Rockefeller University
Digital Commons @ RU

Student Theses and Dissertations

1997

Protein/Peptide Recognition Modules in Cellular Signaling and HIV Pathogenesis

Chi-Hon Lee

Follow this and additional works at: https://digitalcommons.rockefeller.edu/student_theses_and_dissertations

Part of the Life Sciences Commons

Recommended Citation
Lee, Chi-Hon, "Protein/Peptide Recognition Modules in Cellular Signaling and HIV Pathogenesis" (1997). *Student Theses and Dissertations*. 451.
https://digitalcommons.rockefeller.edu/student_theses_and_dissertations/451

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact nilovao@rockefeller.edu.

LD4711.6
L4771
c.1
RES

THE LIBRARY

LD 4711.6 L4771 1997 c.1 RES
Lee, Chi-Hon.
Protein/peptide recognition
modules in cellular

Rockefeller University Library
1230 York Avenue
New York, NY 10021-6399

Protein/Peptide Recognition Modules in Cellular Signaling and HIV Pathogenesis

A thesis presented to the faculty of
The Rockefeller University
in partial fulfillment of the requirements of
the degree of Doctor of Philosophy

by

Chi-Hon Lee

April, 1997

For My Parents

Acknowledgments

I would like to thank my advisor John Kuriyan in whose laboratory the work described in this dissertation was conducted. As an advisor, John has guided me closely, being patient with my inexperience, and shaping my thinking process. I am indebted to him for his support, insight and creativity, and for the freedom to let my work take me wherever it led.

I would like to thank all the present and past members of his laboratory for interesting discussions and advice. In particular I would like to thank Jonathan Goldberg and Jacqui Gulbis for their invaluable advice, for their patience in teaching me how to run several crystallography programs. Thanks to David Wilson for much needed advice and for invaluable friendship. Thanks to T.S.R. Krishna and Dorothea Kominos for their help in the early phase of the crystallography work. Thanks to Xiaomin Chen, Frank Sicheri, and Brian Guenther for helpful advice. Also thanks to David Jeruzalmi and Elena Conti for showing me how to work with CCP4 programs. Thanks to Ramakoti Suresh and Jeff Bonanno for their help with data collection, and to Steven Jacques and Huguette Viguet and for their help with the project. Special thanks to Ronnie Gulli and Anne Roche for making my life easier in Rockefeller.

Many thanks to Stephen Burley for being so supportive over the past several years and for many interesting discussions. I also

thank him for advising me to improve my language skills. Thanks also to the members of the Burley laboratory who provided me with invaluable advice, Xiaoling Xie, Adrian Ferre-D'Amare, Elaine Halay and Joe Kim. In particular, thanks to Xiaoling for teaching me many skills in crystallography.

I would also like to thank Kalle Saksela for his help and advice with the Nef project and faithful friendship. Thanks to the past member of Saksela lab, Benjamin Leung, for his help with Nef protein purification. Special thanks to David Cowburn for allowing me to work and to use many expensive equipment in his lab and for invaluable advice and help. I am most grateful to the members of the Cowburn lab for much helpful advice, particularly, Jie Zheng and Sean Cahill. Many thanks to Brian Chait and Urooj Mirza for help with masspectrometry measurements.

Many thanks to the members of my thesis committee, Stephen Burley, Ulrike Gaul, Hidesaburo Hanafusa and Tony Pawson for careful reading of this document. Thanks specially to Ulrike Gaul for many interesting discussion in cell signaling and for supporting me to move in the field of fly genetics.

Thanks specially to Peter Model for his guidance during my training at Rockefeller and for introducing me to Marcelo Magnasco. I am forever indebted to Marcelo Magnasco for teaching me the language of mathematics and physics. I thanks Andrej Sali for help

and advice with molecular modeling work. Thanks to Roberto Sánchez for teaching me how to run Modeler. Thanks to Roderick MacKinnon and Seth Darst for much needed advice and many interesting discussions.

I would like to extend my thanks to my previous advisors Joseph Schlessinger and Edward Skolnik who brought me into the filed of signal transduction and continue to support and advise me in the past five years. I thank Joseph Schlessinger for his support and insight into the SH2 and PTB projects. Many thanks to the past and present members of Schlessinger lab. In particular, thanks to Zhongtao Zhang for providing me the X11 crystals and for his help with this project, to Valsan Mandiyan for help and advice in the binding study, to Mark Lemmon for the ITC experiments and for many memorable discussions, and to Benjamin Margolis and Ivan Dikic for many interesting discussions.

Table of Contents:

Acknowledgments ----- iii
Table of Contents ----- vi
Table of Figures ----- x
Table of Tables ----- xiv
Abstract ----- xv

1. Introductions

1.1 Signaling by Protein/Peptide Recognition Modules----- 1
1.2 The SH2 Domains ----- 4
1.3 The PTB Domains ----- 9
1.4 The SH3 Domains ----- 14
1.5 The HIV-1 Nef Protein ----- 21

2. Materials and Methods

2.1 Structure Determination of the SH-PTP2 SH2 Domain ----- 23

2.1.1 Protein Expression and Purification. ----- 23
2.1.2 Crystallization and Data Collection ----- 24
2.1.3 Molecular Replacement and Model Refinement ----- 27

2.2 Structure Determination of the X-11 PTB domain ----- 33

2.2.1 Space Group Determination and Data Collection ----- 33
2.2.2 Heavy Atom Derivatives ----- 34
2.2.3 Phasing by MAD/MIR ----- 35
2.2.4 Model Building and Refinement ----- 36

2.3 Characterizing the Interaction between the HIV-1 Nef Protein and Src Family SH3 Domains ----- 42

2.3.1 Divergence Analysis and Molecular Modeling ----- 42

2.3.1.1 Sequence Comparison ----- 42

2.3.1.2 Molecular Modeling ----- 45

2.3.2 Biochemical and Biophysical Characterization ----- 47

2.3.2.1 Protein Expression and Purification---- 47

2.3.2.2 Surface Plasmon Resonance Analysis--- 50

2.3.2.3 Isothermal Titration Calorimetry ----- 51

2.3.2.4 Tryptophan Fluorescence Spectroscopy ----- 52

2.3.2.5 Circular Dichroism (CD) Spectroscopy -- 54

2.3.2.6 Co-precipitation Assay ----- 54

2.4 Structure Determination of HIV-1 Nef - FYN(I) SH3 complex ----- 56

2.4.1 Identification of Domain Structure of Nef and Optimization of Crystal Quality ----- 56

2.4.1.1 Protein Expression and Purification---- 60

2.4.2 Crystallographic Methods ----- 62

2.4.2.1 Data Collection ----- 62

2.4.2.2 Heavy Atom Derivatives ----- 62

2.4.2.3 MAD Phasing and Density Modification ----- 65

2.4.2.4 Model Building and Refinement ----- 66

3. Results ----- 71

3.1 Phosphopeptide Recognition by the SH-PTP2 SH2 Domain ----- 71

3.1.1 General Architecture ----- 71

3.1.2 Recognition of Phosphotyrosine ----- 75

3.1.3 Sequence Specific Recognition ----- 83
3.1.4 Peptide Binding Affinity and Specificity ----- 92
3.1.5 Summary ----- 96

3.2 Crystal Structure of the X11 PTB Domain Complexed with the APP Peptide ----- 97
3.2.1 The Binding of the X11 PTB Domain to the APP Peptide by Biosensor Analyses ----- 97
3.2.2 Overall Structure and Peptide Binding Mode -- 99
3.2.3 Interactions with NPxY Motif ----- 103
3.2.4 The C-terminal Specificity ----- 108
3.2.5 The N-terminal specificity ----- 109
3.2.6 Summary ----- 113

3.3 Biochemical Characterization of HIV-1 Nef - SH3 Domain Interaction ----- 114
3.3.1 High-Affinity Binding of Nef to Hck-SH3 Revealed by SPR Analysis ----- 114
3.3.2 Sequence Comparison of SH3 Domain and Molecular Modeling ----- 120
3.3.3 RT-loop Determines the Differential Affinity of Hck and Fyn SH3 Domains ----- 125
3.3.4 Nef PxxP Peptide has Low Affinity and Specificity ----- 130
3.3.5 Nef-PxxP Motif-mediated SH3-binding Examined by a Co-precipitation Assay ----- 132
3.3.6 Summary ----- 135

3.4 Crystal Structure of HIV-1 Nef-FYN(I) SH3 Complex -- 136

3.4.1 Overall Structure of Nef ----- 137
3.4.2 The Interaction between Nef PxxP Motif and the SH3 Domain ----- 142
3.4.3 Tertiary Interactions between Nef and SH3----- 147

3.4.4 Flexibility of the Nef:SH3 Interface ----- 152
3.4.5 Sequence Variation in Nef and its Implication in the Nef-Host Proteins Interactions ----- 153
3.4.6 Summary ----- 159

4. Discussions ----- 160

4.1 The Binding Specificity of the SH2 Domains ----- 160
4.1.1 The SH2 Specificity and its Prediction ----- 160
4.1.2 The Discrepancy between Structural and Binding Data ----- 163
4.1.3 Does the *in vitro* Specificity Observed for Peptide Account for the *in vivo* Specificity ---- 165

4.2 Peptide Recognition by the X-11 PTB Domain ----- 167
4.2.1 The Comparison with the SH2 Domains ----- 167
4.2.2 Biological Implications of the X11-APP Peptide Complex Structure ----- 171

4.3 The SH3-mediated Interactions ----- 173
4.3.1 Interactions with the PP-II Helix: the SH3 and WW Domains ----- 173
4.3.2 The Affinity and Specificity of the SH3-mediated Interactions ----- 174

4.4 The Biological Functions of Nef: the Implication of the Nef Structure ----- 179

Appendix A. Specificity, Selectivity, and Affinity ----- 185

Appendix B. Molecular Replacement ----- 190

References ----- 196

List of Figures

Figure 1.2.1 Schematic diagram of the SH2 structure of the Src-Y*EEI complex. ----- 8

Figure 1.3.1 The solution structure of the Shc PTB domain. ----- 13

Figure 1.4.1 The SH3-PxxP interaction, for the GRB2/Sem-5 SH3 domain complexed with a Sos peptide bound in the minus orientation. ----- 17

Figure 1.4.2 Alignment of the PxxP motif of SH3 ligands. ----- 18

Figure 1.4.3 Two types of packing geometries, with distinct preferences for proline vs. non-proline residues. ----- 19

Figure 2.1.3.1 Ramachandran plot of the 2.1 Å refined SH-PTP2/PDGFR-1009 model. ----- 31

Figure 2.1.3.2 Electron density maps for (a) the SH-PTP2/ PDGFR-1009 complex, and (b) the SH-PTP2/IRS1-895 complex. ----- 32

Figure 2.2.3.1 Ramachandran plot of the 2.3 Å refined X11/APP peptide model. ----- 40

Figure 2.2.3.2 Experimental electron density map of X11/APP peptide complex derived from MAD analysis. ----- 41

Figure 2.3.1.1.1 Conservation analysis of the SH3 domains. ----- 44

Figure 2.3.1.2.1 Surface representation of the conservation score and the electrostatic potential of Hck/Nef peptide model. ----- 46

Figure 2.3.2.4.1 Binding of Nef PxxP peptide to Hck SH3 domains. ----- 53

Figure 2.4.1.1 Nef constructs for crystallization trials. ----- 59

Figure 2.4.2.2.1 (a) Isomorphic difference Patterson map (b) Anomalous difference Patterson map. ----- 63

Figure 2.4.2.3.1 Electron density maps in a region of the hydrophobic core of Nef_{core}. ----- 64

Figure 2.4.2.4.1 Ramachandran plot of the 2.5 Å refined Nef_{core}/FYN(I) model. ----- 70

Figure 3.1.1.1 Schematic diagram of the SH-PTP2 SH2 domain in complex with a peptide from the insulin receptor substrate-1 (IRS-1). ----- 73

Figure 3.1.1.2 Alignment of SH2 sequences and definition of the residues notation. ----- 74

Figure 3.1.2.1 Stereo views of the phosphotyrosine binding sites of (a) SH-PTP2 and (b) Src. ----- 79

Figure 3.1.2.2 Schematic diagram showing interactions between the phosphotyrosine and the SH2 domain, for SH-PTP2 (a) and for Src (b) ----- 80

Figure 3.1.3.1 Comparison of the SH-PTP2 and Src SH2 domains. ----- 84

Figure 3.1.3.2 The conserved water molecule that bridges the peptide and the SH2 domain in the SH-PTP2, Src and Lck structures. ----- 84

Figure 3.1.3.3 Stereo view of the +1 and +3 peptide binding sites of (a) SH-PTP2/PDGFR-1009 complex and (b) Src/YEEI complex. ----- 87

Figure 3.1.3.4 The molecular surface of the SH-PTP2/PDGFR-1009 complex. ----- 87

Figure 3.1.3.5 Stereo view of the +3 peptide binding site of (a) SH-PTP2 PDGFR-1009 complex and (b) Src pYEEI complex. -- 88

Figure 3.1.3.6 Cutaway view of the SH-PTP2 SH2 domain, showing the interactions with the IRS1-895 peptide. ----- 88

Figure 3.2.2.1. Ribbon representation of the structure of the X11 PTB domain in complex with a 14 residue APP peptide. ----- 101

Figure 3.2.2.2. Stereodiagram of the interactions between the X11 PTB domain and the APP peptide. ----- 101

Figure 3.2.2.3 Sequence alignment of PTB and PH domains. ----- 102

Figure 3.2.3.1 Schematic depiction of contacts between the APP peptide and the X11 PTB domain. -----106

Figure 3.2.3.2 Comparison of phosphotyrosine/tyrosine binding sites of PTB domain and SH2 domain. ----- 107

Figure 3.2.5.1 Molecular surface of the X11 PTB-APP peptide complex. ----- 110

Figure 3.3.1.1. Determination of the equilibrium dissociation constant for the Nef/Hck-SH3 interaction using surface plasmon resonance (SPR). ----- 115

Figure. 3.3.1.2 Comparison of binding of wild-type Nef and the PxxP defective mutant Nef-PA1 to Hck-SH3. ----- 116

Fig. 3.3.2.1 (a) Alignment of a consensus Nef PxxP motif (RPxVPLR; amino acids 71-77 in HIV-1 NL4-3 Nef) with other peptides binding to different SH3 domains in "plus" or "minus" orientations. ----- 121

Fig. 3.3.2.1 (b) Alignment of SH3 domain amino acid sequences. ----- 122

Fig. 3.3.2.1 (c) Molecular model of interactions between Hck-SH3 and the Nef PxxP motif. ----- 123

Fig. 3.3.3.1 Differential binding of Nef to Hck, Fyn-I and Fyn SH3 domains. ----- 127

Fig. 3.3.5.1 Nef/SH3 co-precipitation assay. ----- 131

Figure 3.4.1.1. Structure of Nef:SH3 complex. (a, b). Stereo-diagrams of the polypeptide backbones of Nef_{core} and Fyn(R96I) SH3. ----- 139

