups in the free pY1021
phosphopeptide and the PLCC SH2 domain/pY1021 complex as a function of pH with the
calculated pKa values are shown. A pKa value of 6.1 is observed for the free pY1021
phosphate group. Since the chemical shift of the 31P signal of the pTyr in the PLCC SH2
domain/pY1021 complex is unchanged between pH 5.3 and 7.5, the pKa must be shifted
by more than one pH unit upon binding to the PLCC SH2 domain. Using the 31P
Figure 7 -- Low pKₐ Value for the pTyr Phosphate in the PLCC SH2 Domain/pY1021 Complex. pTyr $^{31}$P chemical shifts for the PLCC SH2 domain/pY1021 complex and the free pY1021 phosphopeptide as a function of pH with calculated pKₐ values. Solid lines represent the best fit to the data.
31P Chemical Shift (ppm)

SH2/pY1021 (pKa ~ 4.0 ± 0.2)

pY1021 free (pKa = 6.1 ± 0.1)

pH

4.4 4.9 5.4 5.9 6.4 6.9 7.4 7.9
chemical shift of the -1 state for the free peptide to estimate the chemical shift of the -1 state for the pTyr in the PLCC SH2 domain complex, we obtain a pK\(_a\) value of 4.0.

This shift in pK\(_a\) can be explained by the stabilization of the -2 state in the highly basic pTyr binding pocket of the SH2 domain. However, a similar change of pK\(_a\) has not been observed for the src SH2 domain-pYEEI complex, in which a pK\(_a\) of 5.5 was measured, as compared to a value of 5.9 for the free peptide (Xu et al., 1995). This difference may be due to the greater number of basic amino acids in the pTyr binding pocket of the PLCC SH2 domain (four arginines) than in the pTyr-binding region of the src SH2 domain (two arginines and one lysine). This result is significant since the different electrostatic potentials in the pTyr binding sites of different SH2 domains may contribute to their specificities indirectly, by a partitioning of the free energy of interaction with different amounts of the binding energy from pTyr contacts and the balance from other interactions. The differential dependence on non-pTyr residues could possibly be exploited to disrupt particular classes of SH2 domain interaction and not others. Aid in the interpretation of this low pK\(_a\) value based on computer simulations has been hindered by the difficulty in parameterization of the electrostatic portions of the force fields for pTyr in these protein complexes (Feng et al., 1996).

D. One-bond \(^1\)H-\(^{15}\)N HSQC Spectra

Backbone amide \(^1\)H-\(^{15}\)N correlations were used to confirm the histidine pK\(_a\) values determined by the HMBC spectra. It is important to note that the chemical shift changes are much smaller in these HSQC spectra than in the HMBC spectra, with differences in \(^{15}\)N chemical shifts often less than 2 ppm in HSQC spectra compared to 30-70 ppm in the HMBC spectra. This is due to the direct effect of the protonation state on the imidazole ring nitrogen as opposed to the through space effect of the electrostatic field of the ionizable group on backbone amides. Therefore, the accuracy of the determination of pK\(_a\) values
from the HMBC spectra is much greater. However, these spectra can also be used to derive information about Cys sulfhydryl groups which are ionizable in the pH range of 5 to 8, as well as Asp or Glu carboxyl groups which may be protonated at the low end of this pH range. While sulfhydryl and carboxyl group ionizations would have the greatest effects on the directly-bonded Cys $^{13}$Cβ or $^{13}$C-carboxyl resonances, respectively, chemical shift changes can also be observed for nearby backbone amide $^1$H and $^{15}$N resonances which are easily detected in $^1$H-$^{15}$N HSQC spectra of $^{15}$N-labeled samples.

Amide $^1$H resonances that change by greater than 0.15 ppm and $^{15}$N resonances that change by greater than 0.6 ppm in the HSQC spectra recorded over the pH range of 5.0 to 8.0 and that fit a simple sigmoidal curve are given in Table 1 with calculated pKa values. Many of the backbone amide $^{15}$N and $^1$H resonances from residues in the N- and C-termini undergo significant chemical shift changes as a function of pH, likely due to effects of nearby titrating histidines. Amide resonances of Ile 5 and Lys 9 are strongly affected by titration of the His 6 imidazole group (pKa=5.6), resonances of Tyr 12, His 13 and Ala 14 are affected by the His 13 imidazole group (pKa=5.9) and resonances of His 23 and Met 24 are affected by the His 23 imidazole group (pKa=6.3).

In addition, the effect of pH on resonances of His 23, Met 24 and Leu 25 were useful in assignment. Original assignment of these residues was made at pH 5.8, but these resonances coalesce and then move to new chemical shifts during the course of the titration. Their identities could be established by assuming that the changes in chemical shift fit sigmoidal curves with a pKa value close to that previously determined for the imidazole group of His 23.

There is only one cysteine residue in the PLCC SH2 domain (Cys 58 or Cys βD5) and none in the sequence of the pY1021 PDGFR phosphopeptide. No significant change in the amide $^1$H or $^{15}$N chemical shifts of Cys βD5 or nearby residues (His βD4 and Arg βD6) were observed as a function of pH in these HSQC experiments. Resonances of Ala 46, Ile 47 and Ser 48, close in space to the cysteine, also were not perturbed. Thus the
### Table 1: Phosphate $^{31}$P, Histidine $^{15}$Nδ1, and Amide $^1$H and $^{15}$N pH Titration Data

<table>
<thead>
<tr>
<th>Residue</th>
<th>pKa</th>
<th>HA Chemical Shift</th>
<th>A- Chemical Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phosphate $^{31}$P:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>free pTyr</td>
<td>6.08 (0.02)</td>
<td>-1.072 (0.02)</td>
<td>2.95 (0.02)</td>
</tr>
<tr>
<td>bound pTyr</td>
<td>4.01 (0.18)</td>
<td>-1.0 (1.0)**</td>
<td>2.41 (0.01)</td>
</tr>
<tr>
<td><strong>Histidine $^{15}$Nδ1:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His 6</td>
<td>5.60 (0.02)</td>
<td>178.8 (1.4)</td>
<td>244.9 (0.1)</td>
</tr>
<tr>
<td>His 13</td>
<td>5.93 (0.02)</td>
<td>187.0 (0.6)</td>
<td>238.7 (0.3)</td>
</tr>
<tr>
<td>His 23</td>
<td>6.29 (0.02)</td>
<td>177.9 (0.3)</td>
<td>228.1 (0.2)</td>
</tr>
<tr>
<td>His 87</td>
<td>7.01 (0.02)</td>
<td>174.3 (0.1)</td>
<td>207.3 (0.4)</td>
</tr>
<tr>
<td><strong>Amide $^1$H:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile 5</td>
<td>5.62 (0.04)</td>
<td>7.88 (0.01)</td>
<td>8.30 (0.01)</td>
</tr>
<tr>
<td>Ser 8</td>
<td>5.79 (0.03)</td>
<td>7.83 (0.01)</td>
<td>7.36 (0.01)</td>
</tr>
<tr>
<td>Lys 9</td>
<td>5.64 (0.02)</td>
<td>8.11 (0.02)</td>
<td>7.30 (0.01)</td>
</tr>
<tr>
<td>Glu 10</td>
<td>5.81 (0.04)</td>
<td>8.43 (0.03)</td>
<td>9.22 (0.01)</td>
</tr>
<tr>
<td>Tyr 12</td>
<td>5.97 (0.03)</td>
<td>7.87 (0.01)</td>
<td>7.51 (0.01)</td>
</tr>
<tr>
<td>Leu 16</td>
<td>5.06 (0.08)</td>
<td>7.78 (0.03)</td>
<td>8.09 (0.01)</td>
</tr>
<tr>
<td>Met 24</td>
<td>6.17 (0.05)</td>
<td>8.04 (0.01)</td>
<td>7.87 (0.01)</td>
</tr>
<tr>
<td>Val 36</td>
<td>5.78 (0.04)</td>
<td>9.38 (0.01)</td>
<td>9.69 (0.01)</td>
</tr>
<tr>
<td>Lys 38</td>
<td>5.71 (0.04)</td>
<td>8.24 (0.01)</td>
<td>7.99 (0.01)</td>
</tr>
<tr>
<td>Arg 39</td>
<td>6.66 (0.05)</td>
<td>8.07 (0.01)</td>
<td>8.25 (0.01)</td>
</tr>
<tr>
<td>Ile 81</td>
<td>6.07 (0.06)</td>
<td>7.90 (0.01)</td>
<td>7.74 (0.01)</td>
</tr>
<tr>
<td>Tyr 84</td>
<td>6.71 (0.04)</td>
<td>7.71 (0.01)</td>
<td>7.51 (0.01)</td>
</tr>
<tr>
<td>Ile 99</td>
<td>6.20 (0.05)</td>
<td>8.26 (0.01)</td>
<td>8.43 (0.01)</td>
</tr>
<tr>
<td>Asn 100</td>
<td>6.16 (0.12)</td>
<td>8.60 (0.01)</td>
<td>8.40 (0.02)</td>
</tr>
<tr>
<td>Glu 101</td>
<td>6.20 (0.03)</td>
<td>8.45 (0.01)</td>
<td>8.94 (0.01)</td>
</tr>
<tr>
<td>Ser 104</td>
<td>6.08 (0.07)</td>
<td>8.19 (0.01)</td>
<td>8.01 (0.01)</td>
</tr>
<tr>
<td>Ser 105</td>
<td>6.13 (0.06)</td>
<td>7.94 (0.01)</td>
<td>7.79 (0.01)</td>
</tr>
<tr>
<td><strong>Amide $^{15}$N:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile 5</td>
<td>5.71 (0.18)</td>
<td>119.9 (0.2)</td>
<td>121.5 (0.1)</td>
</tr>
<tr>
<td>Ser 8</td>
<td>5.69 (0.06)</td>
<td>116.2 (0.2)</td>
<td>111.4 (0.1)</td>
</tr>
<tr>
<td>Glu 10</td>
<td>5.81 (0.25)</td>
<td>120.0 (0.3)</td>
<td>121.4 (0.1)</td>
</tr>
<tr>
<td>Tyr 12</td>
<td>6.07 (0.15)</td>
<td>124.0 (0.1)</td>
<td>122.7 (0.1)</td>
</tr>
<tr>
<td>His 13</td>
<td>6.11 (0.18)</td>
<td>127.8 (0.1)</td>
<td>128.9 (0.1)</td>
</tr>
<tr>
<td>Ala 14</td>
<td>6.05 (0.22)</td>
<td>124.7 (0.1)</td>
<td>125.6 (0.1)</td>
</tr>
<tr>
<td>Ser 15</td>
<td>6.15 (0.16)</td>
<td>111.2 (0.1)</td>
<td>112.4 (0.1)</td>
</tr>
<tr>
<td>Leu 16</td>
<td>5.58 (0.14)</td>
<td>125.5 (0.2)</td>
<td>127.2 (0.1)</td>
</tr>
<tr>
<td>His 23</td>
<td>6.28 (0.11)</td>
<td>116.6 (0.1)</td>
<td>118.3 (0.1)</td>
</tr>
<tr>
<td>Met 24</td>
<td>6.15 (0.27)</td>
<td>118.2 (0.1)</td>
<td>117.5 (0.1)</td>
</tr>
<tr>
<td>Leu 25</td>
<td>5.99 (0.21)</td>
<td>118.2 (0.1)</td>
<td>117.2 (0.1)</td>
</tr>
<tr>
<td>Glu 63</td>
<td>5.61 (0.35)</td>
<td>129.3 (0.2)</td>
<td>129.9 (0.1)</td>
</tr>
</tbody>
</table>
Ser 76  4.99 (0.28)  107.2 (0.8)  105.4 (0.1)
Val 78  5.87 (0.16)  117.3 (0.1)  116.0 (0.1)
Glu 85  5.32 (0.19)  116.3 (0.4)  118.0 (0.1)
His 87  7.06 (0.10)  120.1 (0.1)  121.9 (0.1)
Ile 99  6.21 (0.14)  123.8 (0.1)  125.1 (0.1)
Asn 100 5.91 (0.13)  126.5 (0.2)  123.3 (0.3)

\textsuperscript{a}Standard deviations of the data fit using an in-house Monte Carlo program (see Materials and Methods) are reported in brackets.

\textsuperscript{b}A value of -1.082±1.0 ppm was used as an estimate for the chemical shift of the -1 charge state of the pTyr phosphate in the PLCC SH2 domain/pY1021 complex based on the chemical shift of the -1 state in the free phosphopeptide, since no value for this chemical shift could be accurately determined from the data.
pKₐ of Cys βD5 must be 9 or greater, higher than the pKₐ of free cysteine (pKₐ=8.3). This increased pKₐ value is likely due to burial of the sulphydryl group in the hydrophobic binding interface of the peptide complex. Cys βD5 has been shown by structural studies (Pascal et al., 1994) to be involved in phosphopeptide recognition at the +1 position relative to the pTyr leading to burial of the sulphydryl group. Cys βD5 is conserved between the N- and C-terminal SH2 domains of PLCγ, both of which have a strong preference for hydrophobic amino acids at the +1 position (Songyang et al., 1993). In general the identity of the residue at the βD5 position is very important in SH2 domain recognition of phosphopeptides at the +1 position (Songyang, 1994).

The carboxyl groups of free Asp and Glu residues have pKₐ values of 4.0 and 4.4 respectively (Nozaki & Tanford, 1967). In the PLCC SH2 domain/pY1021 complex, significant changes in ¹H or ¹⁵N chemical shifts over the pH range of 5 to 8 were observed for backbone amides of Glu 10, Glu 63 Glu 85 and Glu 101. Changes in chemical shifts of Glu 10 and 101 may be due to titration of His 6 and/or His 13 as titration curves for these residues yield pKₐ values close to 6. However, chemical shift changes of backbone ¹H or ¹⁵N resonances of a number of residues in the small β-sheet (strands βD' to βF) and the αB helix could be fit with pKₐ values significantly lower than the pKₐ values of His 6 and His 13. The Glu 85 ¹⁵N resonance is affected by an ionizable group which was fit to a pKₐ of 5.3, most likely the Glu 85 carboxyl group itself with a pKₐ value approximately 1 pH unit higher than that of free glutamic acid. Glu 85 lies near the C-terminus of the αB helix and the negative potential due to the helical dipole may raise the pKₐ of the carboxyl group of Glu 85, as suggested previously for the imidazole group of His 87.