Figure 3.4.1.2. Molecular Surface of Nef. ----- 140

Figure 3.4.1.3 (a,b) Architecture of Nef ----- 141

Figure 3.4.2.1 The footprint of Nef on the surface of the Fyn(R96I) SH3 domain. ----- 141

Figure 3.4.2.2 (a, b) Comparison of the Nef:SH3 and peptide:Crk-SH3 interfaces. ----- 143

Figure 3.4.3.1. (a, b) Tertiary interactions between Nef_{core} and Fyn(R96I) SH3 domain. ----- 148

Figure 3.4.4.1 Sequence alignment of Nef_{core}. ----- 154

Figure 3.4.4.2 The putative thioesterase binding site. ----- 158

Figure 4.4.1 Proposed model for the pleiotropic functions of Nef. ----- 182

List of Tables

Table 2.1.3.1 X-ray data collection and refinement statistics for the SH-PTP2 SH2 domain ----- 30

Table 2.2.3.1 Statistics for data collection of the X11/APP complexes ----- 37

Table 2.2.3.2 Statistics for refinement of the X11/APP complexes ----- 38

Table 2.4.2.4.1 Statistics for data collection and refinement for Nef/FYN(R96I) complex ----- 69

Table 3.1.4.1 Relative binding affinities of phosphopeptides to the SH-PTP2 SH2 domain. ----- 95

Table 3.2.1.1 Binding Affinity of the APP peptides for the X11 PTB domain ----- 98

Table 3.3.3.1. Summary of the quantitative Nef/SH3-binding data ----- 126

Abstract

The peptide-recognition domains play a key role in Eukaryotic signal transduction by mediating sequence-specific interactions with their protein/peptide ligands. In this thesis, I illustrate in molecular detail the recognition mechanisms utilized by three peptide-recognition domains: SH2 (Src homology 2) domains, PTB (phosphotyrosine binding) domains, and SH3 (Src homology 3) domains. The SH2 domains recognize tyrosyl-phosphorylated peptides using a binding mode that is conserved among the SH2 family members. In particular, the structure of the SH-PTP2 SH2 domain (described in this thesis) has revealed that sequence specificity can extend across the five residues immediately C-terminal to the phosphotyrosine. On the other hand, the PTB domains interact with peptide residues that are N-terminal to the phosphotyrosine in the sequence context, NPxY*. Although phosphorylation on tyrosine residue of ligand is absolutely required for interacting with the SH2 domains, some PTB domains (such as the X11 PTB domain described in this thesis) are able to mediate phosphorylation-independent interaction. The three-dimensional structure of the X11 PTB/peptide complex has revealed a general mechanism utilized for recognizing both phosphorylated and nonphosphorylated peptide ligands. In combination with *in vitro* binding study, the structural findings suggest that the PTB domains should be regarded as peptide recognition modules rather than phosphotyrosine binding module (as for the SH2 domain). While local sequences in the peptide ligands appear to determine the specificity in binding to the SH2 and PTB domains, SH3/peptide interactions are in general of low specificity and low affinity. As demonstrated by my biochemical study of

HIV-1 Nef/SH3 domain interaction, regions of intact SH3 binding proteins outside the PxxP motif can enhance both affinity and specificity by targeting the highly divergent RT-loop of the SH3 domains. The crystal structure of Nef in complex with a Src-family SH3 domain has provided the first example of high affinity SH3-protein interaction and revealed a novel mechanism by which SH3-mediated interaction achieves high affinity and specificity by the presentation of the PxxP motif on the tertiary structure of Nef. Furthermore, the structure of HIV-1 Nef suggests that Nef functions as an adaptor to subvert cellular signaling processes for the benefit of the invading virus.

1. Introduction:

1.1 Signaling by Protein/Peptide Recognition Modules

> Glendower: I can call spirits from the vasty deep.
> Hotspur: Why, so can I, or so can any man; But will they come when you do call for them?
>
> William Shakespeare, *King Henry IV*

The formation of specific protein-protein interactions is one of the key mechanisms for intracellular signaling. The biological consequences of these inter-molecular interactions include targeting signaling proteins to particular cellular locations, facilitating substrate-enzyme interactions, and modulating the enzymatic activity (reviewed in Pawson and Schlessinger, 1993; Cohen *et al.*, 1995). Thus, the signals encoded in the macromolecular format can be propagated further. In order to ensure the fidelity of the signaling processes that is required for coordinating various cellular activities, these interactions are controlled in a highly qualitative ——and most likely —— quantitative and temporal manner. Analogous to the quotation at the beginning of this introduction, the question depends not on whether any particular interaction can occur, but rather on whether only "specific" interactions can take place when needed. Understanding how specificity can be achieved and describing the molecular mechanism used has been the focus of my graduate study.

Many protein-protein interactions, particularly those responding to tyrosine-phosphorylation signals, are mediated by

small protein domains that recognize short stretches of peptide sequence (reviewed in Pawson, 1992; Cohen *et al.,* 1995; Kuriyan and Cowburn, 1997). Such peptide recognition domains are usually modular in both structural and functional respects. They are capable of folding correctly when removed from the parent proteins, and they can usually recognize their targets even when isolated. As a result, these domains can be integrated into larger molecules of which they are parts, while remaining functional with relatively little steric constraints imposed by the rest of the molecules. The "modular design" is both economic and efficient. Thus, it might account for the wide spread utilization of peptide recognition modules in various cell processes such as cellular signaling, sorting, and transcription (reviewed in Pawson, 1995).

The first peptide recognition modules to be identified were the SH2 and SH3 domains (Src homology 2 and 3 domains), so named because they share similarity with two separate non-catalytic regions of the Src family tyrosine kinases (Sadowski *et al.*, 1986; Moran, *et al.*, 1990). The SH2 and SH3 domains are well known for their critical roles in signal transduction by receptor tyrosine kinases (Pawson, 1992). Their functions involve the targeting of proteins to sites that contain phosphotyrosyl residues (for SH2) and proline-rich sequences (for SH3). More recently, the PTB/PI (phosphotyrosine binding/phosphotyrosine interacting) domains were identified by their ability to interact with phosphopeptides containing NPXY* motifs (Y*, phosphotyrosine) (Kavanaugh and Williams, 1994; Blaikie *et al.*, 1994). Although both SH2 and PTB domains recognize

phosphotyrosine, they share neither a structural folding nor a peptide recognition mode. Most strikingly, the architecture of the PTB domain closely resembles that of another signaling module, the plekstrin homology (PH) domain without any detectable sequence similarity (Zhou *et al.*, 1995; Eck *et al.*, 1996; Zhou *et al.*, 1996). Members of the PH domain family have rather divergent biological functions and often share low sequence homology. The structural resemblance between PH and PTB domain has led to tentatively including the PTB domains in the PH domain superfamily. Furthermore, the PDZ domains (which are known to recognize peptide sequence at the carboxyl-termini of membrane associated proteins) share a similar peptide binding mode with the PTB domains (Doyle *et al.*, 1996).

With the remarkable advances in molecular cloning and genome sequencing, the families of the peptide-recognition domains have been expanding rapidly. Some of them now include close to a hundred members. As each family member exhibits a different binding preference to its ligands, the emerging scheme for intracellular signaling networks that employ closely related members of these families has become exceedingly complicated. A *major* question arises as to how different domains achieve distinct specificities given their shared structural folds within their family members. Three examples are offered in this thesis to answer this question. These include: (i) the N-terminal SH2 domain of the tyrosine phosphatase SH-PTP2, (ii) the PTB domain of the X11 protein and finally (iii) the HIV Nef protein/Hck SH3 domain

complex. A separate introduction is given below for each domain. Since the main scheme concerns "specificity" which has been loosely used to describe a wide variety of phenomena, a definition for this term is provided in the Appendix A.

1.2 The SH2 Domains

The SH2 domain was first recognized as a phosphotyrosine-binding domain during the experiments designed to resolve the signaling pathways actuated by viral oncogenes (Sadowski, *et al.*, 1986; Matsuda, *et al.*, 1990; Mayer *et al.*, 1991). Subsequent experiments demonstrated that an individual SH2 domain targets specific regions of tyrosine-phosphorylated proteins such as particular sequences in the cytoplasmic regions of activated receptor tyrosine kinases (review in Pawson, 1992). The characteristic feature of the SH2 domain is thus its ability to bind to phosphorylated tyrosine residues with some measure of sequence specificity. This feature is the basis for the critical roles played by the SH2 domains in regulating the localization and the activity of the proteins of which they are parts. The SH2-mediated initiation of protein-protein interactions is one of the immediate consequences of the activation of cell surface receptors that signal by way of tyrosine kinases.

Following the above biochemical findings, structures of the SH2 domains were determined by NMR spectroscopy (for the Abl tyrosine kinase (Overduin *et al.*, 1992) and p85 subunit of PI3 kinase (Booker *et al.*, 1992) and by X-ray crystallography (for the Src tyrosine

kinase (Waksman *et al.*, 1992)). The crystal structure of Src SH2 domain, determined as complexes with low affinity phosphotyrosyl peptides, revealed a general mechanism of phosphotyrosine recognition. The subsequent question concerns the binding specificity of SH2 domain and the identification of critical residues responsible for the specificity.

Comparisons of SH2 target sequences in tyrosine-phosphorylated proteins (such as platelet derived growth factor receptor (PDGFR) and the polyoma virus middle T antigen) indicate that residues surrounding the phosphotyrosine in a linear fashion determine the binding specificity to the SH2 domains (Cohen *et al.*, 1990; Kazlauskas *et al.*, 1992; Talmadge *et al.*, 1989; Cantley *et al.*, 1991). However, a general picture of SH2 target specificity did not emerge until an exhaustive investigation was carried out using a peptide library approach (Songyang *et al.*, 1994; Songyang *et al.*, 1993). This established that the three residues immediately C-terminal to the phosphotyrosine are key determinants of specificity. The structure determination of the high affinity peptide complex of Src (and the closely related Lck) provided the first view of peptide recognition by the SH2 domains (Waksman *et al.*, 1993; Eck *et al.*, 1993) (Figure 1.2.1). By combining the structural information with selectivity data from the peptide library study, the sequence preference can be correlated with particular residues in the SH2 domain (Songyang *et al.*, 1994; Songyang *et al.*, 1995). Subsequently, the structures of peptide complexes of the SH2 domains of the tyrosine phosphatase SH-PTP2 (described in this

thesis, Lee *et al.*, 1994), phospholipase C-γ (Pascal *et al.*, 1994) as well as the adapter proteins, GRB2 (Rahuel *et al.*, 1996) and Shc (Zhou *et al.*, 1995) collectively further clarified the mechanism of peptide recognition and extended our understanding of the SH2 specificity.

An additional level of complexity was added when the biochemical and structural analysis extended toward larger signaling molecules containing more than one domain. Structures of the adapter protein GRB2 (Maignan *et al.*, 1995) and the regulatory unit of the Abl tyrosine kinase (Gosser *et al.*, 1995) have provided insights into spatial arrangements of multiple domains. Furthermore, structural analysis of the multi-domain constructs of ZAP-70 (Hatada *et al.*, 1995), Lck tyrosine kinases (Eck *et al.*, 1994), and the tyrosine phosphatase SH-PTP2 (Eck *et al.*, 1996) revealed the cooperative recognition of peptides by larger signaling molecules which contain multiple domains. The recently determined structures of two Src family tyrosine kinases (Hck and Src) further revealed how enzyme activity can be regulated by a complicated network of intramolecular interactions mediated by both the SH2 and SH3 domains

* * *

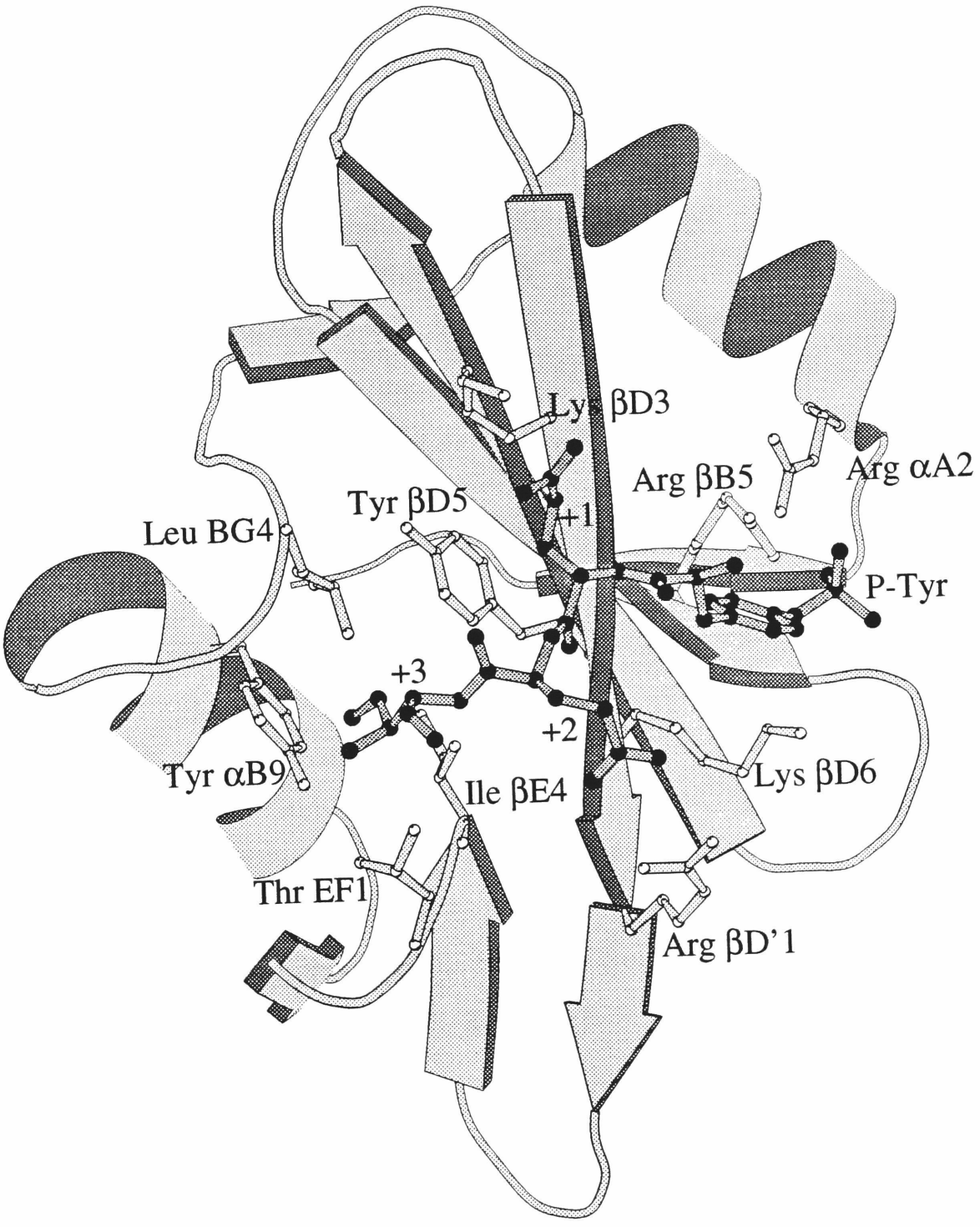
At the time when I started this project, there were only two structures of high affinity peptide complexes available (the closely related Src and Lck SH2 domains). In these structures, the modes of peptide recognition were found to be virtually identical. To gain further insight into the SH2 specificity, I chose to determine the structure of peptide complexes of an SH2 domain from the protein tyrosine phosphatase SH-PTP2 (also known as PTP1D, PTP2C or Syp).

The SH-PTP2 belongs to a class of phosphatases (including SH-PTP1, SH-PTP2 and Csw) that contain two SH2 domains N-terminal to the catalytic domain (Shen *et al.*, 1991; Plutzky *et al.*, 1992; Yi *et al.*, 1992). The SH2 domains of SH-PTP2 are implicated in localizing the phosphatase to the phosphorylation sites of activated receptors (Feng *et al.*, 1993; Kalzlauskas *et al.*, 1993). In addition, peptide binding to the SH2 domain stimulates the catalytic activity of the phosphatase domain (Lechleider *et al.*, 1993; Sujimoto *et al*, 1994), thereby suggesting a regulatory role for these SH2 domains.

The SH2 domains of tyrosine phosphatases are relatively distant in sequence from Src and Lck (~30% overall sequence identity). A notable feature of the phosphatase SH2 domains is the lack of a relatively conserved arginine residue at the phosphotyrosine binding site (it is replaced by glycine). In structures of Src and Lck, this arginine sidechain has been shown to interact with the phosphotyrosine by hydrogen-bonding and amino-aromatic interactions (Waksman *et al.*, 1992; Eck *et al.*, 1993; Waksman *et al.*, 1993). Thus, the structure of a phosphatase SH2 domain is of interest in terms of understanding how the binding site compensates for the loss of an apparently important interaction. Furthermore, the binding affinities of this SH2 domain with a number of phosphotyrosyl peptides from the platelet derived growth factor (PDGF) receptor, the epidermal growth factor (EGF) receptor and the insulin receptor substrate-1 (IRS-1) have been studied in detail (Case *et al.*, 1994). Hence, analysis of the peptide complex structures of the relatively

divergent SH-PTP2 SH2 domain provides an opportunity to extend our understanding of these molecular interactions.

Figure 1.2.1 Schematic diagram of the SH2 structure of the Src-Y*EEI complex. The view is from the peptide-binding surface and illustrates the secondary structure elements and the notation used. The peptide is shown in a ball-and-stick representation and comprises phosphotyrosine (pTyr), residue +1, residue +2 and so on. α-helices and β-strands are shown as ribbons and arrows, respectively.


Ly βD3
Arg βB5
Arg αA2
Tyr βD5
+1
Leu BG4
P-Tyr
+3
+2
Tyr αB9
Ile βE4
Lys βD6
Thr EF1
Arg βD'1

1.3 The PTB Domains

The phosphotyrosine binding or phosphotyrosine interaction (PTB or PI) domain was first identified as the component of the adapter protein Shc (Src homology 2 /collagen homology) that is responsible for its (Shc's) association with activated and tyrosine-phosphorylated receptors (Kavanaugh *et al.*, 1994; Kavanaugh *et al.*, 1995; Dikic *et al.*, 1995; Batzer *et al.*, 1995; Pratt *et al.*, 1996; Ravichandran *et al.*, 1996). The interactions mediated by the PTB domains are critical for the tyrosyl-phosphorylation of the adaptor proteins, Shc and IRS-1 (insulin receptor substrate-1) (Milia *et al.*, 1996; Isakoff *et al.*, 1996; Sawka-Verhelle *et al.*, 1996; O'Bryan *et al.*, 1996; Ravichandran *et al.*, 1996). These phosphotyrosine-containing sequences in turn serve as docking sites for other signaling molecules (Pelicci *et al.*, 1996; White *et al.*, 1996). The PTB domains are distinct from the SH2 domains in both structural and functional aspects. While the SH2 domains recognize the phosphotyrosine and peptide residues immediately following the phosphotyrosine (Waksman *et al.*, 1993), the PTB domains (such as Shc and IRS-1) interact with peptide residues that are N-terminal to the phosphotyrosine in the sequence context NPxY* (N, Asn; P, Pro; x, any amino acid; Y*, phosphotyrosine) (Kavanaugh *et al.*, 1995; Zhou *et al.*, 1995; Batzer *et al.*, 1995).

The unique nature of the PTB/peptide interaction was first revealed by the solution structure of the Shc PTB domain in complex with a phosphopeptide ligand (Zhou *et al.*, 1995) and was further illustrated by the crystallographic and NMR studies of the IRS-1 PTB

domain (Eck *et al.*, 1996; Zhou *et al.*, 1996). In these structures, the PTB domains incorporate their peptide ligands into the structure by extending one of the β-sheets, using anti-parallel hydrogen bonding interactions (Figure 1.3.1). The NPxY* motif, characteristic of the PTB domain ligands, is at the C-terminus of this β-strand and forms a β-turn. Mutation analysis of the PTB/peptide interactions indicated that Shc and IRS-1 PTB domains have distinguishable but overlapping peptide-binding specificity (Wolf *et al.*, 1995). A few interactions between the PTB domains and the peptide residues N-terminal to the NPxY* motif were identified and some of them can be attributed to their differences in the binding specificity. Based on the binding and structural studies of Shc and IRS-1 PTB domains, it was suggested that the PTB domains might exhibit a limited repertoire of sequence recognition (Eck *et al.*, 1996).

An unexpected finding from these structural studies is that the PTB domains share the same structural fold with the pleckstrin homology (PH) domains without any detectable sequence homology between them (Zhou *et al.*, 1995). The PH domains are small protein modules that have rather divergent functions, such as binding to proteins or phospholipids (Gibson *et al.*, 1994; Ferguson *et al.*, 1995). The PH domain family now has more than 90 members which share low sequence homology (some with just 10~15% sequence identity) within their family members. Likewise, the sequence homology between different PTB domains is often beyond the detection of conventional sequence comparison programs. Based on (i) the sequence and functional diversity of the PH domains and (ii) the

structural similarity between the PTB domains and PH domains, it was suggested to include the PTB domains as a subfamily of the PH domains (Lemmon *et al.*, 1996).

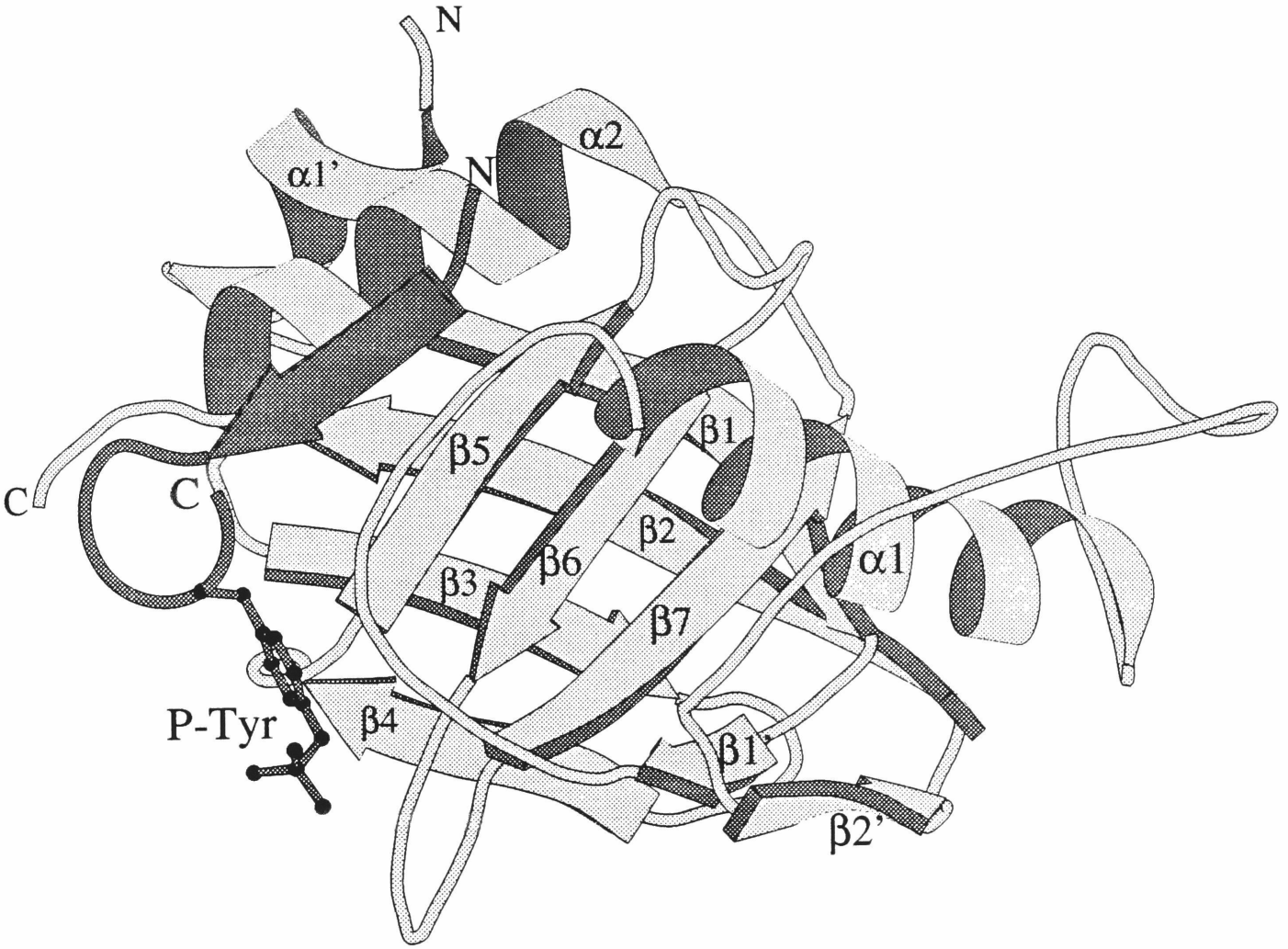
Although members of the PTB domain family have rather divergent amino acid sequences, a database search for proteins homologous to the Shc PTB domain has led to the identification of a small number of novel PTB domains (Borg and Margolis, 1995). A neuronal-specific protein, X11, was found to contains a novel PTB domain, as well as two PDZ domains (Duclos *et al.*, 1993; Borg and Margolis, 1995). It has been recently reported that the X11 protein via its PTB domain interacts consistently with the cytoplasmic domain of the β-amyloid precursor protein (β-APP) *in vivo* (Borg *et al.*, 1996). Although its biological consequence(s) remains unclear at present, the interaction between the X11 PTB domains and β-APP appears to be of high affinity and high specificity. In culture cells, β-APP is the major protein ligand for the X11 PTB domain. In addition, the X11 PTB domain discriminates against other potential PTB-binding proteins that contain the NPxY motif. In contrast to the Shc or the IRS-1 PTB domains, the X11 PTB domain does not interact with a number of tested growth factor receptors at either activated or resting states (Margolis *et al.*, unpublished results).

Previous biochemical characterizations of the X11/APP interaction indicated that a NPTY motif located at the carboxyl-terminus of β-APP is essential for its association with the X11 PTB domain. A 14mer peptide (APP peptide) encompassing the sequence

QNGEYNPTYKFFEQ efficiently competes with full-length β-APP in the binding to the X11 PTB domain, suggesting that this local sequence accounts for the binding. This region is also required for the internalization of β-APP, a process that leads to the degradation of β-APP and the formation of the pathological β-amyloid peptides (Aβ) (Haass *et al.*, 1993; Lai *et al.*, 1995). Most interestingly, the interaction between the X11 PTB domain and β-APP (as well as the APP peptide) appears to be independent of phosphorylation. Also, the phosphorylation-independent interaction between β-APP and the FE65 PTB domain has been reported (Fiore *et al.*, 1995) though it appears to involve regions of β-APP distant from the NPTY motif (Margolis *et al.*, unpublished). The phosphorylation-independent interactions have been reported for the Shc PTB domain (Charest *et al.*, 1996), even though the tyrosyl phosphorylation of the ligands is often required for interacting with the PTB domains of Shc and IRS-1. Thus, the dispensability of phosphorylation for binding is not unique to the X11 PTB domain.

In an effort to understand the interaction between the X11 PTB domain and the APP peptide, I have determined the structure of the X11 PTB domain in complex with an unphosphorylated APP peptide. The determined structure, in combination with the biosensor binding studies (provided by Mandiyan, *et al.*) revealed the structural basis for the X11 PTB specificity in binding to the APP peptide (see Results section).

Figure 1.3.1 The Solution Structure of Shc PTB domain. The bound peptide is shown in darker gray with the phosphotyrosine represented in a ball-and-stick model.


N
α2
α1'
N
β1
β5
C
C
β2
β6
α1
β3
β7
P-Tyr
β4
β1'
β2'

1.4 The SH3 Domains

The SH3 domains are small modules composed of around 60 amino acids. Their functions are to mediate protein-protein interactions involved in signal transduction and cytoskeleton modulation (reviewed in Pawson, 1993). The characterization of cellular proteins that bind to the SH3 domains has led to the identification of SH3 ligands as short proline-rich peptide sequences with a minimal PxxP motif (Cicchetti *et al.*, 1992; Ren *et al.*, 1993).

Structures of the SH3 domains were first determined for Src (Yu *et al.*, 1992) and spectrin (Musacchio *et al.*, 1992) in an unliganded form by NMR and X-ray crystallography, respectively. The highly conserved SH3 fold is composed of two small antiparallel β sheets that pack against each other to form a barrel-like structure. A noticeable feature of the unliganded SH3 structures is a very shallow groove lined with several highly conserved aromatic residues. This groove forms the ligand binding site. The structural determination of p85 (Yu *et al.*, 1994) and Abl SH3 domains (Musacchio *et al.*, 1994) in separate complexes with peptide ligands established that the proline-rich peptide adopts a polyproline-type II (PP-II) helix that interacts with the SH3 domain, as anticipated by a previous modeling study (Lim *et al.*, 1994). Structural analyses of the peptide complex of the Sem-5 (Lim *et al.*, 1994) and Src SH3 domains (Feng *et al.*, 1994) reveal that the peptide can bind in two opposite orientations (Figure 1.4.2) which are determined by a general mechanism (discussed below). Furthermore, biochemical and

structural analyses of the interaction between HIV-1 Nef protein and Src family SH3 domains reveal how tertiary interactions can augment the binding affinity and specificity of the SH3 domains (Lee *et al.*, 1996; Lee *et al.*, 1995; see Result section).

In binding to the SH3 domain, the proline-rich peptides adopt a PP-II helix conformation which exhibits a three-fold pseudo-symmetry in cross section. When it binds to the SH3 domain, two of its three edges provide six peptide residues (denoted as P_{-3}, P_{-2}, P_0, P_{-1}, P_{+2} and P_{+3}) that fit in corresponding binding pockets on the SH3 surface (Figure 1.4.1). The interface is thus pseudo-symmetrical (given that P_{-1} and P_0 are equivalent, so are P_{+2} and P_{+3}). An interesting consequence is that the peptide can bind in two opposite orientations —— namely, plus and minus (Figure 1.4.2). The orientation is primarily determined by an ionic interaction between a conserved acidic residue and a basic residue (usually an Arg) at P_{-3} position. Peptide residues at P_{-1}, P_0, P_{+2}, and P_{+3} positions which interact with the hydrophobic binding surface of the SH3 domains are often either proline or other hydrophobic residues.