Large chemical shift changes for the backbone ¹⁵N resonances of Glu 63/βD'3, Ser 76/αB1 and Val 78/αB3 are observed over the pH range of 5 to 6. Based on the structure of the PLCC SH2 domain/pY1021 complex, these residues are most likely responding to the sum of the effects of ionization of the carboxyl groups of Glu 63, Glu 73, Glu 75 and
Asp 79 and also to the nearby His 6 imidazole group. This region of the PLCC SH2 domain contains a large concentration of negative charge and has been speculated to be a region for Ca\textsuperscript{2+} binding (Mahadevan et al., 1994).

E. Amide Exchange with Solvent

The rate of amide exchange with water increases as a function of pH from a minimum around pH 3 (Wüthrich, 1986). The lowest rates of exchange are for protons buried in the protein interior or protected by stable hydrogen bonds. The resonances broaden as the rate increases from the slow exchange limit at which sharp peaks can be observed in $^1$H-$^{15}$N HSQC spectra, to the intermediate exchange limit. Finally, the resonances merge with the water peak in the fast exchange limit. The effect of slow or intermediate exchange with water on the intensity of HSQC peaks can be minimized by using experiments which do not perturb the water resonance (Grzesiek & Bax, 1993; Stonehouse et al., 1994; Zhang et al., 1994) even at pH values slightly above neutral. The arginine side chain guanidinium resonances were examined as a function of pH in order to probe the extent to which the different guanidinium side chains are protected from exchange by hydrogen-bonding or burial in the PLCC SH2 domain/pY1021 complex. No change in chemical shifts for Arg 18, 37, 39, and 59 (Arg $\alpha$A2, $\beta$B5, $\beta$B7, $\beta$D6) guanidinium resonances were observed, indicating the stability of the arginine-pTyr interactions over this pH range (5-8). The arginine resonances do, however, experience differential exchange broadening behavior as a function of pH (Figure 8). The $H_e$ resonances of three arginines not involved in phosphopeptide binding (Arg 27, Arg 30 and Arg 96) disappear most rapidly, at pH values below 7, due to exchange with solvent, as does that of Arg 91 whose aliphatic group only is involved in peptide binding at the $+4$ to $+6$ positions of the phosphopeptide (Pascal et al., 1994). This exchange behavior indicates that any hydrogen-bonding interactions involving the $H_e$ of these four residues must be weak and/or transient.
Figure 8: Arginine $H_\varepsilon$ Exchange Broadening Behavior as a Function of pH.
The pH at which the intensity is 20% of maximum is compared to (a) the measured rate of $H_\varepsilon$ exchange with solvent and (b) the $^1H-^{15}N_\varepsilon$ order parameter (both sets of data from Pascal et al., 1995). Note that for two residues, Arg 37 and Arg 39, the $H_\varepsilon$ peaks are not significantly broadened by pH 8.0 and these residues are arbitrarily plotted at a pH value greater than 8.0.
In contrast, the $H_E$ resonances of Arg $\beta B5$ and Arg $\beta B7$ show no exchange broadening even at pH 8.0, indicating that they are buried by the phosphopeptide and/or make strong hydrogen-bonding contact with the phosphopeptide. It has been shown that Arg 39 is involved in hydrogen-bonding interactions both with free phosphate and with the pY1021 phosphate group due to a strong downfield shift of the $H_E$ resonance in phosphate buffer relative to imidazole buffer (Pascal et al., 1995). Such a downfield shift was not observed for the Arg $\beta B5$ $H_E$ resonance; in fact, a slightly upfield shift was observed. However, recent structure calculations have shown that the Arg $\beta B5$ $H_H$ protons are involved in hydrogen-bonding with the phosphate group, and in making these contacts, the Arg $\beta B5$ $H_E$ and $N_E$ are completely buried from solvent.

The remaining three arginine $H_E$ protons have exchange behavior between these extreme cases. The $H_E$ resonance of Arg $\beta D6$ broadens and disappears by pH 6.7. The aliphatic side chain positions of Arg $\beta D6$ contact the phenyl ring of the pTyr, but the guanidinium group is not likely to be involved based on results of $^{15}N$ relaxation experiments probing ps-ns motion of the $^{1}H$-$^{15}N_E$ bond vectors (Pascal et al., 1995). The $H_E$ resonance of Arg 50 broadens beyond detection at pH 7.1. Based on the $^{1}H$-$^{15}N_E$ order parameter and the rate of exchange, Arg 50 may be involved in hydrogen-bonding interactions with carbonyl groups in the AB loop. Arg 27 and Arg 30 are found within this AB loop, but as discussed previously are surface-exposed and are probably not involved in any intramolecular hydrogen-bonding interactions. Finally, the $H_E$ resonance of Arg $\alpha A2$ broadens to 20% of its maximum intensity by pH 8.0. Data from previous exchange and relaxation experiments did not indicate a clear difference between the $H_E$ resonance of Arg $\alpha A2$ and Arg $\beta B7$ in terms of protection from exchange (Pascal et al., 1995). Relaxation measurements, which monitor motion of the guanidinium group and cannot distinguish between stabilizing interactions formed by hydrogen bonds involving the $H_E$ and $H_H$ protons, in fact showed a slightly higher order parameter for the $^{1}H$-$^{15}N_E$ of Arg $\alpha A2$ than Arg $\beta B7$. However, chemical shift arguments are supportive of Arg $\beta B7$ $H_E$ being
involved in stronger hydrogen-bonding interactions than the Arg αA2 Hε, in that a much larger chemical shift change for the Arg βB7 Hε resonance occurs upon phosphate binding. While the timescales of motion that influence 15N relaxation and solvent exchange can be significantly different, the correlation between the 1H-15N order parameter and the exchange behavior (Figure 8b) is intriguing and some insights may be gained by their comparison.

Data for the arginine Hη protons are not shown as most of these resonances are highly overlapped, and those which are distinct are very broad due to conformational exchange involving hindered rotation about the Ne-Cζ and Cζ-Nη bonds. The majority of the Hη resonances disappear due to exchange broadening by pH 7.0, with the exception of the two Arg βB7 Hη resonances which are protected from exchange until pH 7.2 and the Arg βB5 Hη resonances. While only two of the four potential Hη resonances of Arg βB7 are observed, likely due to rapid rotation about the Cζ-Nη bonds, Arg βB5 is unique in the PLCC SH2 domain/pY1021 complex in that all four Hη resonances are observed (Yamazaki et al., 1995). These four resonances can be stereoassigned by considering intra-residue distances between the Hε and Hη protons in arginine guanidinium groups (Pascal et al., 1997). The Arg βB5 Hη11 and Hη12 resonances are very weak and broad, but there is no pH dependence of the intensities. The resonances of Hη21 and Hη22 are sharper and have strong intensities throughout the titration, with no significant broadening by pH 8.0. These results further confirm the presence of stable hydrogen-bonding interactions involving the Arg βB5 guanidinium group Hε and Hη protons.

In addition to arginine NH and NH2 guanidinium resonances, a number of amide backbone resonances broaden and disappear over the course of the titration. These resonances are from residues in the disordered N-terminal and C-terminal portions of the SH2 domain (Gly 4, Ile 5, Ser 8, Glu 10, Glu 101 and Glu 102) or from loops (Asn 40 and Asn 43 in the BC loop, Gly 64 and Gln 65 in the D'E loop, Gly 70 and Asn 71 in the EF loop, Leu 77 near the FB loop and Arg 91 in the BG loop). Both the N- and C-terminal
residues as well as the BC, D'E and EF loops are poorly defined due to a lack of NOE restraints in structure calculations of the PLCC SH2 domain/pY1021 complex (Pascal et al., 1994). Some of these residues are in regions where backbone $^{15}$N relaxation measurements show low $^1$H-$^{15}$N order parameters (Farrow et al., 1994), i.e. the BC loop and N- and C-terminal residues, confirming greater mobility. Leu 77 which is completely exchange-broadened prior to pH 8.0 has a high order parameter of 0.85 but a very large $R_{ex}$ component indicating the presence of millisecond time scale motions. Thus, the exchange data is correlated with absence of hydrogen-bonding, surface exposure and conformational mobility either on the ps-ns timescale, giving rise to low order parameters in NMR relaxation measurements, or on the millisecond timescale.

IV. Summary:

In summary, the pH titration behavior of residues of the PLCC SH2 domain/pY1021 complex has been investigated. In this complex, the pTyr of the pY1021 peptide and His 57/$\beta$D4 residue of the PLCC SH2 domain have anomalously low pK$_A$ values (<5.0). Stabilization of the -2 state of the phosphate of the pTyr in the highly positively charged binding pocket is consistent with the theory of a strong electrostatic attraction leading to extremely rapid on-rates for SH2 domain interactions. Burial of His $\beta$D4 and stabilization of the N$_e$-protonated tautomer through a hydrogen-bonding network similar to that observed in the crystal structures of other SH2 domain-phosphopeptide complexes leads to a lowering of the pK$_A$ value in both the peptide-complexed and free SH2 domains. This result provides additional insights into the role of the highly conserved His $\beta$D4 residue in positioning a critical arginine to allow pTyr-SH2 domain binding. His 87 as well as Glu 85 yield pK$_A$ values elevated from the free amino acid pK$_A$ by approximately 1 pH unit due to interaction with the helix dipole at the C-terminus of helix $\alpha$B. These titration experiments also provide information concerning the exchange with solvent of NH and
NH\textsubscript{2} protons of both backbone and arginine side chains as a function of pH. Most arginine resonances broaden beyond detection due to exchange by neutral pH, however the H\textsubscript{E} and the H\textsubscript{\Pi} protons of Arg \beta B5 and the H\textsubscript{E} proton of Arg \beta B7 are not significantly broadened during the course of the titration. The H\textsubscript{E} proton of Arg \alpha A2 did broaden but not significantly until pH 8.0, consistent with weaker hydrogen-bonding interactions involving this residue than those of Arg \beta B5 and Arg \beta B7.
Chapter 6: Investigations of Arginines and pTyr in the PLCC SH2 Domain/pY1021 Complex

I. Introduction

SH2 domain interactions can be subdivided into two separate recognition regions, one involving the pTyr and the other involving specific sequences C-terminal to the pTyr residue. pTyr recognition is crucial in an SH2 interaction, as it has been shown by in vitro binding assays and mutational analysis that binding to non-tyrosine phosphorylated peptides or proteins is weak or non-existent (Anderson et al., 1990; Kazlauskas et al., 1992; Moran et al., 1990; Rotin et al., 1992; Valius et al., 1993). While NOEs obtained from half-filtered experiments (see Chapter 4) and subsequent structure calculations (Pascal et al., 1994) yielded much information concerning recognition by the PLCC SH2 domain of residues C-terminal to the pTyr in the pY1021 phosphopeptide, data on pTyr recognition was more limited. A number of NOEs were observed between the pTyr and SH2 domain resonances, including the backbone and side-chains of Arg 39, Ala 46, His 57 (excluding the imidazole ring), Arg 59 and the Cys 58 backbone. However, these resonances were from non-exchangeable protons, and thus the data yielded little concerning the charge-charge interactions involving the pTyr and the four possible arginine guanidinium groups which were found to occupy the binding pocket (Pascal et al., 1995).

NMR structure determination depends in large part on $^1$H-$^1$H NOEs, and such NOEs can easily be found for hydrophobic packing in cores of proteins, particularly involving interactions of methyl groups. However, charge-charge or hydrogen-bonding interactions are much more difficult to discern by solution NMR methods as polar groups may not be protonated (e.g. carboxyl groups) or may have protons in fast exchange with solvent such that their resonances cannot be observed (e.g. Ser OH, lysine $\text{NH}_3^+$). In addition, these
interactions are often on the protein surface, where side chain positions are more poorly defined in NOE-based structure determinations and where solvent interactions may compete. As an example of the problems in detection of charge-charge interactions in NOE-based structure determination, the nearest non-exchangeable protons of the peptide to the Arg αA2 and Arg βB5 guanidinium groups in the crystal structure of the src SH2 domain/YEEI complex (Waksman et al., 1993) were within a distance of > 6 Å. For the -2 charge state, there are no protons on the phosphate group, so in an NMR structure determination, no protein-peptide NOEs would be observed linking these two arginine side chains with the phosphate group. In addition, for technical reasons, we were unable to observe NOEs from the arginine Hɛ and Hη protons to the phosphopeptide (see Chapter 4).

Another possibility for investigation of the charge-charge interaction involving the pTyr may be heteronuclear 31P-1H NOEs. However, the magnitude of this interaction depends on γp2γH2, and since the γ of 31P is roughly 40% as large as that of the 1H γ, one would expect the 31P-1H NOE to be 16% as large as a 1H-1H NOE at best. Secondly, 31P-1H dipolar relaxation is strongly dependent on the value of J(ωH-ωP), which is not as large as J(0), which predominates in dipolar relaxation in proteins. Finally, relaxation of 31P nuclei also includes contributions by chemical shift anisotropy, which increases as higher fields are used (Vogel, 1989). For that reason, heteronuclear 31P-1H NOEs are not commonly measured.

Thus, we investigated the Arg-pTyr interactions by a number of different methods. It has been observed that a free phosphate ion can bind to an SH2 domain in the same position as a pTyr phosphate group (Waksman et al., 1993). Comparison of samples of the free PLCC SH2 domain in imidazole and phosphate buffers to those of the pY1021 complex allowed determination of the effects of the phosphate group on the chemical shifts of the protein. The analysis of these experiments was performed by myself, Toshio Yamazaki and Steve Pascal. In addition, I reanalysed spectra to find NOEs involving
resonances of the Arg guanidinium groups in the PLCC SH2 domain/pY1021 complex in order to improve their precision in structure calculations and help discern their role in pTyr binding.

II. Materials and Methods

Assignment of arginine guanidinium resonances was based on previous assignment of these resonances by Toshio Yamazaki using a doubly-labeled SH2 domain in complex with the pY1021 phosphopeptide and the experiments described in Yamazaki et al. (1995). I prepared the imidazole sample and phosphate samples by taking 250 μl of 15N-labeled protein for each sample from fractions from a MonoS column, reducing the protein and exchanging the sample using Amicon Centricon 3 ultracentrifugation units into a buffer containing argon-purged 100 mM imidazole, pH 6.0, or 100 mM sodium phosphate, pH 6.0, to a final volume of 500 μl. Following collection of HSQC spectra, the pY1021 phosphopeptide was titrated into the imidazole-buffered sample until a 1:1 protein/peptide ratio was obtained. HSQC experiments were run using the pulse sequence described in Zhang et al. (1994) or using a modified HSQC to enhance sensitivity for the arginine NH2 resonances (Pascal et al., 1995). These HSQC experiments typically employ a 79 ppm sweep width centered around 100 ppm in F1 in order to observe both the arginine guanidinium and backbone amide resonances. With these parameters, the Trp indole 15NH is folded, but all other resonances fall within the sweep width.

The sample used for the 13C-NOESY and CN-NOESY was a 1.5 mM sample of 13C,15N-labeled PLCC SH2 domain to which excess pY1021 was added, dissolved in 100 mM sodium phosphate buffer, pH 6.3. Parameters for the 3D 13C-NOESY included 16 transients, 128 and 32 points in t1 and t2, respectively, and sweep widths of 5849.7, 1650 and 8000 Hz in t1, t2 and t3, respectively, with the 13C offset set to 43 ppm. Parameters for the 3D 15N-NOESY included 16 transients, 128 and 32 experiments in t1
and t2, respectively, and sweep widths of 5000, 3000 and 8000 Hz in t1, t2 and t3, respectively, with the 15N offset set to 119 ppm. For the 3D CN-NOESY, 16 transients, 128 points in t1, 32 points in t2, and 832 complex points in t3 were collected with sweep widths of 5000, 3000 and 8000 Hz in t1, t2 and t3, respectively, with the 13C offset set to 43 ppm and the 15N offset set to 119 ppm. All three experiments employed a 150 ms mixing time. These experiments were processed using the NMRPipe (Delaglio et al., 1995) program, with the data apodized using 60-90° phase-shifted sine-bell squared functions. Processed data was then analyzed using the programs PIPP and STAPP (Garrett et al., 1994) on Sun Sparc Workstations.

NOESY experiments which select for protons close in space to arginine guanidinium protons were developed by Toshio Yamazaki, and were run using the same 13C,15N doubly-labeled sample of the PLCC SH2 domain/pY1021 complex used for the 3D 13C-NOESY and CN-NOESY. These experiments encompass a 2D version with proton detection in both dimensions and a 3D version involving proton detection in t1 and 15N detection in t2. Parameters for the 2D experiment include the collection of 256 t1 points, 1024 complex t2 points and 128 transients, sweep widths of 6000 and 8000 Hz in t1 and t2, respectively, and a mixing time of 100 ms. Selection of arginine guanidinium 15N magnetization was achieved by SEDUCE shaped pulses applied following t1 and the 100 ms mixing time. In the 3D experiment, 8 transients, 128 and 32 points in t1 and t2, respectively, and 1024 complex points in t3 were collected, with a sweep width of 4000 Hz in 15N centered at 100 ppm. All experiments were run on Varian Unity 500 or 600 MHz Spectrometers equipped with a triple-resonance, pulsed field gradient probe with an actively shielded z-gradient and a gradient amplifier unit.

In the interpretation of our data, we used structures calculated as described in Chapter 4, the publicly-available structures of SH2 domains and SH2 domain/phosphopeptide complexes (see Chapter 4), and the average coordinates of a 130-ps simulation of the
PLCC SH2 domain/pY1021 complex described in Feng et al. (1996). Molecular visualization was done using the program QUANTA (Molecular Simulations Inc., 1993).

III. Results and Discussion

A. Arginine Resonances of the PLCC SH2 Domain: Sample Conditions

In order to obtain information about interactions involving the pTyr phosphate group, we employed a chemical shift-based approach in which the chemical shifts of the SH2 domain in phosphate buffer were compared to the SH2 domain in imidazole buffer and to a phosphate-free PLCC SH2 domain/pY1021 complex. We have made two assumptions: (1) that the phosphate ion binds to the SH2 domain and causes very similar chemical shift changes as the pTyr phosphate group and (2) that the differences in chemical shift between the PLCC SH2 domain in imidazole and phosphate buffer are due only to specific phosphate ion binding in the pTyr-binding pocket. The first assumption is based on the X-ray structure of the free src SH2 domain (Waksman et al., 1993), in which a phosphate ion was observed to bind in one of the four molecules in the asymmetric unit; the position of the phosphate group was identical to the position of the phosphate group of pTyr in the src SH2 domain/phosphopeptide complexes (Waksman et al., 1992; Waksman et al., 1993).

In Figure 1 are shown spectra of the 15N-labeled PLCC SH2 domain in (a) imidazole buffer, (b) phosphate buffer and (c) imidazole buffer to which the pY1021 phosphopeptide has been added in a 1:1 ratio to the PLCC SH2 domain. The spectrum in figure 1c is identical to that observed in a previously prepared doubly-labeled sample of SH2 domain in phosphate buffer to which phosphopeptide had been added to excess. This older sample was previously used for assignment of arginine H€ and Hη protons (Yamazaki et al., 1995).
Figure 1 -- $^{15}$N-$^1$H HSQC Spectra. (a) imidazole buffer, (b) phosphate buffer, and (c) imidazole buffer with pY1021 phosphopeptide to yield a 1:1 SH2 domain/phosphopeptide complex. All samples are at pH 6.0. The figure is taken from Pascal et al. (1995).
a) Imidazole

b) $P_i$

c) pY1021
In the imidazole spectra (Figure 1a), a number of H_ε and H_η protons with very similar chemical shifts can be observed. Each arginine has a single resonance line for all four H_η protons. However, upon addition of phosphate (Figure 1b) or phosphopeptide (Figure 1c), a number of chemical shift changes occur with some arginines having more than one H_η resonance. The most dramatic differences were observed for Arg 37. As found previously (Yamazaki et al., 1995), four H_η proton resonances are present for Arg 37 in the SH2 domain/pY1021 complex. One of these four resonances (at 7.45 ppm) was shifted downfield upon binding to the phosphopeptide, two were very strongly shifted downfield (to 9.47 and 9.67 ppm) and a fourth was shifted upfield to 5.75 ppm. The 9.47 ppm and 7.45 ppm resonances were much sharper and stronger than the other two H_η resonances. The 5.75 and 9.67 ppm resonances can only be observed in free SH2 domain in phosphate buffer at 10° C (Toshio Yamazaki, personal communication). In addition to the changes in resonances of Arg 37, the H_η resonance of Arg 39 splits into two upon binding the pY1021 peptide. Arg 18 at 30° C has one observable upfield-shifted resonance (δ_Η 6.08 ppm, δ_Ι5N 71 ppm), but a second resonance was observed at 10° C at δ_Ι5N ~73.2 ppm by Toshio Yamazaki (Yamazaki et al., 1995).

Considerable motion exists about the C_ζ-N_η bond and N_ε-C_ζ bonds in arginine despite their partial double-bond character, such that in NMR spectra of free arginine at 30° C, the four H_η signals coalesce into a single resonance broad line at approximately 6.6 ppm with exchange rates in the microsecond to low millisecond time scale (Kanamori & Roberts, 1983, Yamazaki et al., 1995). A barrier to rotation of 14 ± 0.2 kcal/mol was measured for rotation about the N_ε-C_ζ bond, with rotation about the C_ζ-N_η having an even lower barrier of rotation (Yamazaki et al., 1995). In the free SH2 domain in imidazole buffer, none of the arginine guanidinium groups are involved in a stable hydrogen-bonding arrangement as is evident from the one broad H_η resonance observed for each Arg and the near random-coil chemical shifts for the H_η resonances which are between 6.5 and 6.8 ppm (see Figure 1). Similarly, single H_η resonances are observed in
phosphate buffer and in the pY1021 complex for Arg 27, 30, 50, 59 and 96, and their $^{15}$N$_{\gamma}$, H$_{\gamma}$, $^{15}$N$_{\varepsilon}$ and H$_{\varepsilon}$ resonances have similar chemical shifts among the three samples. This provides evidence that these guanidinium groups are not involved in any strong hydrogen-bonding interactions in any of the sample conditions.

When an H$_{\gamma}$ is involved in a hydrogen bond, the energy from hydrogen-bonding should increase the barrier to rotation about the C$_{\varepsilon}$-N$_{\gamma}$ and N$_{\varepsilon}$-C$_{\varepsilon}$ bonds, and several H$_{\gamma}$ resonances may be observed. Arginine guanidinium groups have five possible hydrogen-bond donors (the H$_{\varepsilon}$ and each of the four H$_{\gamma}$ protons) which can make monodentate or bidentate hydrogen bonds with hydrogen-bond acceptors. Bidentate interactions of Arg guanidinium groups can occur in one of two possible arrangements (see Figure 2). Arg 37 (Arg $\beta$B5) probably is involved in the bidentate interaction shown in Figure 2a with one or two of the phosphate pTyr oxygen atoms, as observed in the crystal structure of every SH2 domain/phosphopeptide complex. Spectroscopically, this interaction results in observation of four H$_{\gamma}$ protons, of which two protons (H$_{\gamma}12$ and H$_{\gamma}22$) are shifted strongly downfield by > 2.5 ppm, as demonstrated also for the C-terminal SH2 domain of p85 in complex with a high-affinity binding phosphopeptide (Breeze et al., 1996). However, Arg $\beta$B5 is completely buried from solvent in SH2 domain/phosphopeptide complexes, and additional stabilization of the Arg $\beta$B5 guanidinium group may occur through the Arg $\beta$B5 H$_{\gamma}11$/His $\beta$D4 N$_{\varepsilon}$ hydrogen bond. In cases in which this interaction may be surface-exposed and competing with solvent, NMR studies indicate that this bidentate interaction is weaker. For example, this interaction is observed in the crystal structure of the ternary complex of dihydrofolate reductase (DHFR), NADPH and methotrexate (Filman et al., 1982) involving the Arg 57 guanidinium group and a carboxyl group of methotrexate; four Arg 57 H$_{\gamma}$ resonances are observed and the inner H$_{\gamma}$ protons resonate at approximately 9.5 and 10 ppm (Nieto et al., 1996), but only at 0° C. Arg 46 of ribonuclease H is involved in a intra-molecular interaction of this type with Asp 102 as demonstrated in the
Figure 2 -- Arg Guanidinium Group Bidentate Hydrogen-Bonding Interactions. Geometry of bidentate hydrogen bonds involving arginine guanidinium proton donors and carboxyl or phosphate oxygen acceptors. (a) Involving the two inner Hη protons. (b) Involving the Hε and Hη11 protons.
crystal structure, but only two downfield-shifted H\(_1\) resonances (at 8.2 and 8.4 ppm) are observed (Yamazaki et al., 1993b).

With regard to the HE and NE resonances, a summary of \(^1\)HE and \(^{15}\)NE chemical shift changes in the arginine guanidinium region upon addition of phosphate or phosphopeptide is shown in Figure 3. These chemical shift differences are significant only for arginine residues involved in pTyr binding, which are Arg 18, Arg 37, Arg 39 and Arg 59. These four Arg guanidinium groups are affected to different extents by the phosphate group or by the rest of the phosphopeptide, most likely the phenyl group of the pTyr residue. For example, chemical shift differences for the Arg 37 HE, and \(^{15}\)NE resonances are much larger between spectra in imidazole and phosphate buffer than between spectra with phosphate and with phosphopeptide, although the chemical shift changes are upfield which is reverse to the expected chemical shift change for a hydrogen bond. Thus these protons are affected to a greater extent by the phosphate group. In contrast, chemical shifts of the Arg 59 HE and \(^{15}\)NE resonances are predominantly affected by the pTyr phenyl ring. Arg 18 and Arg 39 HE and \(^{15}\)NE resonances show very significant effects for both the phosphate group and the phenyl ring, although the effects on the Arg 39 resonances are larger than those of Arg 18.

The largest downfield shift of the HE resonance upon phosphate binding, involving Arg 39 (1.5 ppm), indicates that Arg 39 may be involved in a second type of bidentate hydrogen bonding interaction involves the HE and H\(_{\eta 11}\) protons (see Figure 2b) with the phosphate group both with free phosphate and the pTyr group in the pY1021 phosphopeptide. The presence of a hydrogen-bond involving the H\(_{\eta 11}\) proton hinders rotation both about the N\(_e\)-C\(_{\zeta}\) and C\(_{\zeta}\)-N\(_{\eta 1}\) bonds, but not to the extent that four H\(_{\eta}\) resonances are observed; instead one or two H\(_{\eta}\) resonances as observed. However, a large downfield shift of the HE proton is observed: the Arg 106 HE resonance is shifted to 8.56 in ribonuclease H, and the Asp 43 HE proton resonance is shifted to 9.1 ppm in the ternary DHFR/NADPH/methotrexate complex (Nieto et al., 1996a). In the PLCC SH2
Figure 3 -- Arg $^1H_\varepsilon$ and $^{15}N_\varepsilon$ Chemical Shift Differences. $^1H_\varepsilon$ (a) and $^{15}N_\varepsilon$ (b) chemical shift differences of arginine guanidinium groups of the PLCC SH2 domain of phosphate buffer versus imidazole buffer (black boxes), and the pY1021 complex versus free PLCC SH2 domain in phosphate buffer (gray boxes).
domain/pY1021 complex, two resonances are observed for the four H\(_{\eta}\) protons, indicative of an intermediate barrier to rotation about the N\(_{e}\)-C\(_{\zeta}\) and C\(_{\zeta}\)-N\(_{\eta}\) bonds, and the H\(_{e}\) resonance is shifted downfield by 1.5 ppm. In a molecular dynamics study of the PLCC SH2 domain/pY1021 complex, energy minimization of the NMR structure led to formation of this type of bidentate interaction involving the pTyr phosphate and Arg 39 guanidinium group which was maintained over the course of the simulation (Feng et al., 1996).