The pseudo-symmetry of the PP-II helix breaks down in the presence of non-proline residues in the helix. For example, the P_{-1} and P_0 are no longer equivalent (Figure 1.4.3). Non-proline residues at one particular edge will adopt preferred conformation for tight packing to SH3 domain only in one of the two orientations. In other words, the preference of proline residues over non-proline at certain positions is intimately linked with the orientation of the bound

peptide. To illustrate, when the peptide binds to the SH3 domain in the minus orientation, non-proline residues are tolerated only at one edge (corresponding to positions P_{-1} and P_{+2}) since at the other edge (P_0 and P_{+3}) the sidechain of the non-proline residue will extend away from the binding surface (Figure 1.4.3). Thus, proline residues are required on one edge (P_0 and P_3) in the plus orientation but on the opposite edge (P_{-1} and P_2) in the minus orientation. As a result, this leads to the consensus sequence, PxxP, that defines the SH3 ligands. To summarize in the simplest form, the peptide containing motif "PxxPxR" is likely to bind to the SH3 domains in a "minus" orientation while peptides with the sequence "RxxPxxP" will bind in a "plus" orientation because of the positions of the basic residue and the proline residues (Figure 1.4.2).

The binding of PxxP-containing peptides to SH3 domains is rather weak in general, with K_D around 2 to 50 μM (Yu *et al.*, 1994). This may be due to the relatively small interface (~380 $Å^2$) observed between peptides and SH3 domains. Residues of SH3 domains that are involved in the interaction are highly conserved among different SH3 domains. Two highly variable loops, the RT- and n-Src loops (so named because of the critical Arg and Thr for Src SH3 function and a insertion observed in the neural form of Src, respectively) make limited contacts with the bound peptide although they border the PxxP binding surface. Because of the conserved nature of the binding surface, SH3-PxxP interactions are relatively promiscuous. The observed low affinity and low specificity raise the question of how SH3 domains mediate specific interactions.

Figure 1.4.1 The SH3-PxxP Interaction, for the GRB2/Sem-5 SH3 domain complexed with a Sos peptide bound in the minus orientation (Lim et al., 1994). The bound PxxP peptide is shown in a ball-and-stick representation. The conserved hydrophobic residue that form the peptide binding surface and the acidic residue that interacts with Arg at P_{-3} of the peptide are shown in light-gray stick. This figure is based on a previous report (Lim, *et al.* 1994).

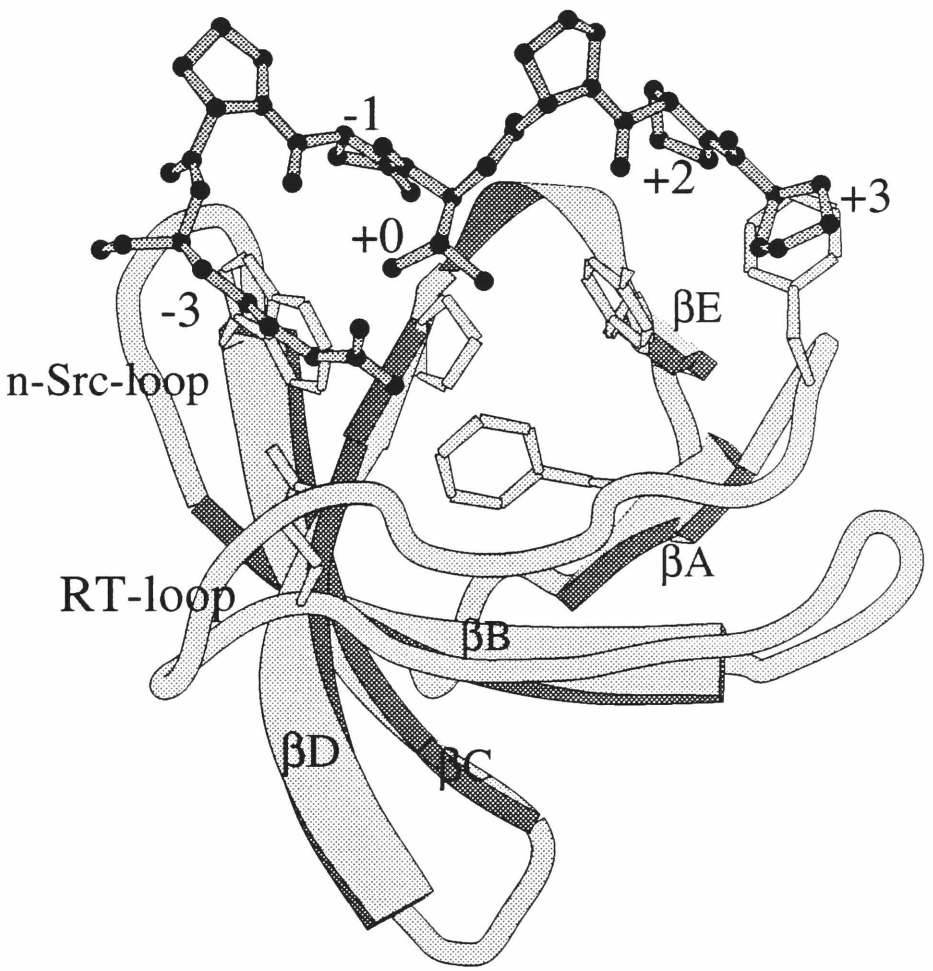
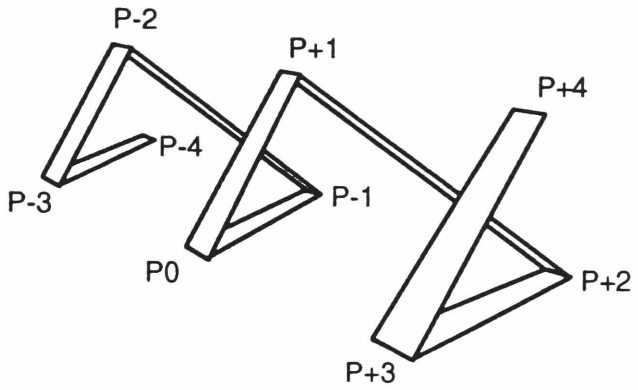
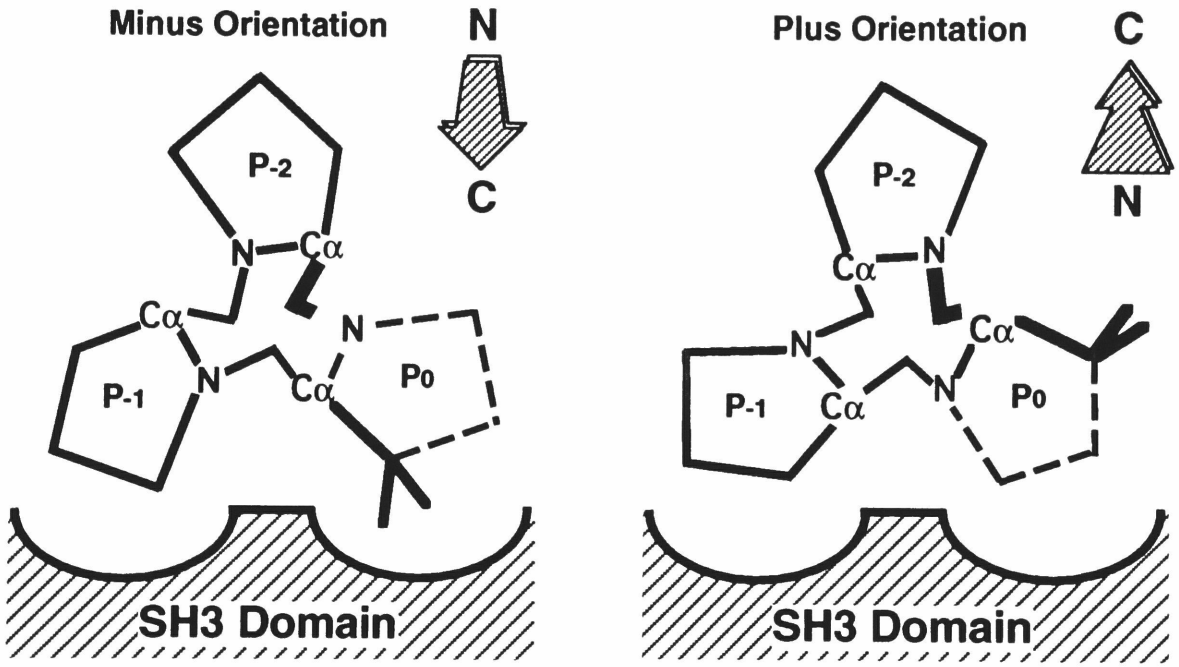

-1
+2
+3
+0
-3
βE
n-Src-loop
βA
RT-loop
βB
βD
βC

Figure 1.4.2 Alignment of the PxxP motif of SH3 ligands. The positions P_{-3}, P_{-1}, P_0, P_2 and P_3 contain the ligand residues which interact with the SH3 domain. The spacing of positions are also shown in the ribbon diagram representing a left-handed PPII helix. This figure is taken from a previous report (Lee, *et al.* 1994).



Orientation		P-3	P-2	P-1	P0	P1	P2	P3		Origin	SH3 domain
	N-	R	K	L	P	P	R	P	-C	*Screen	PI3K
+	N-	R	A	L	P	P	L	P	-C	*Screen	Src
	N-	M	P	P	P	L	P	P	-C	3BP1	Abl
	C-	R	P	P	V	P	P	P	-N	mSos1-1	Grb2/Sem5
	C-	R	R	P	L	P	P	H	-N	*Screen	Src
-	C-	R	S	P	V	P	P	A	-N	Dynamin	p85 PI3K
	C-	**R**	L	**P**	**V**	x	**P**	R	-N	Nef	Hck

Figure 1.4.3 Two types of packing geometries, with distinct preferences for proline vs. non-proline residues. In the minus orientation (left panel), non-proline residue can adopt preferred conformation at P_0 for tight interaction with SH3 domain while in plus orientation (right panel), non-proline residue at P_0 position will extend its sidechain away form the binding surface of the SH3 domain, resulting in unfavorable interaction. This figure is based on a previous report (Lim, *et al.* 1994).


Minus Orientation
N
C
P-2
Cα
N
Cα
N
P-1
N
Cα
P0
SH3 Domain
Plus Orientation
C
N
P-2
Cα
N
N
Cα
P-1
Cα
N
P0
SH3 Domain

A degree of specificity has been reported previously for certain cases of SH3-peptide interactions. For instance, the Abl SH3 binds preferentially to peptides containing hydrophobic residues at P_{-3} position while other SH3 domains favor basic residues at the same position (Cheng *et al.*, 1994; Rickles *et al.*, 1994). This appears to be due to the lack of an acidic residue (replaced by a Thr in Abl SH3) found in most SH3 domains, which mediates the ionic interaction with the peptide residue Arg P_{-3} (as discussed above). In addition, the N-terminal SH3 domain of Crk binds specifically to a peptide derived from a guanine nucleotide exchange factor C3G and this interaction also involves selectivity at the P_{-3} position (Knudsen *et al.*, 1995). C3G peptide contains a Lys (instead of an Arg) at the P_{-3} position which is tightly coordinated to three acidic residues on the RT-loop of the Crk SH3 domain (Wu *et al.*, 1995). The recognition of Lys appears to be specific: replacement of Lys P_{-3} with even Arg reduced binding affinity by at least 10 fold. An X-ray structure of the Crk SH3 in complex with the Sos peptide (containing Arg at P_{-3}) reveals that the Arg binds in a suboptimal conformation (Wu *et al.*, 1995).

As discussed above, the reason for the lack of specificity in SH3-peptide interactions is due to limited variation in the interaction surface. Selection with longer peptide segments and non-natural analogs has demonstrated that higher affinity and selectivity can be obtained by exploiting non-conserved regions of SH3 (Combs, *et al.*, 1996). However, the specificity observed in these studies and others is generally less significant than other peptide-domain interactions

such as that of the SH2 domains. Thus, it is likely that many SH3 domains can only achieve high affinity and high specificity by cooperative interaction and/or tertiary interaction.

1.5 The HIV-1 Nef Protein

The nef gene of human and simian immunodeficiency viruses (HIV-1, HIV-2, SIV) codes for a protein that is critical for the development of AIDS (reviewed in Trono, 1995). Rhesus monkeys inoculated with pathogenic strains of SIV develop an AIDS-like disease and eventually die as a consequence. Remarkably, animals infected with an isogenic strain of SIV missing Nef remain healthy and acquire long-term immunity that protects them from subsequent challenge with pathogenic strains of SIV containing Nef (Kestler *et al.*, 1991; Daniel *et al.*, 1992). Likewise, viruses defective in the nef gene have been isolated from some humans who are experiencing long-term non-progressive HIV-1 infection (Kirchhoff *et al.*, 1995; Deacon *et al.*, 1995).

Although Nef is clearly essential for disease progression, the mechanism of its action is poorly understood. Nef (i) enhances viral infectivity and replication in primary cells, (ii) alters the state of T-cell activation, and (iii) reduces the surface expression of CD4, the major receptor for HIV (reviewed in Trono, 1995). Nef has no known catalytic activity and is, therefore, likely to function through interactions with cellular proteins ——particularly those involved in cellular activation and signaling. In particular, Nef has been shown to

associate with a Ser/Thr protein kinase, PAK68 (Sawai *et al.*, 1994). PAK68 (also known as PAK1) is a member of p21-associated kinases, which physically interact with Rho-family GTPases and respond to stress or proliferation signals. Furthermore, it was demonstrated that Nef also interacts with the SH3 domains of certain Src-family tyrosine kinases, such as Hck and Lyn (Saksela et al., 1995).

The interaction of Nef with Src-family tyrosine kinases is intriguing because these proteins have diverse and critically important roles in many cellular signaling pathways (Bolen, 1993). The Src-family kinases have in common a modular architecture composed of a conserved catalytic domain and two other "Src-homology" domains known as SH2 and SH3 (reviewed in Pawson, 1995 and Cohen *et al.*, 1995). The physiological relevance of the interaction of Nef with the SH3 domains is unknown. However, mutagenesis of Nef showed that the PxxP motif is essential for interacting with the SH3 domain as well as for optimal spread of HIV-1 virus in primary cell cultures (Saksella *et al.*, 1995). Recent animal studies showed that SIV viruses carrying mutant Nef that is unable to interact with Src-family tyrosine kinases fail to cause disease in Rhesus monkeys (Sawi *et al.*, 1997). These results suggest that the virus has evolved to exploit SH3-mediated interactions with cellular proteins in order to somehow enhance its replication (Saksela *et al.*, 1995).