Large downfield chemical shifts for the two inner H\(_{\eta}\) protons of Arg 37 and the H\(_{e}\) of Arg 39 are the rationale for hydrogen-bonding restraints involving these protons and those of the phosphate oxygen atoms. These hydrogen bond restraints are shown in Figure 4 in the format required for input into the program X-PLOR (Brünger, 1992) for structure calculations. Note that the three hydrogens are restrained to be within 1.8 to 3.0 Å of any one of the four phosphate oxygens. This is a very conservative restraint as a 3 Å distance between a NH proton and an oxygen atom is only a very weak hydrogen bond.

**B. Backbone Amide Chemical Shift Changes**

Chemical shift changes of the backbone amide NH and \(^{15}\)N resonances were also observed between the three SH2 domain samples and are shown in Figures 5 and 6. Differences in chemical shift between the phosphate- and imidazole-buffered samples indicate that backbone NH and \(^{15}\)N atoms close to the phosphate-binding site are found in three regions, the N-terminus of the \(\alpha\)A helix and the loop N-terminal to this helix (the AA loop) involving residues Ser 15, Leu 16, Arg 18 and Glu 22, the C-terminus of the \(\beta\)B strand and the loop following it (the BC loop) involving residues Lys 38, Arg 39 and Glu 41, and the \(\beta\)C strand involving residues Ala 46 and Ser 48.

The chemical shift changes in backbone resonances are not due to direct hydrogen bonding to pTyr phosphate oxygens but to more long-range effects. The largest shift of a backbone NH between imidazole and phosphate buffer is for Glu 41, with an upfield shift
Figure 4 -- Phosphate-Oxygen Hydrogen Bond Restraints. Format of the hydrogen-bond restraints involving the H\textsubscript{n12} and H\textsubscript{n22} protons of Arg 37 and the H\textsubscript{e} proton of Arg 39 with any one of the four phosphate oxygen atoms. Distances between these atoms were restrained to between 1.8 and 3.0 Å, and between 2.8 and 4.0 Å for the nitrogen atoms attached to these protons. Note that 'segi "a"' is used to define the PLCC SH2 domain while 'segi "b"' is used to define the phosphopeptide, 'resi' represents residue number, ne, he, nh, and hh represent epsilon nitrogens and protons and eta nitrogens and protons, respectively, of the guanidinium groups. The oxygens attached to the phosphorus atom are named oh, oj1, oj2 and oj3, with oh being the phosphate oxygen attached to the $C\zeta$ atom of the phenyl ring.
assi (segi "a" and resi 37 and name hh12 ) (segi "b" and resi 4 and (name oj1 or name oj2 or name oj3 or name oh)) 2.3 0.5 0.7
assi (segi "a" and resi 37 and name nh1 ) (segi "b" and resi 4 and (name oj1 or name oj2 or name oj3 or name oh)) 3.3 0.5 0.7
assi (segi "a" and resi 37 and name hh22 ) (segi "b" and resi 4 and (name oj1 or name oj2 or name oj3 or name oh)) 2.3 0.5 0.7
assi (segi "a" and resi 37 and name nh2 ) (segi "b" and resi 4 and (name oj1 or name oj2 or name oj3 or name oh)) 3.3 0.5 0.7
assi (segi "a" and resi 39 and name he ) (segi "b" and resi 4 and (name oj1 or name oj2 or name oj3 or name oh)) 2.3 0.5 0.7
assi (segi "a" and resi 39 and name ne ) (segi "b" and resi 4 and (name oj1 or name oj2 or name oj3 or name oh)) 3.3 0.5 0.7
Figure 5 -- Backbone NH Chemical Shift Differences. $^1$H chemical shift differences of backbone NH amide resonances of the PLCC SH2 domain in (a) phosphate buffer versus imidazole buffer, and (b) in complex with the pY1021 peptide in imidazole buffer versus free in phosphate buffer.
Figure 6 -- Backbone $^{15}$N Chemical Shift Differences. $^{15}$N chemical shift differences of backbone NH amide resonances of the PLCC SH2 domain in (a) phosphate buffer versus imidazole buffer, and (b) in complex with the pY1021 peptide in imidazole buffer versus free in phosphate buffer.
15N Chemical Shift Difference (ppm)

Residue Number

254
(not indicative of hydrogen bonding). The largest downfield shift for a backbone amide is the Arg 18 NH, which, based on the fact that the NH of the homologous residue in other SH2 domain/phosphopeptide complexes (Eck et al., 1993; Waksman et al., 1992; Waksman et al., 1993) is 4.2-4.7 Å away from the closest phosphate oxygen, is not representative of a hydrogen bond interaction.

Residues whose backbone resonance change between the imidazole and phosphate samples are not localized only to the surface contacted by the phosphate group, but also include atoms in the AA and BC loops which are distant from the phosphate group in calculated structures. This may be due to fluctuations of the position of the phosphate group relative to the peptide backbone or conformational adjustments which occur upon phosphate binding, such as changes in hydrogen bonds. In MD simulations of the PLCC SH2 domain, changes in hydrogen bonding around residues Leu 16 and Ser 48 were observed between the free and pY1021 peptide complexed domains (Feng et al., 1996). The $^{15}$N-$^1$H order parameters for the backbone amide of Leu 16 also decreased upon addition of the phosphopeptide (Farrow et al., 1994). However, the change in chemical shift for Glu 41 cannot be explained from the structure, from measurement of order parameters or from the MD simulations.

Differences in chemical shift involving the peptide-bound SH2 domain in imidazole buffer versus the free SH2 domain in phosphate buffer arise in a number of regions. Chemical shift changes were observed in the backbone resonances near the EF loop (Leu 69) and in the BG loop (Leu 89 and Tyr 90), most likely due to interaction of residues +1 to +5 of the phosphopeptide with either backbone Hα or side-chain protons of these residues in the SH2 domain. These results are expected in light of the protein-peptide interactions observed (see Chapter 4; Pascal et al., 1994). Changes in chemical shift involving NH and $^{15}$N resonances of residues in the pTyr-binding pocket such as in the βB (Arg 39), βC (Ala 46) and βD (Lys 56, His 57, Cys 58 and Arg 59) strands are possibly due to ring current effects from the pTyr phenyl ring or from effects of the
pY1021 peptide backbone, likely involving the peptide bond between the -1 and the pTyr residue, and the pTyr and the +1 residue. The specific causes of each chemical shift change cannot be discerned, but it would be expected that chemical shift effects of the phenyl ring of the pTyr are important because NOEs have been observed from the pTyr phenyl ring protons to Hα protons of these residues. In addition, structure calculations have indicated that backbone atoms of these residues contact or are close (<6 Å) to the pTyr phenyl ring.

C. Arg 37 Stereoassignment

In general, definition of the position of Arg guanidinium groups using NOE restraints is expected to be poor due to the inability to distinguish which of the four Hη protons gives rise to NOEs. This is the result of motion about the partial double bonds in the Arg guanidinium group which leads to observation of a single Hη resonance in most cases (Kanamori & Roberts, 1983; Yamazaki et al., 1995). In such cases, a pseudoatom correction of 2.4 Å is used in X-PLOR calculations. However, since Arg 37 has four distinct Hη resonances, stereoassignment of the Arg 37 Hη resonances can be performed relying on distances to the Hε, as demonstrated in Figure 7. If the positions of the arginine Hη's are held fixed, distances of 2.26, 3.46, 3.98 and 3.45 Å are observed between the Hε and the Hη11, Hη12, Hη22 and Hη21 nuclei, respectively. Thus I used the small distance between the Hε and the Hη11 protons in stereoassignment.

Figure 8 shows a portion of a 2D NOESY spectrum of a 13C-15N labeled sample of the PLCC SH2 domain/pY1021 complex which selected for NOEs to protons attached to 15N nuclei of guanidinium groups. From the HSQC shown in Figure 1c, it can be inferred that Hη resonances at 5.75 and 9.67 ppm, and at 7.47 and 9.47 ppm are attached to the same Nη. In Figure 7, a large NOE is observed between the Arg 37 Hε and the Hη at 5.75 ppm, this being the largest NOE from the Arg 37 Hε. Thus, Hη11 was assigned to the
Figure 7 -- Geometry of the Arg Guanidinium Group. Generally, the H$_\eta$ protons in Arg residues average to a single resonance due to motion about the N$_{e}$C$_{z}$ and C$_{z}$N$_{\eta}$ bonds. However, if the guanidinium group of the Arg is held in place by hydrogen bonds such that four different resonances can be observed for the H$_{\eta}$ protons, the resonances can be stereoassigned by making use of the distance relationships shown. The following distances are found for Arg residues using CHARMM22 parameters (Brünger, 1992).
Figure 8 -- Arg 37 Stereoassignment. Region of the Arg-specific 2D $^{15}$N-NOESY showing intra-residue NOEs involving the guanidinium group of Arg 37. Note the NOE involving the Arg 37 $\mathrm{H_e}$ and an Arg 37 $\mathrm{H_\eta}$ resonance at 5.75 ppm, which was useful in stereoassignment of the Arg 37 $\mathrm{H_\eta}$. 
resonance at 5.75 ppm and \( H_\eta 12 \) was assigned to the resonance at 9.67 ppm.

The other two \( H_\eta \) protons (at 7.47 and 9.47 ppm) can be assigned to one of two \( N_\eta 2 \)-attached protons. A large NOE between the \( \eta 12 \) and \( \eta 22 \) protons is expected, as they are separated by a distance of approximately 2.3 Å (see Figure 7). However, most magnetization from the \( H_\eta 12 \) proton would be expected to transfer to the geminal proton, and only some (approximately 6-fold less assuming an \( r^{-6} \) dependence) would transfer from \( H_\eta 12 \) to \( H_\eta 22 \). In addition, chemical exchange due to rotation about the \( C_\zeta -N_\eta \) bond would be expected to reduce the intensity of this NOE as well. Such an NOE is observed at 6-fold poorer intensity, but at this signal-to-noise level it is difficult to rely on this for assignment. However, one can also use chemical shift arguments to identify the \( H_\eta 22 \) proton. A bidentate hydrogen bond from a phosphate oxygen would be expected to interact with the two inner \( H_\eta \) protons of the FLVR arginine as has been found from crystal structures (Waksman et al., 1993), and this hydrogen bonding should result in a strong downfield shift of both protons. Thus, one can assign the Arg 37 resonances at 7.47 and 9.47 to \( H_\eta 21 \) and \( H_\eta 22 \) respectively. This stereoassignment, however, is not consistent with ab initio chemical shift calculations performed on the Arg 37 guanidinium group in the PLCC SH2 domain/pY1021 complex (Feng et al., 1996), in which the resonances at 7.47, 9.47, 9.67 and 5.75 were assigned to \( H_\eta 11 \), \( H_\eta 12 \), \( H_\eta 21 \) and \( H_\eta 22 \), respectively. The reason for this discrepancy between calculated chemical shifts and experimental assignments is not clear.

The weak NOE involving the Arg 37 \( H_\eta 12 \) and \( H_\eta 22 \) protons was also accompanied by an Arg 37 \( H_\eta 12/H_\eta 21 \), and the two NOEs are of similar intensity. The \( H_\eta 21 \) is an additional 1.2 Å from \( H_\eta 12 \), and thus this NOE must arise from chemical exchange due to rotation about the \( C_\zeta -H_\eta \) bonds. Thus, of the four Arg 37 \( H_\eta \) resonances, the \( H_\eta 11 \) and \( H_\eta 12 \), and the \( H_\eta 21 \) and \( H_\eta 22 \) protons, respectively, were considered equivalent for purposes of structure calculations, and 1 Å pseudoatom corrections were applied.
Figure 9 -- Intra-residue Arg Guanidinium Group NOEs: The same region of the Arg-specific 2D $^{15}$N-NOESY as shown in the previous figure highlighting intra-residue NOEs involving the guanidinium groups of Arg 18, 27, 39, 50 and 59.
For other arginine guanidinium group protons, intra-residue interactions of some of these protons are shown in Figure 9. A number of arginine H_e/H_\eta NOEs are observed with two separate H_\eta resonances for Arg 39 having NOEs of different intensity. This information is not sufficient for stereoassignment for Arg 39 H_\eta resonances, and any NOE involving the Arg 39 H_\eta nuclei was interpreted with all protons considered equivalent and a 2.4 Å pseudoatom correction was applied.

D. NOEs Involving the Arginine Guanidinium Group Protons

I also used the Arg-specific NOESY spectrum in assignment of inter-residue NOEs involving arginine H_\eta protons. In Figure 10 the same region of the Arg-specific NOESY spectrum as shown in Figures 7 and 8 is presented, highlighting a number of inter-residue NOEs. In particular, NOEs were found between the guanidinium groups of Arg 18 and Arg 37. These NOEs were very important in placing the Arg 18 guanidinium group in the pTyr-binding region, because of the lack of protein-peptide NOEs involving the Arg 18 backbone or sidechain (see Chapter 3 and Pascal et al., 1995). In addition, NOEs involving the arginine guanidinium protons of Arg 18 and Arg 37 and the His 57 imidazole ring were useful in interpretation of the reduced pK_a observed for His 57 (see previous chapter, Singer & Forman-Kay, 1997) and confirmation of interactions observed in crystal structures of other SH2 domains thought to play a role in this reduced pK_a.

In addition, I observed NOEs to hydroxyl groups of Ser residues, between Arg 37 H_e and Ser 48 OH and between Arg 39 H_\eta and Ser 44 OH. These NOEs were predicted from the average structure of a 130 ps molecular dynamics simulation of the PLCC SH2 domain/pY1021 complex (Feng et al., 1996). In addition, the Arg 37 H_e/Ser 48 OH NOE confirms interactions discussed in the previous chapter surrounding the imidazole ring of His 57. The presence of these OH resonances was confirmed by observation of intra-residue H_\alpha/OH and H_\beta/OH NOEs in the 3D CN-NOESY for both residues. Ser OH
Figure 10 -- Inter-residue Arg Guanidinium Group NOEs: Arg-specific 2D $^{15}$N NOESY. The same region of the Arg-specific 2D $^{15}$N-NOESY as shown previously, highlighting a number of inter-residue NOEs involving stereoassigned guanidinium protons of Arg 37. In addition, NOEs involving the His 57 imidazole ring and guanidinium protons of Arg 37 and Arg 59 are also shown.
groups are generally in rapid exchange with water (Wüthrich, 1986), however if the Ser OH group is buried or involved in hydrogen-bonding interactions, its resonance becomes visible. Since these NOEs were only recently assigned, they were not included in previous structure calculations.

In addition to the NOEs shown in this figure, there are several inter-residue NOEs involving the guanidinium group protons in the phosphate-binding region (i.e. Arg 18, 37, and 39) and aliphatic protons. Because this is a two-dimensional experiment, the identity of these protons could not be established with certainty. Thus I re-analyzed the 3D 13C-NOESY, gradient-enhanced 15N-NOESY and a gradient-enhanced CN-NOESY. Figure 11 shows strips from the 13C-NOESY highlighting NOEs to the Arg 37 and Arg 39 guanidinium groups and His 57 imidazole group. The Glu 22 HΔ/His 57 HE1, Ser 48 Hβ2/His 57 HE1 and Arg 37 HE/Ser 48 OH NOEs (see Figures 10 and 11) are consistent with the hydrogen-bonding network shown in the previous chapter. In addition, one can observe a number of NOEs to the Arg 37 Hη21 proton (at 7.47 ppm) which are useful in defining the position of this guanidinium group, although these NOEs are input in structure calculations as involving protons Hη21 and Hη22, as discussed previously.

The NOEs I observed between the Arg 37 guanidinium groups and His 57 imidazole groups observed in the arginine-specific 2D 15N-NOESY and the 3D experiments are summarized diagrammatically in Figure 12. Arg 37 contacts side chains of residues on one side of the βB strand, including the Leu 35 methyl groups and the Arg 39 guanidinium groups, as well as the βC strand (specifically the Ala 46 methyl group). However, Arg 37 also interacts with NOEs to the side chain and backbone protons of the N-terminus of the αA helix. Similarly, contacts were observed between the His 57 guanidinium group and side chains of residues on the same side of the βD strand, specifically involving the methyl groups of Ile 55 and the guanidinium group of Arg 59, as well as the βC strand (Ser 48), and across to the αA helix, including Arg 18 and Glu 22.
Figure 11 -- Inter-residue Arg Guanidinium Group NOEs: 3D $^{13}$C NOESY. Strips from the $^{13}$C NOESY of the PLCC SH2 domain/pY1021 complex. A number of NOEs involving the Arg 18 and Arg 37 guanidinium groups and the His 57 imidazole ring are highlighted. Unrelated peaks from close $^1$H and $^{13}$C shifts are marked with an X.
Figure 12 -- Summary of Arg 37 and His 57 Inter-residue NOEs. The figure was generated using the program ChemDraw (Rubenstein, 1988).
Figure 13 -- Superposition of Residues Arg 18, Arg 37 and Arg 39 from the Most Recent Structure Calculations. The Arg 37 guanidinium group occupies a single conformer, while multiple conformers are observed for the guanidinium groups of Arg 18 and Arg 39.
Inclusion of these Arg-specific and His 57 NOEs has slightly increased the precision of determination of the PLCC SH2 domain/pY1021 structure in the pTyr-binding region. In Figure 13 is shown the superposition of the side chains of Arg 18, 37 and 39 in the 25 lowest-energy structures. Arg 37 is well defined, but greater variability is observed in the positions of Arg 18 and Arg 39, with the side chain of Arg 18 being the most poorly defined.

E. Solvent Protection of the Arginine Guanidinium Groups

A 3D version of the Arg guanidinium-selective NOESY shown in Figures 8-10, in which the NOE crosspeaks are separated in F2 by the $^{15}$N chemical shift, was performed by Toshio Yamazaki. This gave essentially the same results as the 2D experiment, however, this spectrum also yielded NOE between water and guanidinium protons. These crosspeaks primarily reflect exchange between water and guanidinium protons and thus can be used as a measure of surface accessibility of the guanidinium groups. In Figure 14 the intensities of the water signal arising from various protons is plotted, and from this I observed that most arginine $H_e$ and $H_\eta$ protons are solvent exposed i.e. have a large crosspeak to the water resonance. The $H_e$ protons of Arg 18 and Arg 39 show a smaller peak to water indicating greater burial from the bulk solvent. In addition, all protons from the Arg 37 guanidinium group with the exception of Arg 37 $\eta_{11}$ are protected from exchange. In NMR structures of the PLCC SH2 domain/pY1021 complex, the $H_{\eta_{11}}$ proton is buried from solvent. This is evidence of the presence of either a bound water or a hydroxyl group in close contact with the $H_{\eta_{11}}$ proton. Bound waters to the arginine guanidinium groups could not be detected in the PLCC SH2 domain/pY1021 complex using the techniques described in Pascal et al. (1997) because of their lower signal-to-noise relative to the NH groups (Steve Pascal, personal communication). In addition, the $H_{\eta_{11}}$ proton is severely broadened so that its intensity is among the weakest of all of the arginine
Figure 14 -- Guanidinium Group Solvent Protection. Intensities of crosspeaks from water to each of the Arg guanidinium proton resonances, with the exception of the Arg 18 Hη and Arg 96 Hε resonances since these were very broadened, due to conformational exchange in the case of Arg 18 Hh and exchange with solvent in the case of Arg 96 Hε (see the experiments on solvent exchange as a function of pH in Chapter 5) Note that stereoassignment of the Arg 39 Hη and 15Hη resonances was not performed, but that in this figure, the Hη resonance at $\delta = 7.30 \text{ ppm}$ is designated $H_{\eta a}$ and the Hη resonance at $\delta = 6.75 \text{ ppm}$ is designated $H_{\eta b}$.
H$_7$ proton resonances. The presence of a bound water exchanging at an intermediate time scale may explain its broadening. In the structures of high-affinity binding phosphopeptide complexes of syp N-terminal (Lee et al., 1994), src (Waksman et al., 1993) and ick (Eck et al., 1993) SH2 domains, the Arg $\beta$B5 H$_{\eta}$11 atom is buried and not within 4 Å of a bound water molecule. However, in the molecular dynamics simulations of Feng et al. (1996), one bound water is found hydrogen-bonding to the Arg 18 H$_{\eta}$ and close to Arg 37 H$_{\eta}$11, while another is found hydrogen bonding to Arg 37 H$_{\eta}$21. The results of NH exchange protection studies are consistent with the presence of a bound water near Arg 37 H$_{\eta}$11 but not Arg 37 H$_{\eta}$21.

F. Role of the Arginine Residues

Four arginine residues are present within the phosphate-binding pocket of the PLCC SH2 domain. Since the phosphate group has a net charge of -2 (see Chapter 5 and Singer & Forman-Kay, 1997) and each of the four arginine guanidinium groups has a +1 charge, an excess of positive charge would result. In the following section, I will discuss the role of each residue in pTyr binding based on evidence in this chapter, previous chapters and other studies, both of the PLCC SH2 domain/pY1021 complex and other SH2 domains.

Arg 37 (Arg $\beta$B5) is highly conserved among SH2 domains and structural studies indicate this residue forms a very strong bidentate interaction with the pTyr phosphate group of the form shown in Figure 2a in all SH2 domain/phosphopeptide complexes studied. The role of this residue in pTyr binding has also been confirmed by mutagenesis studies (Marengere & Pawson, 1992, Mayer et al., 1992). This interaction is especially strong because the Arg guanidinium group is completely buried from solvent, as shown in calculated structures and studies demonstrating that the Arg 37 guanidinium group exchanges very poorly with solvent (Pascal et al., 1995, Singer et al., 1997). Thus motion about rotatable bonds in the Arg 37 guanidinium groups is restricted, as demonstrated by
the presence of four H\textsubscript{\Pi} proton resonances and a high order parameter (S\textsuperscript{2}) for the 15N\textsubscript{E}-H\textsubscript{E} of Arg 37 of 0.80 ± 0.01 (Pascal et al., 1995). Resonances of the Arg 37 aliphatic side chain appear to be broad indicative of molecular motion on the μs to ms timescale (Pascal et al., 1995), so the results of structure calculations showing the Arg 37 side chain adopting a single conformation is somewhat surprising. In fact, NOEs involving this residue may be present only a fraction of the time, indicating that structure calculations involving time-averaged (Torda et al., 1990) or ensemble-averaged NOEs (Bonvin et al., 1994) may be more appropriate (see Chapter 7). It is clear that Arg 37 is a very important residue in pTyr phosphate recognition.

Arg 39 is also involved in hydrogen bonding with the pTyr phosphate group, and one of the hydrogen bonding partners in this interaction is the H\textsubscript{E} proton. The interaction of the Arg 39 guanidinium group with the phosphate group may be of the form shown in Figure 2b. This interaction is not as strong as that shown for the Arg 37 guanidinium group, since it shows two, not four H\textsubscript{\Pi} resonances, and the 15N\textsubscript{E}-H\textsubscript{E} S\textsuperscript{2} is also lower (S\textsuperscript{2} ~ 0.7, Pascal et al., 1995). In addition, the Arg 39 guanidinium group is not as buried from solvent, based on solvent exchange experiments (see also Chapter 5 and Pascal et al., 1995). However, chemical shift evidence presented in this chapter, as well as protein-peptide NOEs between the pTyr ring and the Arg 39 H\textsubscript{\Phi} protons (see Chapter 4) indicate that the guanidinium group of this residue is also in contact with the pTyr ring.

The role for Arg 18 in pTyr binding is less clear. Chemical shift evidence indicate its role both in both the pTyr phenyl ring and the pTyr phosphate group. However, the chemical shift changes due to phosphate are not as large as observed for Arg 37 and Arg 39, and thus it is not clear what interactions Arg 18 is making with the phosphate group. Similar order parameters and exchange rates were reported for Arg 39 and Arg 18, indicating that the interactions may be of similar strength (Pascal et al., 1995). However, there is evidence that the Arg 18 interactions are weaker than those of Arg 39, namely that the H\textsubscript{\Pi} resonance is split into two only at low temperature, as well as the exchange
experiments as a function of pH described in the previous chapter. Molecular dynamics calculations of the PLCC SH2 domain/pY1021 complex have shown Arg 18 to make only one hydrogen bond to the pTyr phosphate group and another to the Asp -1 carboxyl group (Feng et al., 1996). Structure calculations indicate the proximity of the Arg 18 guanidinium group to the carboxyl group of the Asp -1 residue of the pY1021 residue of the phosphopeptide. Binding studies have shown Asp -1 to be important in pY1021 binding, and replacement of Asp -1 by a Ala results in 5.6-fold poorer binding to the PLCC SH2 domain (see Chapter 4, Table 1). Since the pTyr phosphate exists in the -2 state in the PLCC SH2 domain/pY1021 complex (see Chapter 5), the role of Arg 18 helps to answer questions in terms of the balance of charges in the pTyr binding pocket; 3 Arg residues (+1 X 3 = +3 ) interact with the pTyr phosphate group and a carboxyl group of a residue N-terminal to the pTyr (-2 + -1 = -3).

Finally, the guanidinium group of Arg 59 is not involved in phosphate binding based on the evidence presented in this chapter. As well, Arg 59 appears to be mobile in solution and solvent exposed. In the structures of other SH2 domain/phosphopeptide complexes, this residue (position βD5) is frequently a Lys residue which in fact forms a weak hydrogen bonding interaction involving the Ne amide with the center of the pTyr ring. Arg 59 may be making hydrophobic and weak hydrogen bonds with the pTyr phosphate group based on the large number of protein-peptide bonds observed for Arg 59 (see Chapter 4) as well as chemical shift evidence presented earlier that the Arg 59 He and Ne is close to the pTyr phenyl ring. However, the single resonance for the Arg 59 Hη protons as well as its near random chemical shift and the poor solvent protection of the Arg 59 guanidinium group indicates that the interaction of the guanidinium group with the pTyr phenyl ring is transient and competes with interactions with solvent.
IV. Summary

The interaction of the arginine guanidinium groups of the PLCC SH2 domain with the phosphate group of the pTyr was studied in detail by a number of methods motivated by the lack of information concerning this interaction from peptide-protein NOEs. We analyzed Arg guanidinium resonances and the mobility of the guanidinium groups and protection from exchange with solvent. A number of NOEs were observed involving these guanidinium groups which were important in specifying their position in the PLCC SH2 domain/pY1021 complex. Chemical shift changes between the free SH2 domain in imidazole and in phosphate buffer as well as the PLCC SH2 domain/pY1021 complex in imidazole buffer have been interpreted in terms of the contributions to the chemical shift of the phosphate group and the pTyr aromatic ring or peptide backbone. We conclude that Arg 37 (Arg B5) is essential in pTyr binding, forming at least two strong hydrogen bonds with the phosphate group which are not affected by exchange with water, since this residue is buried from solvent. Arg 39 (Arg B7) is also important in pTyr binding, forming at least one strong hydrogen bond and possibly two hydrogen bonds with the phosphate group. Arg 18 (Arg a2) may interact more weakly with the phosphate group, but also may interact with the carboxyl group of the -1 residue. Finally, Arg 59 (Arg D6) is more important for recognition of the pTyr phenyl ring.
Chapter 7: Summary and Future Directions

I. Summary

In previous chapters, I have described the solution structure of the PLCC SH2 domain pH1021/pH1021 complex. In the discussion of the solution structure, I emphasized the many interactions responsible for binding specificity. SH2 domain specificity is a key event in the early stages of signal transduction, as binding of different SH2 domain-containing proteins dictates which downstream pathways in signal transduction, such as those involving ras or protein kinase C, will be activated following growth factor stimulation. SH2 domains contain similar folds, however they contain a wide variety of binding interfaces, from the flat hydrophilic surface found for GRB2, to surfaces predominantly flat but contain small binding pockets for hydrophobic amino acids as found among the SH2 domains of the cytoplasmic tyrosine kinases such as src and lck, to SH2 domains which contain extensive hydrophobic surfaces such as observed for the syp SH2 domains and the PLCC SH2 domain. In the PLCC SH2 domain, the presence of a large hydrophobic pocket dictates a hydrophobic binding site, and further sequence specificity within the large hydrophobic binding groove is observed at the +1 and +3 positions.

pTyr binding is common among SH2 domains, and thus contributes to the binding but not specificity of the PLCC SH2 domain. Because of the large role of ionic interactions, description of the mechanism of pTyr binding could not be made solely on the basis of NOEs, and a number of other methods were used. We observed four Arg residues in the pTyr-binding pocket, of which two are involved in hydrogen-bonding with the phosphate group (Arg βB5 and Arg βB7), one is involved strictly in interactions with the pTyr aromatic ring (Arg βD6), and one is involved in a number of interactions including with the pTyr phosphate group and aromatic group and possibly the Asp -1 carboxyl group.
(Arg αA2). A study of the electrostatics of this complex showed that (a) the pTyr phosphate group is bound in the -2 state, and (b) that a low pKₐ value is observed for the histidine near the pTyr binding site (His βD4). This low pKₐ is due to a hydrogen-bonding network involving a number of residues well-conserved among SH2 domains which serves to position Arg βB5 for pTyr binding.