2. Materials and Methods

2.1 Structural Determination of SH-PTP2 SH2 Domain

2.1.1 Protein Expression and Purification

The cDNA of SH-PTP2 was provided by Benjamin Margolis. Polymerase chain reaction was used to amplify the region of the *SH-PTP2* gene encoding the N-terminal SH2 domain (residues 4 to 103). This region was inserted into the pET14b expression vector (Novagen, Inc.) using NdeI- BamHI sites. The protein corresponding to the resulting coding region consists of a hexa-His tag (MGSSHHHHHHSSGLVPRGSHM) fused to residue 4 of SH-PTP2 and extends to residue 103 of SH-PTP2. The complete DNA sequence of the SH2 domain was confirmed by dideoxy sequencing. Ten liter cultures of *Escherichia coli* strain BL21(DE3) transformed with this expression vector were grown at 30° C. After induction with 0.5 μM isopropyl-thio-β-D-galactoside (IPTG) for 3 hours, cells were harvested by centrifugation and re-suspended in 300 ml of PBS (8 mM Na_2HPO_4, 2 mM K_2PO_4, 0.14 M NaCl, 10 mM KCl, pH 7.4) containing 1 mM phenylmethylsulfonyl fluoride (PMSF), 5 μg/ml leupeptin and 5 μg/ml aprotinin. Cells were lysed by four cycles of freeze and thaw, followed by digestion of DNA by addition of DNAse I (final concentration of 0.1 μg/ml) and $MgCl_2$ (final concentration of 15 mM). Cell lysate was centrifuged and the supernatant was loaded onto a 60 ml Ni-ion affinity column (IMAC, Pharmacia) pre-equilibrated with buffer A (20 mM HEPES, pH 8.4, 1 M KCl, 30 mM

imidazole, 1 mM PMSF) and washed with increasing amounts of imidazole in buffer A. The His-tagged SH2 domain was eluted from the resin with 100 mM EDTA in buffer A, and was approximately 95% pure as judged from SDS-polyacrylamide gel electrophoresis and staining with Coomassie blue. The N-terminal histidine containing sequence was removed by proteolytic cleavage using thrombin (0.2U/1mg of protein) at 4°C for 5 hours. The reaction proceeded cleanly, leading to a single major product. A final purification step using cation exchange chromatography (S-Sepharose Fast Flow column, Pharmacia) was carried out to remove the protease and remaining contaminants, and yielded material of approximately 98% purity, as judged by SDS gel electrophoresis. This procedure yields approximately 4 mg of purified protein from 1 liter of culture.

Purified phosphotyrosyl peptides were provided by Steve Shoelson (Harvard University) and were prepared using N^{α}-Fmoc-O-(*O*, *O*-dimethoxyphosphoryl)-L-tyrosine in a stepwise N^{α}-Fmoc synthetic strategy. These peptides were purified by HPLC and lyophilized in volatile buffer. Prior to use in crystallization trials, the lyophilized peptides were reconstituted in 50 mM of Tris buffer (pH 8.0) and pH was adjusted using 2M of NaOH.

2.1.2 Crystallization and Data Collection

The purified protein was concentrated to 20 mg/ml by ultrafiltration (Centricon, Amicon) in a buffer containing 10 mM MES (pH 5.5) and 50 mM KCl. Crystallization conditions were scanned

using the hanging drop method. In the absence of peptide, large crystals (approximate dimensions 0.4 x 0.4 x 0.5 mm^3) were obtained with ammonium sulfate as the precipitant (50% saturated) and MES buffer (100 mM, pH 6.0), at room temperature. The crystals are tetragonal ($P4_32_12$; a=63.2 Å, c=76.15 Å) and show diffraction spots to 2.2 Å spacings. Crystals of complexes of the SH2 domain with a low affinity peptide (PDGFR-740: DGGpYMDMSKDE, encompassing phosphotyrosine 740 of the PDGF receptor) and a high affinity peptide (IRS1-895: SPGEpYVNIEFGS, encompassing phosphotyrosine 895 of the IRS-1 protein) are obtained with tetragonal crystal lattices that are very similar to that for the uncomplexed protein (Table 2.1.2.1). Large tetragonal co-crystals (0.4 x 0.4 x 0.3 mm^3) of PDGFR-740 ($P4_32_12$; a = 62.2 Å, c=74.1 Å) were obtained from a protein drop containing the protein-peptide mixture (ratio 1:1) in 20% polyethylene glycol (PEG, average M_r=4000) and 100 mM MES buffer (pH 6.0), at room temperature. Similar co-crystals ($P4_32_12$, a=62.9 Å, c=77.8 Å) were obtained for the IRS1-895 peptide from 25% PEG monomethyl ether (average M_r=2000) in MES buffer (100 mM pH 6.0) in the presence of 50 mM of ammonium sulfate at 4°C. Crystals of SH-PTP2 SH2 complexed with another high affinity peptide (PDGFR-1009: SVLpYTAVQPNE, encompassing phosphotyrosine 1009 of the PDGF receptor) are quite different in morphology and crystal symmetry ($P2_1$; a=33.8 Å, b=52.7 Å, c=56.4 Å, β=101.1). The crystals (approximate dimensions 0.4 x 0.3 x 0.05 mm^3) were obtained from a protein drop containing the protein-peptide mixture (ratio 1:1) in 15% of PEG (average M_r=4000) and MES buffer (100 mM, pH 6.0), at 4°C.

Diffraction data were measured using a Rigaku RAXIS IIC area detector mounted on a Rigaku RU200 X-ray generator (Rigaku, Japan and Molecular Structure Corporation, USA), using the oscillation method. Diffraction data were indexed and space groups were assigned using software provided by Rigaku and by using the DENZO program (Z. Otwinowski, unpublished). The absence of reflections along the *00l* axis but for the (l=4n) reflections and the absence of the odd (h=2n+1) and (k=2n+1) reflections along the *h00* and *0k0* axes is consistent with space group $P4_12_12$ or its enantiomorph $P4_32_12$. The four-fold axis is along the longest axis of the crystals. Ambiguities in space group assignment were resolved by comparing the results of molecular replacement calculations in each of the alternative space groups (see below). For crystals of the PDGFR-1009 complex, the PDGFR-740 complex and for the uncomplexed form, radiation damage was minimized by flash-freezing the crystals after briefly soaking in solutions containing 20% glycerol in addition to the appropriate components (Hope, 1990). In each of these cases, a single crystal was used for the data collection. Optimal cryoprotection conditions were not found for crystals of the IRS1-895 complex. Data on this complex was collected at 4°C, using a single crystal. Radiation damage reduced the effective resolution of the resulting data from 2.5 Å (at the beginning) to 3.0 Å (at the end). Only data to 3.0 Å spacings were used in the analysis. Statistics of the data collection are given in Table 2.1.3.1.

2.1.3 Molecular Replacement and Model Refinement

The structure of the uncomplexed form of SH-PTP2 was determined first. Molecular replacement calculations were carried out using the program X-PLOR (Brunger, 1992) with the complete structure of the Src SH2 domain (peptide-A complex with the peptide removed) as the search model (Waksman, 1993). This crystal form has one molecule in the asymmetric unit. Rotation function calculations followed by Patterson correlation refinement (Brunger, 1990) using data to 4.0 Å resolution yielded an unambiguous orientation of the search model. Translation functions were calculated in space groups $P4_32_12$ and $P4_12_12$, using data to 4 Å resolution. The value of the translation function was significantly higher in the former case and corresponded to an orientation of the molecule with no bad contacts. Rigid body refinement of the model followed by refinement of all atomic positions by conjugate gradient optimization (using X-PLOR) led to the calculation of electron density maps that were interpretable in terms of changes in the sequence of SH-PTP2 relative to Src. The amino acid sequence of the model was changed to that corresponding to SH-PTP2 in alternating cycles of manual rebuilding, using the program "O" (Jones, 1991). The model was then subjected to least-squares refinement and simulated annealing using the program X-PLOR, which proceeded as described previously (Weis, 1990). The phosphate binding loop (BC) and the longer BG loop of SH-PTP2 are disordered in this structure. As a

result, a satisfactory model could not be built for them. The final R-value of the model to 2.2 Å is 20.0% (Table 2.1.3.1).

Crystals of the IRS1-895 and PDGFR-740 complexes are nearly isomorphous to those of the uncomplexed form (Table 2.1.3.1). Structure determination proceeded by first optimizing the position and orientation of the SH-PTP2 SH2 model, obtained previously, by rigid body refinement, and then carrying out approximately 200 steps of conjugate gradient optimization using X-PLOR. Difference Fourier maps showed very clear density for the phosphotyrosine group and the peptide at the expected positions. Note that this confirms the correctness of the molecular replacement solution, since the peptide was not used in the structure determination and refinement of the uncomplexed SH-PTP2 SH2 domain. Refinement of these models proceeded smoothly, resulting in final R-values of 16.4% for the IRS1-895 structure (3.0 Å) and 18.9% for the PDGFR-740 structure (2.3 Å). The peptide binding loops (BC and BG) are well ordered in the high affinity complex (IRS1-895) but are disordered in the non-specific complex (PDGFR-740). Water molecules were included in the models for the uncomplexed form and for the PDGFR-740 complex, but were not included in the IRS-895 complex due to the limited resolution of the data.

Crystals of the PDGFR-1009 complex are monoclinic rather than tetragonal and have 2 molecules in the asymmetric unit. These crystals are particularly well ordered and diffract strongly to beyond 2.0 Å resolution. The two molecules were placed by rotation and

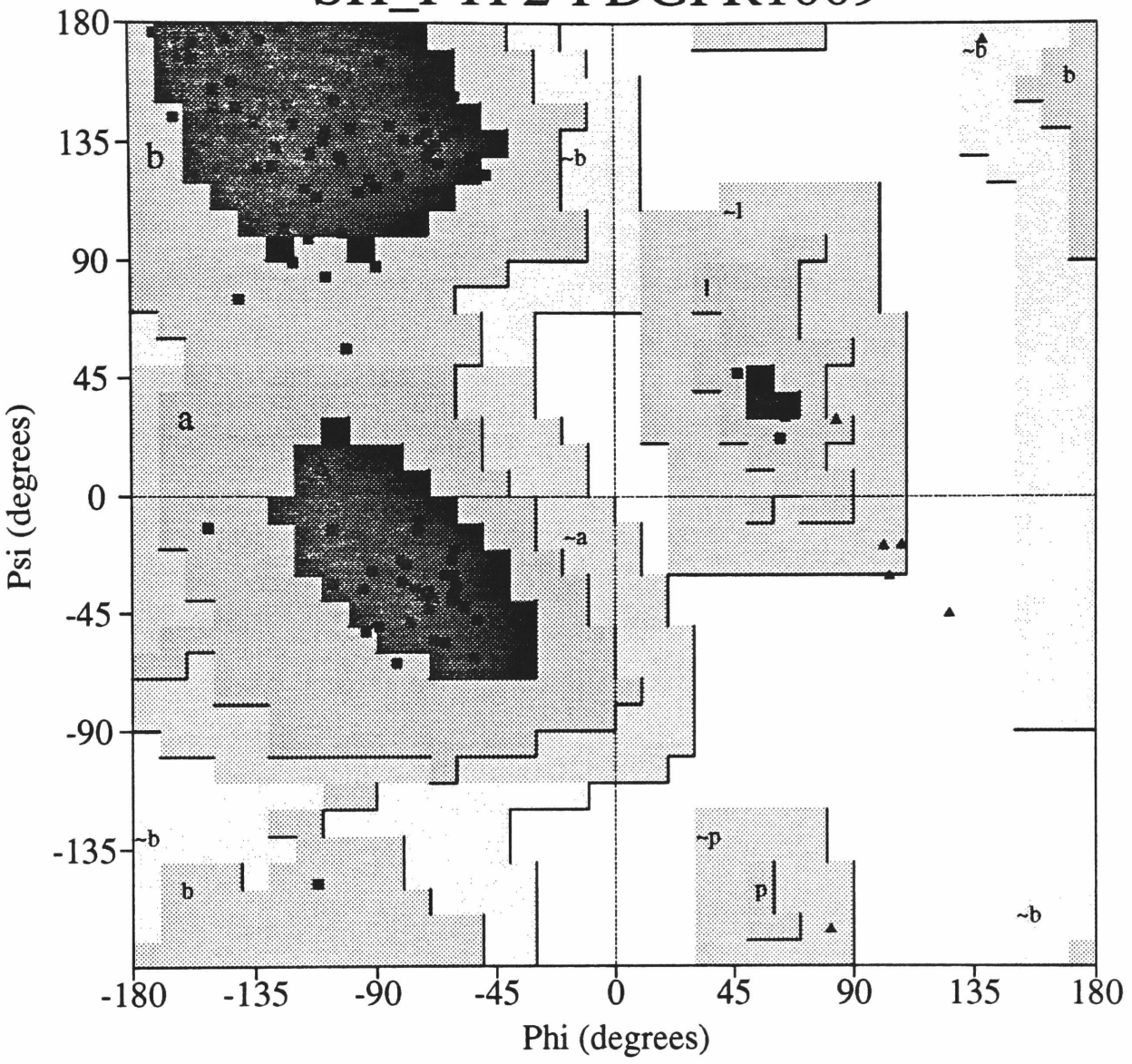
translation functions using the IRS1-895 structure as the search model. Least-squares refinement using X-PLOR resulted in a final model with an R-value of 18.0% (2.0 Å). The final model is of excellent stereochemical quality and all residues lie in the stereochemically allowed regions (Figure 2.1.2.1). Although the 2 molecules in the asymmetric unit are very similar, the temperature factors of one of them are somewhat lower and this molecule (the second one in the model) is used for all analysis.

Table 2.1.3.1 X-ray data collection and refinement statistics for SH-PTP2

	High Affinity Complex PDFGR-1009	High Affinity Complex IRS1-895	Low Affinity complex PDGFR-740	Uncomplexed
Resolution (Å)	30-2.0530-3.0	30-2.3	30-2.2	
Space group	$P2_1$	$P4_32_12$	$P4_32_12$	$P4_32_12$
Unit Cell	a=33.75, b=52.7, c=56.4 β=101.08	a=62.92, c=77.78	a=62.21, c=74.07	a=63.21 c=76.15
Temperature of data collection	-160 °C	4 °C	-160 °C	-160 °C
No. of Oberved Reflections	26661	5965	43918	63420
No. of Unique Reflections	11145	2549	6914	8075
Completeness of all data	90.2%	73%	96%	97%
I/(σI) in all shells	10.6	11.7	36.8	31.7
completeness of outer shell	83.2%(2.12-2.21Å) 73% (2.05-2.12Å)	83.3% (3.05-3Å)	90.9% (2.35-2.3Å)	92.1% (2.24-2.2Å)
I/(σI) in outer shell	3.2 (2.25-2.05 Å)	5.9 (3.05-3Å)	5.8 (2.35-2.3Å)	6.8 (2.24-2.2Å)
Rsym	8.4%	8.9%	5.4%	5.3%
R_{cryst} (|F|>2(σF))	18.0% (6-2.05Å)	16.4% (6-3.0Å)	18.9% (6-2.3Å)	20.0% (6-2.2 Å)
No. of Unique Reflections (|F|>2(σF))	10547 (6-2.05Å)	1858 (6-3.0Å)	5362 (6-2.3Å)	6359 (6-2.2Å)
No. of molecules in asymmetric unit	2	1	1	1
No. of non-hydrogen atoms in final model	1727	853	837	782
No. of water molecules in final model	141	0	63	87
rms deviation of bond lengths(Å)	0.014	0.014	0.012	0.012
rms deviation of angles (degrees)	2.99	3.0	2.89	2.98

Figure 2.1.3.1 Ramachandran plot of the 2.1 Å refined SH-PTP2/PDGFR-1009 model. The glycine residues are indicated by triangles. No residue is in the disfavored region of the plot. The plot is made with the PROCHECK program (Laskowski *et al.*, 1993).