II. Future Directions

In this chapter I discuss a number of future directions which may be undertaken to better understand either the binding of this SH2 domain or the role of this SH2 domain with respect to the rest of the PLC-γ molecule.

A. Electrostatic Interactions

As stated throughout the thesis, NOE-based NMR methods are generally poor for defining side-chain hydrogen-bonding interactions or salt bridges in proteins. This difficulty arises because these interactions rarely involve or can be defined by a pair of non-exchangeable hydrogen atoms. In the study of SH2 domains, the determination of these hydrogen-bonding interactions is necessary for a description of the mode of recognition of the phosphate group of pTyr. In addition, a number of SH2 domains, in particular the group I SH2 domains of Songyang et al. (1994) interact with hydrophilic or acidic residues C-terminal to pTyr by means of surface salt bridges: these would likely be poorly defined in an NOE-based structure determination. In the following section, I will describe three possible strategies to overcome these drawbacks.
1. **Hydrogen-bonding Restraints**

By chemical shift arguments I have shown that pTyr phosphate binding involves contributions principally by two Arg residues. It has been postulated (Feng et al., 1996) that these residues make bidentate hydrogen bonds with the phosphate group, and that these bidentate hydrogen bonds could take two forms, designated Type I and Type II. One may be able to distinguish these interactions from each other and from monodentate hydrogen-bonding interactions based on the number of H\(\eta\) resonances and from chemical shifts of H\(\epsilon\) or H\(\eta\) resonances. However, this is based on a comparison of crystal structures and NMR data of a small number of proteins. In future the comparison of crystal structures to chemical shift behavior of Arg guanidinium group resonances must be expanded. In addition, from examination of high-resolution structures, we may be able to obtain more precise restraints for these bidentate restraints, since the Arg/phosphate hydrogen-bonding restraints used in the calculation of the PLCC SH2 domain/pY1021 solution structure were quite loose.

2. **Inclusion of Electrostatic Energy into Structure Calculations**

As stated previously, NMR structure calculations typically contain energy terms for the experimental restraints as well as additional terms for maintenance of the chemical geometry and a term to prevent atomic overlap. The rationale for this force field is that the structure is based predominantly on experimental restraints, with all other restraints being used to make sure that the calculated structures have reasonable geometries and have no overlapping atoms. In such a force field, terms for the electrostatic energy are ignored, however protocols have been developed for inclusion of electrostatic energy terms within X-PLOR (Michnick et al., 1991). In a preliminary calculation of the PLCC SH2 domain/pY1021 complex using one such protocol, the electrostatic energies were very large and
negative relative to the energies due to the experimental restraints, so that calculated structures had lower overall energies but significantly increased NOE energies. The force constant for the electrostatic energy term in such calculations should therefore be scaled down by a factor of >10. An additional point to note is that electrostatic interactions may be less significant at the protein surface due to screening of charge-charge interactions by solvent. One may include water molecules in such a calculation to mimic this physical situation.

3. Chemical Shift Restraints

Use of hydrogen-bonding restraints for downfield Hε and Hη chemical shifts implies knowledge of the hydrogen-bond acceptor to which the Hε and Hη are interacting. Such information may be inferred from high-resolution crystal structures of identical or homologous molecules or molecular complexes, or from molecular dynamics simulations. However, in these approaches no experimental NMR data per se is being used to determine the hydrogen-bonding partners.

We had observed that interactions of the guanidinium group protons and 15N nuclei with the phosphate group resulted in a number of upfield and downfield chemical shifts; similar shift changes can occur due to interactions with carboxyl and carbonyl groups as well (Williamson & Asakura, 1993). The effect of carbonyl and carboxyl groups on chemical shifts can be determined using a semi-classical formula whose parameters have been determined empirically (Osapay & Case, 1994; Williamson & Asakura, 1993; Williamson et al., 1992), or by quantum mechanical calculations (Le et al., 1995). Recently, a number of authors have suggested the use of these formulae to derive chemical shift restraints in structure calculations. This involves calculation of chemical shifts, with the difference between the calculated and experimental chemical shifts converted into an energy which must be minimized. Such an approach has been developed for X-PLOR
(Kuszewski et al., 1995). However, these calculations are not yet feasible for SH2 domain/phosphopeptide complexes because parameters for phosphate groups have not been developed. In addition, it would be expected that the partially conjugated double bond system of the arginine guanidinium group would have an effect on chemical shifts of nearby NMR-active nuclei. However, parameterization of the effects of guanidinium groups have also not been performed.

B. Time-averaged or Ensemble-averaged NOEs

Typical NMR structure calculations assume that experimental contacts remain fixed or undergo small harmonic oscillations over time. However, this assumption is not valid in regions of the protein which undergo substantial motion. In the case of the SH2 domain/phosphopeptide interaction, a number of such regions were identified. For example, the methyl groups of Leu 69 and Leu 77 have been observed to have very low order parameters and coupling constants indicative of rotamer averaging in spite of their being buried from solvent (Kay et al., 1996). This is of note because the methyl groups of Leu 69 are also observed to be in contact with the phosphopeptide (see Chapter 4). The $N_{\text{e}}$ of Arg 59 has a very low order parameter indicating conformational averaging but high and medium intensity NOEs are observed for the Arg 59 H$_{\gamma}$ and H$_{\delta}$ protons with the pTyr aromatic ring protons. In fact all guanidinium protons showed evidence of motion, since the linewidth of and number of H$_{\eta}$ and $^{15}$N$_{\eta}$ resonances are indicative that hydrogen bonds involving the H$_{\eta}$ protons are transient. Finally, though NOEs are observed between aliphatic protons of the -1 and +4 to +6 residues of the phosphopeptides, the chemical shifts of these protons are virtually identical to random coil values for their equivalent amino acids. Thus the experimental evidence points to significant motion occurring both in the protein and at the protein-peptide interface.
In order to accommodate internal motions of the proteins in structure calculations, two strategies may be employed. The first involves use of NOEs and other restraints as distance restraints which must be satisfied not instantaneously but on average over the course of a molecular dynamics trajectory. Such a trajectory may be exponentially weighted so that the time-averaged distances are weighted to more recent events in the molecular dynamics calculation (Torda et al., 1990). A protocol for use of time-averaged distance, but not dihedral, restraints has been developed for use in X-PLOR. However, this method suffers from the drawback common to all molecular dynamics simulations, in that times on the order of picoseconds to nanoseconds can be covered over the trajectory, and some motions in proteins are on the μs to ms time scale. A second strategy involves calculation of a composite of structures, in which the composite but not the individual structures contain all the experimental restraints. Thus the restraints are not time-averaged, but ensemble-averaged. The structures calculated then can be grouped into families based on how well they superimpose and which restraints are or are not satisfied. Ensemble-averaging protocols for proteins (Bonvin et al., 1994; Fennen et al., 1995) and nucleic acids (Ulyanov et al., 1995) have been developed. This approach attempts to address the sampling problem based on the limited time scale in a time-averaged calculation.

C. Mapping the Binding Surface and Experimental Drug Design

In the previous chapters I have described the solution structure of the PLCC SH2 domain with one high-affinity binding phosphopeptide, the pY1021 peptide from the β-PDGFR. Dr. Yves Aubin is currently determining the structure of the PLCC SH2 domain with the pY992 peptide from the EGFR. Analysis of half-filtered NOESY spectra indicate that the pY992 peptide adopts a different path along the SH2 domain surface beyond the +3 position, in that instead of contacting residues BG1 and BG4 of the SH2 domain, it interacts with the aromatic ring of Tyr αB8. The sequences of the pY1021 and pY992
phosphopeptides diverge beyond the +3 position, so this difference in binding is not surprising. However, this indicates that the PLCC SH2 domain peptide binding surface can accommodate a number of different phosphopeptides. For that reason, it would be of interest to determine how other high-affinity binding phosphopeptides bind to the PLCC SH2 domain surface. Other sites of tyrosine phosphorylation which bind to the PLCC SH2 domain with high affinity include the pTyr 1536 site of the hepatocyte growth factor receptor (Borrello et al., 1996) and pTyr 783 site of PLC-γ (Neil Farrow and Gerry Gish, personal communication). With several structures, one may map a more complete binding surface of the PLCC SH2 domain. Mapping the surface of the PLCC SH2 domain or any other SH2 domain may be useful for designing high-affinity inhibitors of binding for the SH2 domain, since optimal binding should be achieved when a maximum surface is buried and when the various hydrogen-bonding and electrostatic groups on the protein surface are utilized.

D. Mutational Studies

From the study of the PLCC SH2 domain/pY1021 complex, we have identified a number of protein-peptide interactions responsible for high-affinity binding and specificity. From this work, we may be able to perform site-directed mutagenesis in order to alter specificity or increase binding. A mutant of the PLCC SH2 domain in which Cys βD5 is altered to a Tyr was shown to drastically alter specificity from a Type III SH2 domain to a type I SH2 domain, similar to the specificity observed for the src SH2 domain (Songyang et al., 1995). However, a number of other mutations could be made which would alter the PLCC SH2 domain specificity in more subtle ways. I previously mentioned mutations of residues involved in the +1 binding pocket, and in particular stated that mutation of the Tyr BG3 residue would be of interest. In addition, BG loops in SH2 domains are variable both in terms of length and sequence conservation. The BG loop of the PLCN SH2 domain is
longer, and peptides binding the PLCN SH2 domain show higher-affinity binding beyond the +3 residue (see Chapter 4). Thus swapping of BG loops between different SH2 domains would be of interest in this case.

Finally, in Chapter 6, I postulated that of the four Arg residues in the phosphate binding site, Arg βD5 was strictly involved in binding the aromatic ring of the pTyr residue, and of the three remaining Arg residues, Arg αA2 was probably the least important in pTyr binding. This could be tested by mutational analysis.

E. Phosphopeptide Dynamics

$^{15}$N and $^2$H-methyl relaxation measurements on the SH2 domain of the PLCC SH2 domain/pY1021 complex have been performed. Complementary information about motions involved in peptide binding may be obtained by performing relaxation experiments using isotopically-labeled phosphopeptide. Relaxation measurements of backbone $^{15}$N nuclei of the PLCC SH2 domain both free and in complex with the pY1021 phosphopeptide showed that the secondary structural elements were quite rigidly held in place (Farrow et al., 1994), while loop regions were generally more flexible. This has been observed in relaxation studies of many proteins and protein complexes. However, the phosphopeptide may experience a greater range of motions depending on the strength of interaction with the protein. The peptide residues can be subdivided into those which did not bind to the protein (residues -3, -2, +7, +8), those which bound tightly to the protein, as indicated by a large number of protein-peptide NOEs and chemical shift changes upon SH2 domain binding (residues pTyr, Ile +1 and Pro +3) and those which bound more weakly, as indicated by a smaller number of protein-peptide NOEs and random-coil chemical shifts of aliphatic protons (residues Asp -1, Ile +2, Leu +4 to Glu +6). One would expect very low order parameters from $^{15}$N nuclei in the first group, and order parameters approaching SH2 domain values for the second group. Intermediate-binding
residues might be expected to yield order parameters which are intermediate between the two groups. In addition, in a molecular dynamics simulation of the DNA/Antennapedia complex (Billeter et al., 1996), Kurt Wüthrich's group has observed that lifetimes of hydrogen-bonding interactions were longer than those of hydrophobic interactions, which underwent numerous rapid fluctuations. The binding of the pTyr site is heavily influenced by charge-charge interactions, while binding at the +1 and +3 sites is dictated strongly by hydrophobic interactions. Motional differences at these sites may be reflected in differences in order parameters or exchange broadening. The broadening of the NH of Ile +1 and the Hα of pTyr may be relevant in this regard.

However, such experiments require affordable and efficient production of 15N and 13C/15N/2H-labeled peptides; chemical synthesis would be very expensive. In our lab, we have experimented with cloning the pY1021 peptide into the pGEX system as a fusion protein with GST followed by cleavage with thrombin and purification of the subsequent peptide by HPLC. In addition, in vitro phosphorylation of the GST fusion with purified EGF receptor could be performed. Recent work in our lab in synthesizing an unlabeled Pro-rich peptide as a GST fusion gave a yield of approximately 1 mg of peptide per liter of M9 medium (Voula Kanelis, personal communication); other cloning systems may give higher yields.

F. Larger Fragments of PLC-γ

As stated previously, the PLC isozymes are multidomain proteins in which all but one of the recognized domains have been solved in isolation, including the PLC-δ PH domain (Ferguson et al., 1995), the PLCC SH2 domain and the PLC-γ SH3 domain (Kohda et al., 1993), the PLC-δ catalytic domains (Essen et al., 1996) or can be modeled based on the structure of homologous domains (e.g. the PLCN SH2 domain) (see Chapter 1). The only domain whose structure has not been solved in some form is a 234-amino acid region of
PLC-β thought to bind to the Gq class of α-subunits of trimeric G proteins. However, this structural information does not indicate whether the domains are independent or interact in some manner, and whether the domains interact upon ligand binding or phosphorylation. Phosphorylation is of particular interest because phosphorylation of PLC-γ in the region between the C-terminal SH2 domain and SH3 domain increases enzymatic activity (Kim et al., 1991). For that reason, it would be of interest to study pieces of these enzymes which contain 2 or more domains. For example, in our lab, fragments of PLC-γ containing both SH2 domains (22), both SH2 domains and the SH3 domain including the intervening sites of phosphorylation (22Y3) and a fragment containing the split PH domain (P22Y3H) have been subcloned and some expressed in order to study them by NMR. In addition, attempts have been made to phosphorylate the 22Y3 and P22Y3H constructs in vitro using purified EGF receptor. All such multidomain proteins are >200 amino acids, so that triple labeling ($^2$H, $^{13}$C and $^{15}$N) would be required. High levels of deuteration (up to 90%) would be necessary to decrease the pathways for relaxation and therefore decrease linewidth; however, even with extensive deuteration, a limit on the molecular weight still exists due to line broadening. One could estimate a size limit for solution NMR even with extensive deuteration to be 70K (Dr. Kevin Gardner, personal communication); beyond this limit, other techniques for structure determination would have to be used. This size is still smaller than the entire PLC-δ molecule (approximately 90K), which is the smallest of the PLC isozymes. Thus in order to study the structure of the complete PLC protein, one would have to resort to other techniques such as X-ray crystallography or optical techniques such as electron microscopy or atomic force microscopy which at present have poorer resolution. This in fact may be done in the near future as preliminary data on crystals of full-length PLC-δ have been collected (Essen et al., 1996); in addition, a group has recently reported the cloning and purification of active full-length PLC-γ using the baculovirus expression system (Koblan et al., 1995).
G. Summary

In conclusion, in the previous chapters I have described the structure and electrostatic properties of the PLCC SH2 domain/pY1021 complex. In this chapter I have outlined a number of possible directions for future work, which involves further studies of this complex, other PLCC SH2 domain/phosphopeptide complexes and larger fragments of the PLC-γ molecule. With respect to the PLCC SH2 domain/pY1021 complex, chemical shift analysis, mutational analysis and structure calculations involving electrostatic energy terms may be useful for further description of the pTyr binding site. Mutational analysis of the BG loop of this SH2 domain may be of interest in order to understand the specificity of binding of this SH2 domain relative to the PLCN SH2 domain. In addition, structure calculations of this structure using time- or ensemble-averaged NOE restraints may be helpful in understanding the unusual dynamics observed for the methyl groups of Leu 69 and Leu 77. Finally, dynamics studies of the phosphopeptide may be of interest, as this may show a wide range of dynamic behaviors. With respect to the structure of other PLCC SH2 domain/phosphopeptide complexes, these may be of interest to help fully explore the binding surface of this SH2 domain, which may be useful in order to design inhibitors. Finally, the PLCC SH2 domain is only a small portion of the PLC-γ protein. In order to fully understand the role it plays in activation of PLC-γ enzymatic activity, larger fragments of the PLC-γ protein must be studied.
Appendix 1: Lineshape Calculations with Mathematica

Equations for the calculation of lineshape with exchange have been derived from the Bloch equations. The total magnetization when averaged over the proton exchange, taken from equation 3 of Gutowsky and Saika (1956), is as follows:

\[ M = \frac{i \omega_1 M_0 [(\tau_A + \tau_B) + \tau_A \tau_B (\alpha_A p_B + \alpha_B p_A)]}{(1 + \alpha_A \tau_A)(1 + \alpha_B \tau_B) - 1} \]  \hspace{1cm} (1)

where \( \omega_1 \) is the frequency of the applied magnetic field, \( M_0 \) is the static nuclear magnetization at thermal equilibrium, and \( p_A \) and \( p_B \) represent the normalized populations in environments A and B (i.e. free and bound). In addition, \( \alpha_A = (1/T_2) - i (\nu_A - \nu) \) and \( \alpha_B = (1/T_2) - i (\nu_B - \nu) \) where \( T_2 \) is the intrinsic relaxation rate of the nucleus in the transverse plane and \( \nu_A \) and \( \nu_B \) are angular frequencies of nuclei A and B in rad/sec; in order to convert the frequencies in Hz/sec obtained in NMR spectra, one must multiply by \( 2 \times \pi \). The average lifetimes in sites A and B are given by \( \tau_A \) and \( \tau_B \), where \( \tau_A = \tau/\rho_B \), \( \tau_B = \tau/\rho_A \) and \( \tau = \tau_A \tau_B/(\tau_A + \tau_B) \). The equation is encoded in Mathematica® script as shown below: note that the numerator and denominator of equation (1) is encoded separately by the functions "num[tau_]" and "dem[tau_]", and the overall lineshape is given by dividing the numerator and denominator (lineshape[tau_] = num[tau]/dem[tau]). Also note that the lineshape function and many intermediate functions are of the form "function[arg_] :=", such that tau is the argument to the function (see Gaylord et al., 1993). This will enable us to obtain the value for tau (the lifetime in each state) more easily, as will be seen later.

\[
taua[tau_] := tau/pb \\
taub[tau_] := tau/pa \\
delfreqA = 2 Pi (freqA - freq); \\
delfreqB = 2 Pi (freqB - freq); \\
alphaa = 1/T2 - I (delfreqA); \\
alphab = 1/T2 -I (delfreqB); \\
dem[tau_] := (1 + alphaa taua[tau]) (1 + alphab taub[tau]) -1 // N
\]
Values for the parameters for the lineshape can then be input. Below are given parameters for the lineshape in the $^1$H dimension for Ala 46 in an HSQC spectrum of the PLCC SH2 domain/pY992 complex such that the ratio of the peptide-bound and free SH2 domain are 0.66:0.34. The apparent $T_2$ value was obtained from the linewidth at half height for the free Ala 46 $^1$H resonance by the equation $v_{1/2} = 1/(\pi \times T_2)$, where $v_{1/2}$ is the linewidth at half height. The difference in frequency between Ala 46 in the free and bound state has been shown to be 95 Hz (see Table 3, Chapter 2), so that for the sake of the simulation, the frequencies of the bound and free Ala 46 NH protons were given a value of 100 and 195 Hz respectively.

\[ pa = .66 \ ; \ pb = .34 \ ; \ freqA = 100 \ ; \ freqB = 195 \ ; \]
\[ T_2 = .0108 \ ; \ freq1 = 500; \]
\[ Shape[\tau_] := \text{lineshape[\tau]} \] // N // Simplify

The output is the following equation:

\[
(79.5775 (-0.159155 \ I \ \tau + (-162.7 - 14.7366 \ I) \ \tau^2 + 1. \ freq \ \tau^2)) / \\
((-2.3454 + 21.0562 \ I) \ \tau - 0.159155 \ I \ \text{freq} \ \tau + (19282.8 + 4347.29 \ I) \ \tau^2 + (-295. - 29.4731 \ I) \ \text{freq} \ \tau^2 + 1. \ \text{freq}^2 \ \tau^2 )
\]

In addition, one could plot the resulting lineshape, as shown below. The absorption lineshape is given as the imaginary component of the equation above. This can be plotted using a value of \(\tau\), in this case of 2 ms, using the following command:

\[
\text{Plot[Im[Shape[.002]],\{freq,0,300\}]}
\]

yielding the resulting curve:
At this point, we can plot the lineshape of a species undergoing two-site exchange using any input ratios for the two species, a $T_2$ value, a chemical shift difference between the two states and a value for the lifetime of the bound state. However, we would like to be able to extract linewidths from this equation. Therefore the compound function "createLinewidths[tau_]" was created.

```
createLinewidths[tau_] := {
  halfmax = N[Max[Table[Im[Shape[tau]], {freq, 60, 200, .5}], 5]]/2;
  junk = Take[Sort[Table[{Abs[Im[Shape[tau]]-halfmax], freq},
                         {freq, 60, 200, .5}], {1, 4}];
  linewidth = Abs[junk[[1, 2]]-junk[[2, 2]]]
}
Table[createLinewidths[tau], {tau,.0005,.005,.0005}]
```

At this point we can see the advantage of using tau as the argument all functions and compound functions upon which it depends, namely that using the command

```
Table[createLinewidths[tau]]
```
one can determine linewidths over a range of values of tau. The calculation of the linewidth consists of three separate functions. The first ("halfmax") determines 1/2 of the maximum height of the peak within the frequency range given. The next function ("junk") takes a list of points containing the intensity and frequency of absorption over the frequency range (of syntax \{intensity, freq\}), subtracts the intensity by the value of "halfmax", takes the absolute value and sorts the resulting points by the new intensity from lowest to highest. As a result, those points which are near the frequency of half the maximal intensity should have values at or near zero, and the first two points should have the lowest intensities and the frequencies on either side of the peak which represent half the maximal intensity. Subtraction of the frequencies of the first two points then should give the linewidth (as performed by the function "linewidth").

Following this compound function is the command:

\textbf{Table}[createLinewidths[tau], \{tau,.0005,.005,.0005\}]

This allows us to loop through "createLinewidths" and determine the linewidths using different values of tau, starting from 0.5 ms, ending at 5 ms and incrementing by 0.5 ms (for a total of 10 values). The output is shown below:

\textbf{Out[26]}=\{\{42.5\}, \{55.5\}, \{67.5\}, \{77.\}, \{82.5\}, \{85.\}, \{0.5\}, \{0.5\}, \{0.5\}\}

Note that at several of values of tau, the linewidths calculated are clearly too small (0.5 Hz). This is due to the splitting of the resonance into two lines separated by 95 Hz, yielding two resonances of 100 and 195 Hz. A linewidth then cannot be calculated per se over a frequency range of 60 and 200 Hz. Similarly, in Tables 1 and 2 (in the PLCC/pY1021 complex), a number of linewidths of 0.5 Hz are calculated at lower values of tau due to the formation of a single peak from the free and bound peaks. In this case a
linewidth cannot be calculated when the frequency range used is only suitable for the free or bound peak.
Appendix 2: Stereoassignment and Restraints about the Dihedral Angle $\chi_1$ for the pTyr Residue in the PLCC SH2 Domain/pY1021 Complex Using the Program Stereosearch

As stated previously, stereoassignment of methylene protons can lead to greater precision in NMR structures due to the reduced use of pseudoatom corrections and the determination of dihedral restraints during the stereoassignment process. The program Stereosearch (Nilges et al., 1990) was used to accomplish stereoassignment of $\beta$-methylene protons and determination of $\chi_1$ restraints for a number of residues of the SH2 domain in the PLCC SH2 domain/pY1021 complex. Briefly, this procedure employs a conformational search of $\phi$, $\psi$ and $\chi_1$ dihedral space based on input distance ranges and coupling constants. The program is set up for amino acids which have two H$\beta$ protons, so it is more difficult to use for amino acids such as Val and Ile which contain only one H$\beta$ proton. In the following, I discuss my results with respect to stereoassignment of pTyr. Note that this procedure is possible for stereoassignment of $\beta$-methylene protons of proline, but such a complete database search is not necessary because the geometry about the $\beta$-methylene protons is highly constrained by the proline ring.

Below is shown the information in the input file for the pTyr residue. The syntax is as follows; $J$ represents a coupling constant, while $D$ represent a distance: $N$, $A$, $B2$ and $B3$ represent the NH, H$\alpha$, H$\beta_2$ and H$\beta_3$ protons respectively. Distances are intra-residue, with the exception of those terminating with a + sign, which represents the distance involving one proton with the following sequential (i+1) proton. With respect to the NOE data shown below, a strong NOE between the H$\alpha$ and the (i+1) NH proton (see Figure 3) indicates that this residue is in a extended conformation. In addition, I previously stated that the H$\alpha$ proton was found to be closer to H$\beta_3$ than H$\beta_2$, so I stated that the H$\alpha$/H$\beta_3$ distance was quite short (<2.7 Å), the H$\alpha$/H$\beta_2$ distance was longer (<3.5 Å), and that the H$\alpha$/H$\beta_1$ distance was always less than the H$\alpha$/H$\beta_2$ distance by a value of 0.3 Å or
greater. With respect to coupling constant data, a $^{3}J_{HN}H_{\alpha}$ of $>7$ A was estimated due to the presence of NOEs indicative of an extended $\beta$-structure, although $^{3}J_{HN}H_{\alpha}$ coupling constants were not measured for the phosphopeptide. For the $^{3}J_{H\alpha}H_{\beta}$ values, I estimated one to be large ($>9$ Hz) based on the $\sim12$ Hz coupling measured by the method of (Kim & Prestegard, 1989), while another was estimated to be small based on the lack of an $H_{\alpha}/H_{\beta}$ crosspeak in the filtered COSY and PE-COSY experiments.

The output from Stereosearch is also shown below. From these results, one observes that both assignments of the $H_{\beta}$ protons are possible, the first assignment ($\delta\beta_{3} = 3.19$ ppm, $\delta\beta_{2} = 2.83$ ppm) yielding conformers about $\chi_{1} = -60$, the second assignment ($\delta\beta_{3} = 2.83$ ppm, $\delta\beta_{2} = 3.19$ ppm) yielding conformers about $\chi_{1} = 180$. However, the number of conformers found for the first assignment exceeded those of the second assignment by ~20-fold, so the first assignment more likely. In addition, the second assignment covers a very narrow range of $\varphi$ values ($\varphi = 165-185$) which only spans one edge of the $\beta$ region in the Ramachandran plot.