Ramachandran Plot
SH_PTP2-PDGFR1009



Plot statistics

Residues in most favoured regions [A,B,L]	74	84.1%
Residues in additional allowed regions [a,b,l,p]	14	15.9%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	88	100.0%
Number of end-residues (excl. Gly and Pro)	5	
Number of glycine residues (shown as triangles)	10	
Number of proline residues	5	
Total number of residues	108	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

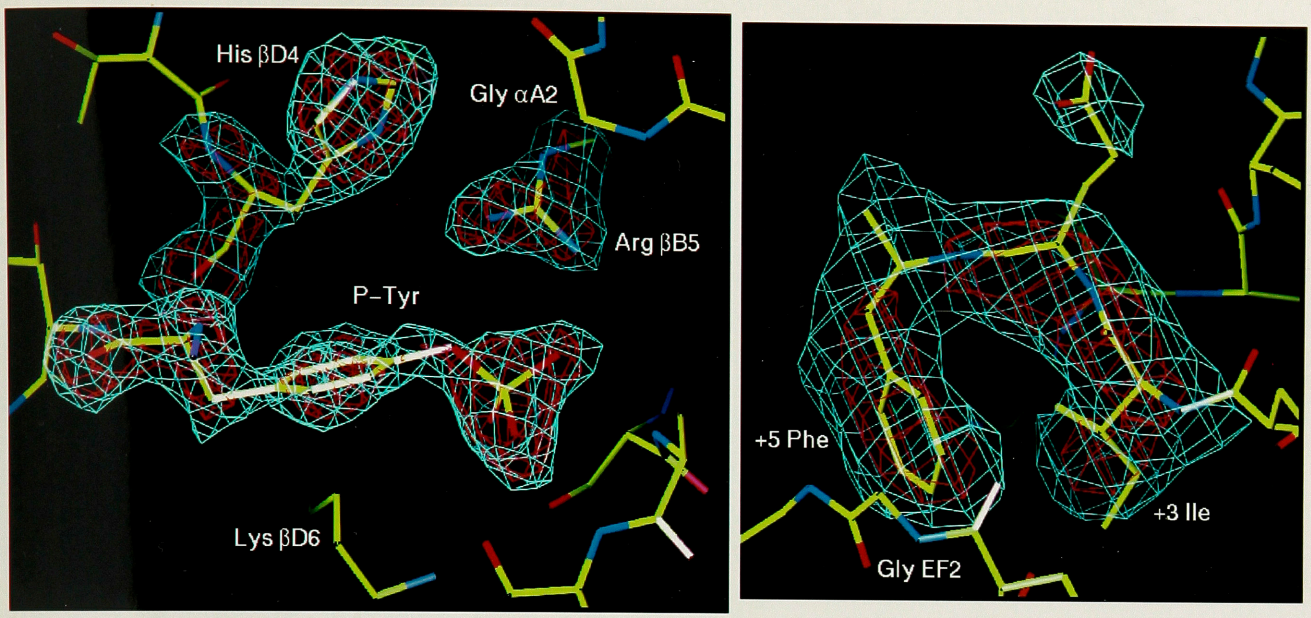


Figure 2.1.3.2 Electron density maps for (a) the SH-PTP2/PDGFR-1009 complex, and (b) the SH-PTP2/IRS1-895 complex. (a) The phosphotyrosine, His βD4 and Arg βB5 were removed from the model of PDGFR-1009 complex and the electron density map was calculated using coefficients $(|F_o|-|F_c|)\exp(-i\alpha_c)$ at 2.1 Å resolution, where F_o is the observed structure factor amplitude and F_c and α_c are the amplitude and phases calculated from the partial model. Blue and red contour lines indicate electron density at 2.7 σ and 4.5 σ, respectively. (b) Ile (+3), Glu (+4) and Phe (+5) were removed from the model of the IRS1-895 complex to calculate a difference map at 3 Å resolution, as in A. Blue and red contour lines indicate electron density at 2.2 σ and 4 σ, respectively. The protein and peptide atoms are shown in stick figure and coloured in yellow for carbon atom, blue for nitrogen atom and red for oxygen atom.

2.2 Structure Determination of the X-11 PTB Domain

2.2.1 Space Group Determination and Data Collection

Crystals of the X11-peptide complexes were provided by Zhongtao Zhang from Schlessinger's laboratory at NYU Medical Center. These crystals were obtained in several days at 21°C with ammonium sulfate as the precipitant (1.4 M ammonium sulfate, pH 6.5, 100 mM MES, 100 mM NaCl, for the 10-residue peptide complex; and 1.8 M ammonium sulfate, pH 6.5, 100 mM MES, for the 14-residue peptide complex). The crystals belong to the tetragonal space group (a=b=74.4/74.6 Å, c=157.1/155.4 Å for 10/14 residue peptide complexes, respectively) and diffract to beyond 2.5 Å using an in-house X-ray source. The space group $P4_12_12$ (or its or its enantiomorph $P4_32_12$) was assigned based on (i) the absence of reflections along the *00l* axis except for the $l=4n$, and (ii) the absence of the odd ($h=2n+1$) reflections along the *h00* axis. The four-fold axis is along the longest axis of the crystals. The unit cell volume is 8.8 X 10^5 $Å^3$. Assuming two molecules per asymmetric unit, the specific volume V_M=2.5 per Dalton of protein corresponds to a solvent content of 50% (v/v) (Matthews, 1968). Crystals of the Se-Met labeled protein diffracted X-rays more strongly than those of native protein. Therefore, the Se-Met crystals were used for all the structural analysis.

All diffraction data was collected from crystals cooled to 100K after cryoprotection in 5%-30% glycerol. X-ray data for MIR analysis

was collected on a Rigaku RAXIS IIC area detector mounted on a Rigaku RU200 X-ray generator (Molecular Structure Corporation, USA). Crystals were derivatized before data collection by soaking in 15 mM of trimethyl lead acetate for 24-36 hours. MAD data were collected from a single frozen crystal (10 residue peptide complex; Se-Met derivative) at the National Synchrotron Light Source (Brookhaven Nation Laboratory) using beamline X25. All data was processed using program DENZO and SCALEPACK (Z. Otwinowski and W. Minor, unpublished).

2.2.2 Heavy Atom Derivatives

Crystals of the 10 residue peptide complex soaked in solutions containing 15 mM trimethyl lead acetate yielded a derivative and still retained high resolution diffraction (2.5 Å). The positions of two bound lead atoms were determined using the program SHELX-90 (Sheldrick *et al.*, 1991). Heavy atom parameters were then refined and initial phases were calculated for two alternative space groups ($P4_12_12$ and $P4_32_12$) in combination with two different heavy atom configurations, using the program MLPHARE. The SIRAS maps calculated from the "best phases" (Blow and Crick, 1959) were basically uninterpretable even after density modification by solvent flattening, consistent with the low figure of merits (0.32).

2.2.3 Phasing by MAD/MIR

A multi-wavelength anomalous diffraction (MAD) experiment was performed on the Se-Met derivative of the 10 residue peptide complex. For MAD data collection, a single Se-Met crystal was mounted ——after flash freezing ——with the c* axis approximately parallel to the spindle in order to facilitate the data collection. The inverse beam method was used to collect Bijvoet pairs of reflection. Data collection at three wavelengths near the Se absorption edge (Table 2.2.3.1) was recorded on a Mar detector and processed with DENZO and SCALEPACK. The positions of 13 Se atoms (out of a total 16 Se atoms) in the asymmetric unit were identified by different Fourier techniques using (i) the phases calculated from the MePbAc derivative and (ii) anomalous (and dispersive) difference of structure factors (at different wavelengths). Heavy-atom parameters were then refined and phases were calculated using the program MLPHARE (Z. Otwinowski) for two alternative space groups ($P4_12_12$ and $P4_32_12$). Summary statistics from the MAD phasing procedure are given in Table 2.2.3.1. Using only MAD data, phases calculated at 2.9 Å resolution with space group $P4_12_12$ yielded an interpretable electron density map which was further improved by solvent flattening and histogram matching using SQUASH (Zhang *et al.*, 1990). The experimental electron density map calculated using only MAD data is shown in Figure 2.2.3.2. The information from the lead derivative was not used further.

2.2.4 Model Building and Refinement

A model accounting for >90% of the structure was built into the MAD map using the program O (Jones *et al.*, 1991) and was refined using X-PLOR (Brünger, 1988) without simulated annealing. The free R value (Brünger *et al.*, 1993) was used to monitor all stages of the refinement. The asymmetric unit contains two X11 PTB/APP peptide complexes which are very similar to each other. Although the non-crystallographic symmetry was not explicitly used in the refinement processes, it provided a useful check on the accuracy of the model. The resolution was later extended to 2.3 Å using a data set collected in-house. Well-ordered solvent molecules were included at this stage, and tightly restrained individual isotropic B-factors were refined. The refinement proceeded smoothly, without the utilization of simulated annealing refinement.

The statistics for data collection, phase determination and refinement are given in Tables 2.2.3.1 and 2.2.3.2. The working R value is 21.0% using data between 6.0 to 2.3 Å and the free R value (10% of the data) is 30.3% for final model using reflections with $I/\sigma I$ >2.0. The current model includes two complexes of X11 PTB/APP peptide (246 residues for two X11 PTB domains and 20 residues for APP peptides), and 222 water molecules. No electron density is present for 14 residues in the $\alpha 1/\beta 2$ loop, 23 residues in the $\beta 6/\beta 7$ loop and 16 residues at the C-terminus of the first X11 PTB molecule, as well as for 7 residues at the N-terminus in addition to 10 residues at the C-terminus of the second X11 PTB molecule. The average

Table 2.2.3.1 Statistics for data collection

*$R_{merge}=100 \times \Sigma_h \Sigma_i |I_{h,i} - \langle I_h \rangle| / \Sigma_h \Sigma_i I_{h,i}$

+$R_{iso}=\Sigma_h |F_{nat,h}-F_{deriv,h}|/\Sigma_h F_{nat,h}$ For MAD analysis, R_{iso} is calculated between data set at two wavelengths. Wavelength 2 is used as reference data set (F_{nat}) in this calculation, and F_{deriv} refers to data at another wavelength.

Statistics of the Crystallographic Analysis: (for X11/10mer peptide complex)

Data set	Resolution (Å)	Reflections measured/unique	Completeness(%) overall/outer shell	Rmerge(%)* overall/outer shell	Riso+	Phasing Power
SIRAS Analysis						
SeMet (Seleno-methionine, in house X-ray source)	30-2.5	70368/15538	97.0/95.6	6.9/27.0		
MePbAc (trimethyl lead acetate, 2 sites, in house X-ray source)	30-2.5	113378/16169	99.8/98.7	8.3/34.7	0.22	0.88
Overall SIRAS figure of merit 0.32 (20~2.7Å)						
MAD Analysis						
SeMet (13 sites, BNL synchrotron)						
λ1(0.9770Å)	30.0-2.6	117094/14337	99.6/97.4	8.3/26.6 (14.4 at 2.9Å)	0.0453	
λ2(0.9793Å)	30.0-2.6	118270/14300	99.9/99.6	8.8/30.2 (15.6 at 2.9Å)	0.0349	
λ3(0.9797Å)	30.0-2.6	118133/14326	99.3/94.3	8.5/27.6 (14.5 at 2.9Å)	-	
Overall MAD figure of merit 0.66 (15~2.9Å)						
Data Sets Used for Refinements						
X-11/10mer peptide (in house X-ray source) SeMet	30-2.3	69525/18927	92.4/80.0	8.3/22.9		
X-11/14mer peptide (in house X-ray source) SeMet	30-2.4	97157/17790	99.3/99.2	5.4/30.2		

Table 2.2.3.2 Statistics for refinement

#Free R-factor was calculated with 10% of the data.

$Note that data with $|F| < 2\sigma F$ were not used in the refinement.

Refinement Statistics:

	X11/APP peptide(14mer)	X11/APP peptide (10mer)
Resolution (Å)	6.0-2.5	6.0-2.3
Completeness (%) (|F|>2.0 σ|F|)	82.7	69.0
R-factor(Free R[#])(%)	21.2 (30.3)	21.0 (30.3)
Completeness (%) (|F|>0 σ|F|)	93.4	91.3
R-factor(Free R[#$])(%)	23.2 (32.5)	24.5 (34.4)
numer of non-hydrogen atoms	2221	2155
numer of residues	285 (390)	266 (384)
number of waters	161	215
Rmsd bond lengths	0.010	0.010
Rmsd bond angles	1.773	1.766
Rmsd B value(bonds)	1.30	1.54
Rmsd B value(angles)	1.65	1.81