Part a: Stereosearch Input for the Residue pTyr:

```
Residue PTY 4
JNA >  8.0  1.0
JAB2 > 10.0 1.0
JAB3 <  3.0  1.0
DAB3 <  2.6  0.1
DAB2 <  3.4  0.3
DNB3 <  4.0  0.5
DNB2 <  3.5  0.2
DB3N+ <  3.5  0.2
DAN+ <  2.6  0.1
DAN <  3.5  0.2
DAB3 < DAB2 -0.3 ! dab3 is smaller than dab2 by more than .3 A
DNB3 > DNB2 -0.3 ! dnb3 is larger than dnb2 by more than .3 A
DAN+ < DNN+ -0.3
END
```

Part b: Stereosearch Output for the Residue pTyr

Results for residue PTY  4

# conf in database: 37649
# conf. input assignments: 245
# conf. swapped assignments: 12
Both assignments possible

Allowed region for PHI:
  input:  -155.0 ... -85.0
  swapped: -155.0 ... -95.0
  both:    -155.0 ... -95.0

Allowed region for PSI:
  input:   105.0 ... -165.0
  swapped: 165.0 ... 185.0
  both:    165.0 ... 185.0

Allowed region for CHI1:
  input:   -95.0 ... -55.0
  swapped: 175.0 ... 185.0
  both:    175.0 ... 185.0
Appendix 3: Measurement of J from Phase-Sensitive COSY Spectra

As stated previously, the measurement of a number of scalar couplings in the pY1021 phosphopeptide was possible from the F1/F2-filtered PE-COSY experiment. However, usually the passive coupling is measured rather than the active coupling for a given crosspeak since it is more accurate. The reason for this is that the active coupling appears in COSY and PE-COSY experiments with an antiphase lineshape, and measurement of the peak separation between the two peaks in the antiphase lineshape can be strongly overestimated due to large resonance linewidths. Note that linewidths of proteins can be quite large and can contribute a large error. I have measured separations between antiphase components of 15-16 Hz for the pY1021 phosphopeptide bound to the PLCC SH2 domain in the double-filtered COSY and PE-COSY experiments, however no $^3J$-coupling has been observed in proteins to be >12 Hz. Measurement of passive couplings in PE-COSY experiments allowed us to obtain a number of coupling constants. However two situations existed where this was not possible. First, in Ile residues, since there is only one $H\beta$ proton, there is no passive coupling to measure in the $H\alpha/H\beta$ crosspeak in the PE-COSY. Second, in the case of residues pTyr and Pro +3, one of the couplings ($^3J_{H\alpha H\beta 2}$) is very small so that one of the crosspeaks ($H\alpha/H\beta 2$) is missing. Thus $^3J_{H\alpha H\beta 2}$ can be measured from the $H\alpha/H\beta 3$ crosspeak but $^3J_{H\alpha H\beta 3}$ cannot be measured from the $H\alpha/H\beta 2$ crosspeak due to its absence.

A method for obtaining $^3J$ values from antiphase doublets was derived by Kim & Prestegard (1989) and coupling constants measured in this way were used in NMR structure refinement of the acyl carrier protein (Kim & Prestegard, 1990). This method involves measurement of peak-to-peak separations of absorptive and dispersive antiphase doublets, the appearance which is shown in Figure 1, and use of the formula shown
**Figure 1 -- Measurement of J from Phase Sensitive COSY Spectra:** The following shows the lineshape and measurement of peak to peak separations of absorptive (on left) and dispersive (on right) antiphase doublets using simulated data with a coupling constant of 8 Hz and linewidths of 0.3, 3, 7 and 16 Hz. The following was taken from Kim & Prestegard (1989).
below,

\[ J^6 - v_a^2 J^4 + \left( -\frac{9}{4} v_a^4 + \frac{3}{2} v_a^2 v_d^2 + \frac{3}{4} v_d^4 \right) J^2 + \frac{81}{64} v_a^6 - \frac{9}{16} v_a^4 v_d^2 - \frac{21}{32} v_a^2 v_d^4 - \frac{1}{16} v_d^6 + \frac{v_d^8}{64 v_a^2} = 0 \]

where \( J \) represents the coupling constant to be measured and \( v_a \) and \( v_d \) represent the measured peak-to-peak separations for the absorptive and dispersive antiphase lineshapes respectively. This formula was encoded by an in-house C program ("j_calc.c"), whose source code is shown below. The left hand side of the equation above is encoded by the variable "funct". However, the form of this equation represents a function which must be minimized through iteration using a starting value of \( J \), in this case given a value of "coup = 20" or 20 Hz. The method used for minimization was Newton's method (Skeel & Keiper, 1993), and in using this method, the derivative of "funct" must be determined. This is done by the variable "d FUNCT". Note that the function is iterated to a minimum value which is >0 (in this case, to a value of 2.0). Initial values tend to be very large \((10^4-10^6)\) prior to minimization. Experimental values for \( v_a \) and \( v_d \) are input into the variables \( v_{\text{abs}} \) and \( v_{\text{disp}} \), and the program is compiled and run.

"j_calc.c"

/* the following is meant to calculate \( J \) from a phase-sensitive COSY */
/* Equation 5 is encoded by the variable "funct". To get the correct */
/* value of \( J \), funct = 0, therefore I also have defined the derivative of */
/* funct -- "d funct", and used Newton's method to solve the problem */
/* x(i) = x(i-1) - f(x-1)/f(x-1) */

#include <stdio.h>
#include <math.h>

main()
{
    float v_abs, v_disp, v_abs_2, v_disp_2, coup, coup_2, funct, d_funct;
    float v_abs_4, v_disp_4, v_abs_6, v_disp_6, v_disp_8, coup_4, coup_6;
    float junk1, junk2, junk3, junk4;
    v_abs = 13.4;
v disp = 35.0;
v abs_2 = v abs*v abs;
v abs_4 = v abs_2*v abs_2;
v abs_6 = v abs_2*v abs_4;
v disp_2 = v disp * v disp;
v disp_4 = v disp_2 * v disp_2;
v disp_6 = v disp_4 * v disp_2;
v disp_8 = v disp_4 * v disp_4;
coup = 20.000;

funct = 3.0;
d funct = 100000;

while(funct > 2.0 ){
coup = coup - funct/d funct;
coup_2 = coup * coup;
coup_4 = coup_2 * coup_2;
coup_6 = coup_4 * coup_2;
junk1 = 0 - 21*v abs_2*v disp_4/32 - v abs_4*v disp_2 * 9/16 +
v abs_6*81/64;
junk2 = v disp_8 / v abs_2/64 - v disp_6/16;
junk3 = junk1 + junk2;
junk4 = 0 - 9*v abs_4/4 + 1.5*v abs_2*v disp_2 + 0.75*v disp_4;

funct = coup_6 - v disp_2*coup_4 + junk4*coup_2 + junk3;
d funct = 6*coup_4*coup_2 - 4*v disp_2*coup_2*coup_2 + 2*junk4*coup;
printf("%5.2f, %15.6f, %15.6f\n", coup, funct, d funct);
}
return 0;

Below are a number of couplings which I measured using this method. The double-
filtered COSY was processed using a 12° phase-shifted sine-bell squared function, and
processed to a resolution of 1.46 and 2.92 Hz per point in t2 and t1, respectively, before
measurement. Measurement was along t2 for all resonances except Ile +1, due to
interference by the water resonance in this dimension. Most of the measured couplings
have been discussed in this chapter. The two couplings shown for Pro +7 measured in this
manner are nearly identical to those measured in the PE-COSY (see Figure 10), and show
that this method was valid for measurement of coupling constants in this system.
<table>
<thead>
<tr>
<th></th>
<th>in-phase separation (Hz)</th>
<th>antiphase separation (Hz)</th>
<th>calculated $^3$J (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro +3 αβ1</td>
<td>14.6, 13.2</td>
<td>30.8</td>
<td>12.3</td>
</tr>
<tr>
<td>Pro +7 αβ1</td>
<td>8.8, 10.3</td>
<td>19.0</td>
<td>8.9</td>
</tr>
<tr>
<td>Pro +7 αβ2</td>
<td>6.5</td>
<td>14.6</td>
<td>5.7</td>
</tr>
<tr>
<td>pTyr αβ</td>
<td>14.6</td>
<td>33.7</td>
<td>12.5</td>
</tr>
<tr>
<td>Ile +6 αβ</td>
<td>14.6</td>
<td>23.5</td>
<td>14.6</td>
</tr>
</tbody>
</table>
Appendix 4: Structural and Sequence Alignment of the BG Loop

The BG loop is highly important in phosphopeptide binding but is of variable length from protein to protein. Structural studies indicate that it is somewhat rigid both in the free state and following peptide binding. In an $^{15}$N relaxation study of the PLCC SH2 domain, order parameters for the BG loop were found to be between 0.80 and 0.87, and varied little and in some cases even decreased upon phosphopeptide binding (Farrow et al., 1994). Energy for stabilization of the BG loop may be due at least partly to burial of hydrophobic residues, and frequently there is conservation of residues in the hydrophobic core of proteins in order to maintain packing interactions. The BG loop of the PLCC SH2 domain undergoes little change upon phosphopeptide binding based both on CD results (Gerry Gish, unpublished data) and NOE patterns in $^{15}$N-NOESY spectra. In addition, the structures of a number of SH2 domains has been solved both free and in solution and show little change in the conformation of the BG loop (Eck et al., 1993; Lee et al., 1994; Waksman et al., 1993). One exception to this is the p85 N-terminal SH2 domain, in which differences in CD spectra were observed upon phosphopeptide binding (Panayotou et al., 1993; Shoelson et al., 1993): recent structural studies have suggested that this change is due to movement of a Tyr ring in the BG loop (Nolte et al., 1996). Because of the general rigidity observed for the BG loop, I examined the crystal and solution structures of a number of SH2 domains and SH2/phosphopeptide complexes which are publicly available. These include the src SH2 domain/YEEI complex of Waksman et al. (1993), the lck SH2 domain/YEEI complex of Eck et al. (1993), the syp N-terminal SH2 domain/pY1009 complex (Lee et al., 1994), the syk C-terminal SH2 domain/IgE phosphopeptide complex (Narula et al., 1995), the PLCC SH2 domain/pY1021 complex and the free SH2 domains of SHC (Mikol et al., 1995) GRB-2 (from the structure of the whole GRB-2 molecule (Maignan et al., 1995)) and the N-terminal SH2 domain of p85 (Booker et al., 1992).
In all SH2 domains studied, two or three hydrophobic residues in the BG loop were found to have side chains buried in the hydrophobic core. The first residue was found to be a Leu residue and will be referred to as Leu BGA; it corresponds to Leu 89 or Leu BG2 in the PLCC SH2. The methyl groups of this residue was found to be close in space to the side chain of the +3 residue of the phosphopeptide, as well as with βC6, βD5 when it corresponds to a bulky hydrophobic amino acid, αB5 and αB8. This Leu residue occupies roughly the same position in the protein and makes the same hydrophobic interactions if it is in position BG2 or BG4: when this residue is the second residue in the BG loop, the N-terminus of the BG loop is in an extended conformation, but if this residue corresponds to BG4, the N-terminus of the BG loop is more irregular in structure.

Another highly conserved residue among SH2 domains (BGC) is also a highly-conserved Leu corresponding to the second last residue of the BG loop (Leu 95 in PLCC SH2). This residue is close in space to the aromatic ring of Phe βB2 (of the FLVR sequence), βC6, a pair of residues in the αB helix, and the previously described BGA residue.

Two residues N-terminal to the BGC residue is a residue which is generally hydrophobic in nature which I refer to as BGB (corresponding to Met 95 in the PLCC SH2). The BG loop adopts an extended strand-like conformation between BGB and BGC in all SH2 structures studied. In SH2 domains with smaller BG loops, the side chain of BGB tends to be only partially buried, but in SH2 domains with longer BG loops (GRB-2, Syp and SHC), this residue is completely buried, making hydrophobic contact with BGA and BGC, as well as residues in the C-terminus of the βC strand.

A sequence alignment of BG loops of SH2 domains can be made based on alignment of residues BGA, BGB and BGC, and such an alignment is shown in Figure 1 for SH2 domains whose structures are solved and coordinates are publicly available, an SH2 domain whose structure is known but whose coordinates are not yet publicly available (the p85 C-terminal SH2 domain (Breeze et al., 1996; Nolte et al., 1996)), and an SH2 domain
Figure 1 -- Sequence Alignment of the BG Loop of a Number of SH2 Domains: To better align these sequences, residues in the final turn of helix $\alpha B$ and in strand $\beta G$ are shown. Alignment generated using the program SeqVu (1995), plotting sequence similarity at a homology level of 70% using the Milik Delta-G HLX Scale.
whose structure has not yet been determined (the PLCN SH2). In the case of SH2 domains with small BG loops, the position of these 3 residues nearly completely determines the fold of this loop: the conserved residues correspond for example in src to Leu BG4, His BG6 and Leu BG8, such that Cys BG5 must be placed in the middle of the turn between BGA and BGB. In SH2 domains with larger BG loops, the position of these residues restricts the conformation of the BG loop to some extent, but the positions of a number of residues are still unknown. For example, in the PLCN SH2 domain, six residues lie between the conserved amino acids in which lies a turn. The PLCN SH2 domain binds hydrophobic residues in the +4 position (see Table 3, Chapter 4): many of the residues in the region between BGA and BGB are generally hydrophilic, however the presence of a Phe at position BG7 may be significant.
References


Cleghon, V, U Gayko, TD Copeland, LA Perkins, N Perrimon, DK Morrison. Drosophila terminal structure development is regulated by the compensatory activities of positive and negative phosphotyrosine signaling sites on the Torso RTK. Genes & Dev. 10:566-577.


320


Larose, L, G Gish, S Shoelson, T Pawson. 1993. Identification of residues in the β platelet-derived growth factor receptor that confer specificity for binding to phospholipase C-γ1. Oncogene 8:2493-2499.


Molecular Simulations Inc. 1993. QUANTA, version 4.0. Waltham, Massachusetts.


demonstration of an intramolecular SH2-phosphotyrosine interaction in the Crk protein.

Rotin, D, D Bar-Sagi, H O'Brodovich, J Merilainen, VP Lehto, CM Cnaessa, BC Rossier,
GP Downey. 1994. An SH3 binding region in the epithelial Na⁺ channel (αENac)
mediates its localization at the apical membrane. EMBO J 13:4440-4450.

Rotin, D, B Margolis, M Mohammadi, RJ Daly, G Daum, N Li, EH Fischer, WH
dehphosphorylation of the EGF receptor: identification of Tyr992 as the high-affinity
binding site for SH2 domains of phospholipase Cγ. EMBO J 11:559-567.

Rozakis-Adcock, J McGlade, G Mbamalu, G Pelicci, R Daly, W Li, A Batzer, S Thomas,
Grb2/Sem5 SH2-containing proteins is implicated in activation of the Ras pathway by

Cambridge, Massachussetts.

Cytoplasmic Protein Kinases Modifies the Kinase Function and Transforming Activity of