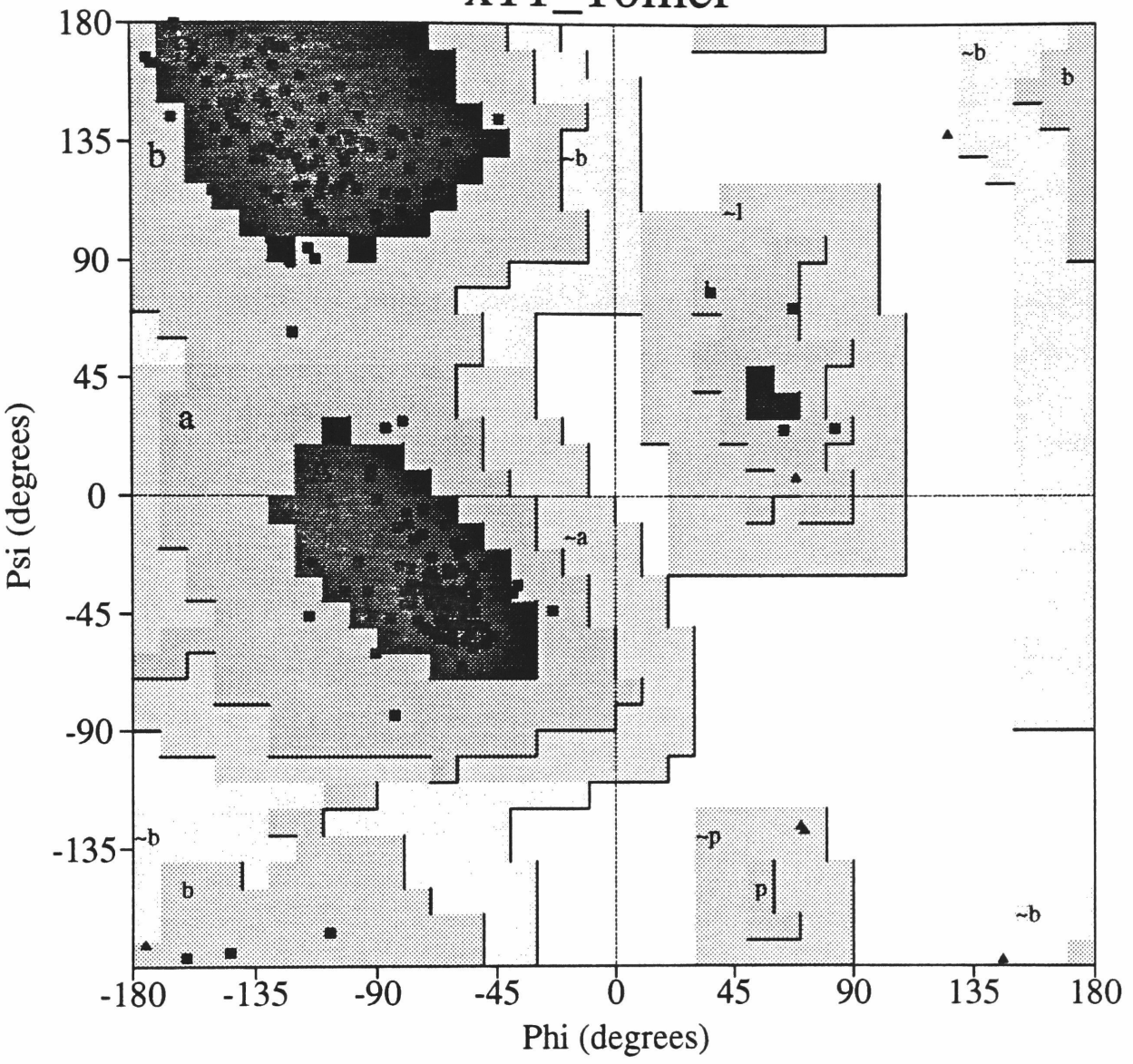
temperature factor for protein atoms is 35.6 $Å^2$. Although the two X11 PTB/APP peptide complexes are very similar (rms deviation for Cα atoms =0.9Å), certain details in the domain/peptide interfaces are slightly different.

Crystals of the X11 PTB domains in complex with the 14 residue APP peptides were also obtained, which belong to the same space group as those of X11 PTB/10 residue APP peptide complexes. The positions of the additional four peptide residues were identified by difference Fourier techniques using phases calculated from the model of X11 PTB/10 residue peptide complex. A model was built accordingly and was refined at 2.5 Å resolution using a data set collected in house. The refinement proceeded smoothly and the final model is essentially identical to the structure of the X11 PTB/10 residue APP peptide complex, apart from the additional peptide residues. 285 residues and 177 water molecules were included in the model. The working R value is 21.2% using data between 6.0 to 2.5 Å and the free R value (10% of the data) is 30.3% for final model using reflections with $I/\sigma I > 2.0$ (Table 2.2.3.2). The final model is of excellent stereochemical quality and no residue of the final model is in the disallowed regions in the Ramachandran Plot (Figure 2.2.3.1).

Figure 2.2.3.1 Ramachandran plot of the 2.3 Å refined X11/APP peptide model. The glycine residues are indicated by triangles. 90% of the residues lie in the most favored regions. No residue is in the disfavored region of the plot. The plot is made with the PROCHECK program (Laskowski *et al.*, 1993).

Ramachandran Plot

x11_10mer

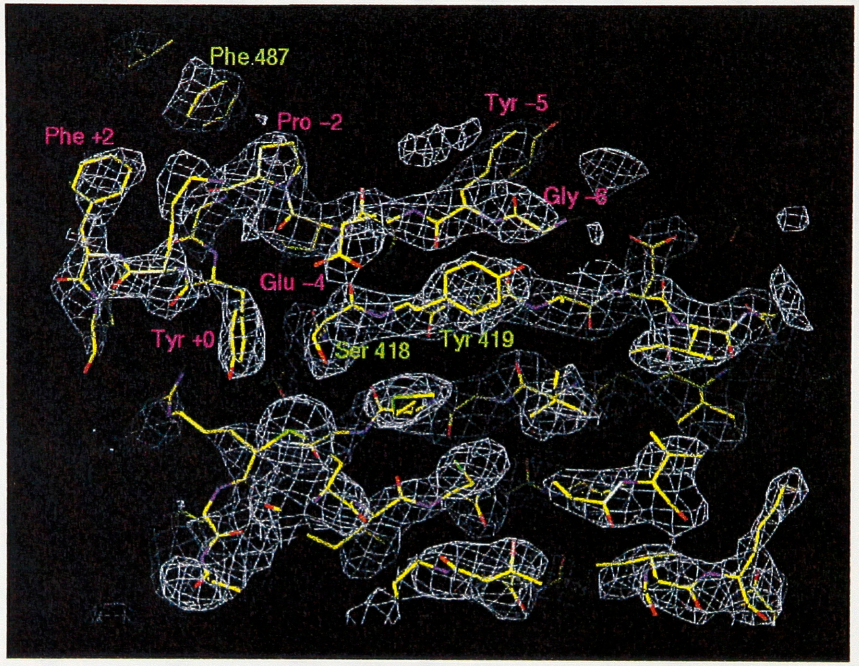


Plot statistics

Residues in most favoured regions [A,B,L]	215	90.0%
Residues in additional allowed regions [a,b,l,p]	24	10.0%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	239	100.0%
Number of end-residues (excl. Gly and Pro)	17	
Number of glycine residues (shown as triangles)	10	
Number of proline residues	5	
Total number of residues	271	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

Figure 2.2.3.2 Experimental Electron density map of X11/APP peptide complex derived from MAP analysis. The electron density map was calculated using coefficients $|F_o|\exp(-i\alpha_c)$ at 2.9 Å resolution, where F_o is the observed structure factor amplitude and α_c is the phases calculated from the MAD analysis. Grey contour lines indicate electron density at 1.2 σ. The protein (residues labeled in green) and peptide (residues labeled in red) are shown in stick figure and colored in yellow for carbon atom, blue for nitrogen atom and red for oxygen atom.


Phe.487
Tyr -5
Pro -2
Phe +2
Gly -3
Glu -4
Tyr +0
Ser 418
Tyr 419

2.3 Characterizing the Interactions between HIV-1 Nef Protein and Src Family SH3 Domains

2.3.1 Divergence Analysis and Molecular Modeling of the SH3 Domains

2.3.1.1 Sequence Comparison

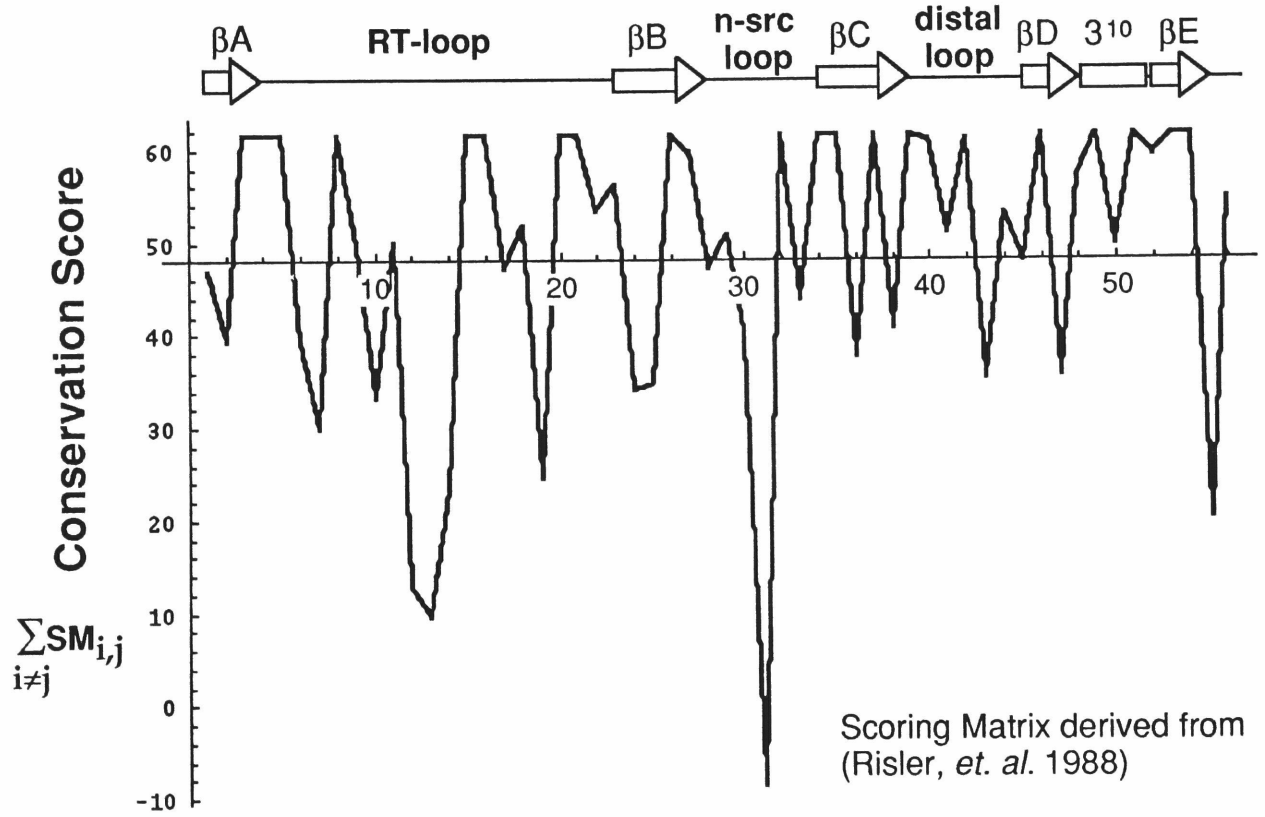
The general purpose of multiple sequence alignment and conservation analysis is to identify the residues that participate in *general* functions and to predict structure. However, in the case of the SH3 domain, both types of information are available: the structure fold is known and the residues involved in binding to PxxP peptide have been identified. However, the question remaining is which residue(s) determine the individual binding specificity of the SH3 domains. Three assumptions were made: (i) different groups of the SH3 domains have distinguishable binding specificity, (ii) differences in primary sequences of the SH3 domains give rise to their differences in binding specificity, and (iii) the specificity-determining residues are likely located near the PxxP binding surface. It follows from the above that the residues that are most polymorphic (or divergent) among different groups of the SH3 domains might be responsible for the individual specificity.

The sequence alignment including over 50 different SH3 domains was first generated using program GAP and then adjusted manually. A matrix of pairwise comparison of sequence identity

among these SH3 domains was calculated. These SH3 domains were then divided into 7 groups according to the sequence identity, their biological functions if known, and the domain structures of the parent proteins —— of which these SH3 domains are parts. Phylogenetic analysis was not performed because the evolution mode for the SH3 domains is not clear. To avoid a complicated weighting scheme, only one representative sequence was chosen from each of the eight SH3 groups (except for Fyn and Hck which both belong to the Src-family group but have different binding specificity) and were subjected to conservation analysis. Although the choice of sequences might appear arbitrary, re-analyzing different and larger sets of sequences revealed similar trends (see below).

Conservation scores, cs, were calculated for amino acid residue at each position as : $cs_n = \sum_{i \neq j} sm_{ij}$ for each residue position, n, in 8 different SH3 domains. The score matrix, sm, is derived from a previous report (Risler *et al.*, 1988) which accounts for multiple physico-chemical properties of amino acids (ex. Lys is hydrophobic as well as basic). A figure that plots cs against residue number is shown (Figure 2.3.1.1.1). As seen, the residues of RT-loop and n-Src loop are not very well conserved among SH3 domains. In particular, residues 10-14 (with sequence EAIHH in Hck SH3 domain) appear to be the most divergent. Examining the structures of various SH3 domains indicated that these residues (sequence EAIHH in Hck) are located near the peptide binding sites (see below), and thus are likely to be responsible for the differences in binding specificity among SH3 domains.

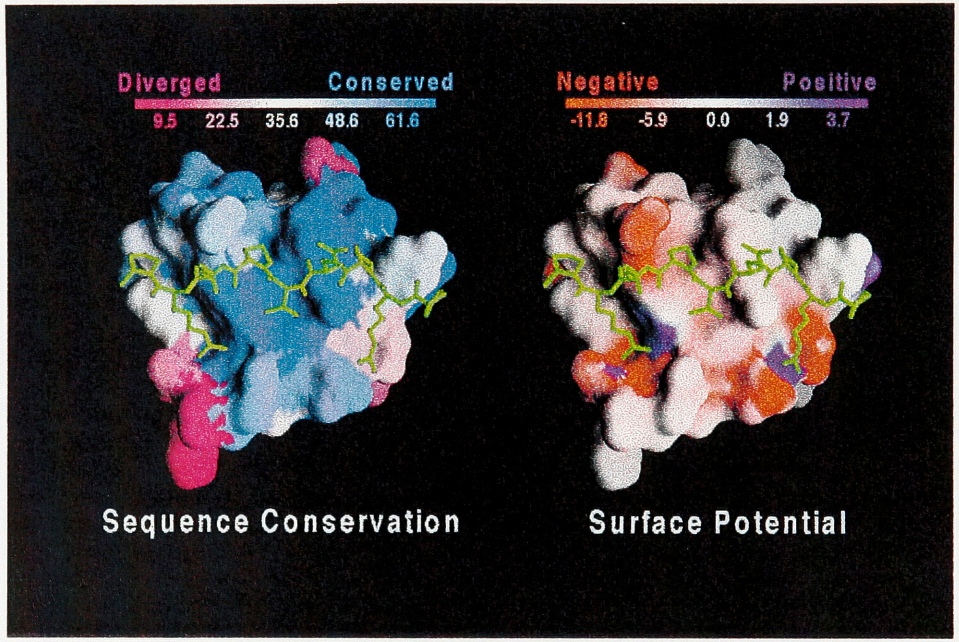
Figure 2.3.1.1.1 Conservation analysis of the SH3 domains
The conservation score (cs) is calculated (as described) for 8 different SH3 domains and plotted against residue number. Noted residues of the RT- and n-Src loops are most divergent among different SH3 domains, as indicated by their the lowest conservation score. The mean value for the overall cs is 48. The secondary structure elements of the SH3 domain are indicated as arrows (for β-strands) and as box (for 3^{10} helix) above the plot.


βA
RT-loop
βB
n-src loop
βC
distal loop
βD
3¹⁰
βE
Conservation Score
$\sum_{i \neq j} SM_{i,j}$
60
50
40
30
20
10
0
-10
10
20
30
40
50
Scoring Matrix derived from
(Risler, *et. al.* 1988)

2.3.1.2 Molecular Modeling

Crystal structures of Fyn SH3 (Noble *et al.*, 1993) and Lck SH3-SH2 domain (Eck *et al.*, 1994) and Crk SH3/C3G complex (Wu *et al.*, 1995) were used for modeling Hck SH3/peptide interaction. The Fyn-SH3 domain (PDB access number 1shf) shows the highest sequence homology (62% identity) to Hck-SH3 among the SH3 domains whose three-dimensional structures have been determined. The Fyn-SH3 structure was, therefore, used to provide the scaffold of our Hck-SH3 model —— except for the region of the n-Src loop in which the sequence of Lck SH3 domain is more similar to Hck. Three SH3 domain structures were superimposed, and residues of Fyn-SH3 that are dissimilar to Hck-SH3 were changed using program "O" (Jones *et al.*, 1991). Peptide coordinates were taken from the structure of Crk SH3 complexed with the C3G peptide (Wu *et al.*, 1995), and the residues were changed according to the Nef PxxP motif. X-PLOR (Brünger, 1988) was used to remove unfavorable van der Waals contacts and to improve the geometry by energy minimization. The resulting model was displayed using QUANTA (Molecular Simulations, Inc.). The resulting Hck SH3-Nef PxxP model resembles the SH3-PxxP structure determined previously. The most divergent sequence, IHH, is located near —— but makes no direct contact with —— the Nef PxxP motif (Figure 2.3.1.2.1).

Figure 2.3.1.2.1 Surface representation of the conservation score and the electrostatic potential of Hck SH3/Nef PxxP peptide model. The molecular surface of the Hck SH3 model (calculated using GRASP) is colored according to (a) conservation score or (b) electrostatic potential (calculated in the absence of the Nef PxxP peptide). The Nef PxxP peptide is shown in green.


Diverged
Conserved
9.5
22.5
35.6
48.6
61.6
Sequence Conservation
Negative
Positive
-11.8
-5.9
0.0
1.9
3.7
Surface Potential

2.3.2 Biochemical and Biophysical Characterization

2.3.2.1 Expression and Purification of Nef Proteins and Nef PxxP Peptide

Construction of plasmids encoding glutathione-S-transferase (GST) fused to the N-terminus of HIV-1 NL4-3-derived Nef proteins containing various mutations has been described previously (Saksela *et al.*, 1995). Expression and purification of GST-Nef-R71 fusion proteins in *Escherichia coli* strain BL21(DE3) were carried out as suggested by the supplier of the components of the GST-fusion protein expression system (Pharmacia). After elution from glutathione-sepharose beads (Pharmacia), GST-Nef fusion proteins were dialyzed against cleavage buffer (20 mM Tris pH 8.4, 150 mM KCl, 2.5 mM $CaCl_2$, 20% glycerol) overnight and subjected to thrombin cleavage (6 U of thrombin/100 mg fusion protein) for 6 h. This resulted in full-length Nef protein with an additional 6 amino acids (GSPEFT) at its amino-terminus. Cleavage was terminated by adding 1 mM phenylmethyl sulfonylfluoride (PMSF) and 10 mM DTT, followed by an overnight dialysis against buffer A (20 mM Tris pH 7.0, 50 mM KCl, 1 mM PMSF, 10 mM DTT). To separate Nef protein from released GST and from uncleaved fusion protein, the dialyzed material was loaded onto a Mono Q HR 10/10 column (Pharmacia), previously equilibrated with buffer B (buffer A with 1 mM DTT), and eluted with a linear NaCl gradient up to 1 M NaCl. Nef protein eluted out at 0.1 M NaCl. The fractions containing Nef were pooled and applied to a 20 ml of glutathione-sepharose column to remove

residual GST (in buffer B). The flow-through material was collected and precipitated with 80% saturated ammonium sulfate, resuspended in a small volume of buffer B, and applied to a Superdex 75 HiLoad gel filtration column pre-equilibrated with buffer B. Nef eluted in a single peak corresponding to the size of a monomer. The fractions containing highly purified Nef protein were concentrated with Centriprep-10 (Amicon) to 30 mg/ml in the same buffer. For each purification, 12 L of culture was typically used, and the final recovery of purified (>99% purity judged by SDS-PAGE) Nef protein was ~30 mg, as determined spectrophotometrically from their extinction coefficients ε280=48790 (experimental value ε280=44190, (Wolber *et al.*, 1992)).

cDNA encoding human Hck (RT-PCR amplified from blood cell RNA from K.S.) and Fyn ((Kawakami *et al.*, 1986); a kind gift from Avery August and H. Hanafusa, the Rockefeller University) were used as templates to amplify the corresponding SH3 domain regions (Fyn residues 86-143) using primers containing sites for EcoRI and Xho I restriction endonuclease for cloning into the corresponding sites in pGEX-1ZT (pGEX-1λT (Pharmacia) with a modified polylinker). The Hck and Fyn SH3 constructs with mutagenized RT-loops (Table 3.3.3.1) were derived by using a long 5' PCR primer extending over this region and containing the designed mutations. All SH3 constructs were confirmed by dideoxynucleotide sequencing. GST-SH3 proteins were expressed from these constructs and purified according to the protocol provided with glutathione-sepharose

(Pharmacia), and biotinylated using the ImmunoPure NHS-LC-Biotin reagent as suggested by the supplier (Pierce).

To produce Hck-SH3 domain (without GST) for ITC (isothermal titration microcalorimetry) experiments, cDNA encoding the SH3 domain of Hck was amplified by PCR using GST-Hck-SH3 construct as template. The amplified DNA fragment was subcloned into an expression vector, pET3a (Novagen) through Nde I and BamH I sites. Expression of Hck-SH3 domain using the resulting plasmid was performed according to the protocol provided by the supplier of pET expression system (Novagen). Protein was purified using standard chromatography techniques, including a DEAE Fast Flow column, a Q-Sepharose Fast Flow column, a Phenyl Sepharose (High performance) column and a Superdex75 HiLoad 16/60 column on an FPLC system (Pharmacia). The protein was concentrated to 10 mg/ml by ultrafiltration (Amicon). For each purification, 6 L of culture was typically used, and the final recovery of purified (>99% purity judged by SDS-PAGE) Hck-SH3 domain was ~12 mg.

Peptide (PVRPQVPLRPMT) was synthesized by the Biopolymer Synthesis Facility at the Rockefeller University, using conventional technology. The peptides were purified by high-performance liquid chromatography and lyophilized prior to resuspension into the buffer used in different experiments. The identity of the peptide was confirmed by mass spectroscopy and its concentration was determined by quantitative amino acid analysis.

2.3.2.2 Surface Plasmon Resonance Analysis

Surface plasmon resonance (SPR) experiments were carried out on a BIAcore Biosensor apparatus (Pharmacia Biosensor). Data analysis was performed with the interactive software (BIAevalution v2.0; Pharmacia Biosensor). Biotinylated GST-SH3 proteins were immobilized through biotin-streptavidin interaction onto a Biosensor Chip SA5 (Pharmacia Biosensor) containing pre-immobilized streptavidin. Immobilizations were performed at 25°C with a flow rate of 5 µl/min in HBS running buffer (10 mM HEPES pH 7.4, 150 mM NaCl, 3.4 mM EDTA, 0.05% Surfactant P20). The amount of biotinylated GST-SH3 fusion protein bound to the chip was monitored by the change in refractive index and controlled to be approximately 450 RU (response units). To measure the equilibrium background response signal (bulk refractive index effect), control experiments were performed using a blank cell at Nef concentration identical to those used for specific binding studies. Equilibrium background response values (~ 0~20% of total response, depending on the concentration of Nef) were subtracted from the values obtained for the specific binding reactions to yield specific response value, cRU (corrected response unit) for each concentration of Nef tested. All binding experiments were performed at 25°C, with a flow rate of 10 µl/min, in BS running buffer with 1 mM DTT. The concentration range of Nef and Nef-PA1 used for binding to various GST-SH3 fusion protein was 6 to 0.06 µM. After completion of an individual binding experiment, the chip was washed for further 20 min with running buffer (HBS with 1 mM DTT) at a flow rate of 15 µl/min, which was

found to be sufficient to regenerate the chip. Due to the extremely high-affinity interaction between biotin and streptavidin ($K_D \sim 10^{-15}$ M, this washing did not result in loss of the immobilized GST-SH3 proteins. The level of the SPR response was examined using a standard Nef solution before and after 20 programmed runs, and revealed no significant change in Nef-binding capacity of the chips.

2.3.2.3 Isothermal titration calorimetry

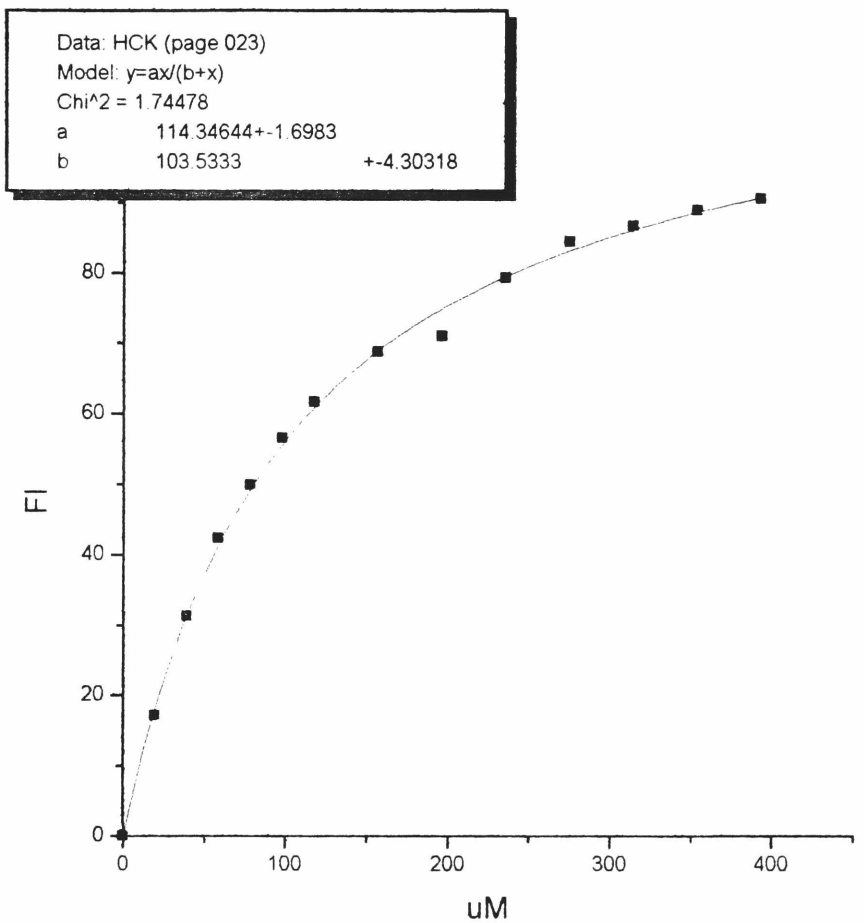
Titrations employed the OMEGA instrument (MicroCal (Wiseman *et al.*, 1989)), and were performed at 25°C in 50 mM HEPES, 150 mM NaCl, 3.4 mM EDTA, 2 mM DTT, pH 7.4. Both Hck-SH3 and Nef were dialyzed exhaustively, in the same container, against this buffer prior to titration. The titration was then performed as described (Lemmon and Ladbury, 1994) with 16 injections of 15 µl each of Hck-SH3 (at 1.09 x 10^{-4} M) into a solution of Nef (1.09 x 10^{-5} M) in the calorimeter cell (1.39 ml). The heat for dilution of the Hck-SH3 solution was determined in a separate titration of SH3 domain into buffer solution present in the cell. The heat per injection remained constant throughout this dilution, and a mean value for this heat was subtracted from heats per injection measured in the binding titration. An additional control was performed for the heat of dilution of buffer into the Nef solution, and this heat was also subtracted from experimental heats of dilution. In all cases the value for [sites]K_D (c-value) was less than 100. Titration curves were fit using ORIGIN software (MicroCal), using a nonlinear least squares algorithm based on a model for a single class of binding site, as

indicated by the shape of the titrations (Wiseman *et al.*, 1989). Stoichiometry, binding constant ($K_B = 1/K_D$) and ΔH were all allowed to float in the fitting procedure.

2.3.2.4 Tryptophan fluorescence spectroscopy

Fluorescence measurements were done at 18°C using a Perkin-Elmer 760-40 fluorescence spectrometer as described previously (Knudsen *et al.*, 1995). The wavelength for excitation was 290 nm and the wavelength for emission was 345 nm. Increasing amounts of 23 mM Nef-PxxP peptide solution were added to 1 cm square quartz fluorescence cuvettes containing 1 ml of 0.5 μM solution of the GST-SH3 domain fusion protein of interest in HBS buffer, and the equilibrium dissociation constants (K_D values) for these interactions were determined based on the observed changes in fluorescence (Figure 2.3.2.4.1). Since the concentrations of SH3 domains were low, the experimental data were fitted using the following equation: $F = F_{max} * [peptide]/([peptide] + K_D)$, where [peptide] and F are the peptide concentration and the measured protein fluorescence intensity at each point, and F_{max} is the observed maximal fluorescence intensity of the protein when saturated with the peptide. Non-linear regression curve fitting using ORIGIN (MicroCal Software, Inc.) was done to fit the experimental data to the above equation. In the case of Hck, purified SH3 domain cleaved free of GST was also similarly tested, and showed binding kinetics very similar to those observed using the GST-Hck-SH3 fusion protein.

Figure 2.3.2.4.1 Binding of Nef PxxP peptide to Hck SH3 domains. The response function shows observed changes in fluorescence intensity (y-axis) after various amounts of Nef PxxP peptide (x-axis, total concentration) was added to the solution containing Hck SH3 domains. Insert: the curve was fit to the equation, $y=ax/(b+x)$ where x and y are the peptide concentration and the measured protein fluorescence intensity at each point, and a is the observed maximal fluorescence intensity of the protein when saturated with the peptide. b is the equilibrium dissociation constant, which was estimated to be 103.5 μM.


Data: HCK (page 023)
Model: y=ax/(b+x)
Chi^2 = 1.74478
a 114.34644+-1.6983
b 103.5333 +-4.30318
FI
0
20
40
60
80
0
100
200
300
400
uM

2.3.2.5 Circular Dichroism (CD) spectroscopy

CD measurements were done using a AVIV 62 DS Circular Dichroism Spectrometer, and data collection and processing were performed using the program provided by the manufacturer. Experiments using 50 μM protein in HBS buffer containing 1 mM DTT were carried out at three different temperatures (4, 15 and 25 °C) using a 0.1 mm pathlength cuvette. Spectra were recorded in wavelength range from 195 to 240 nm with 0.5 nm spacing, and values from 5 individual measurements were averaged and expressed as molar mean residue ellipticity values as a function of the wavelength.

2.3.2.6 Co-precipitation assay

To examine binding of wild-type and mutant Nef proteins to different GST-SH3 fusion proteins in solution, 1 to 4 μg of each GST-SH3 fusion protein was immobilized on glutathione agarose beads, and incubated at 4°C for 20 min with various concentrations of Nef in 20 mM HEPES pH 7.5, 10% glycerol, 1% Triton X-100, 150 mM NaCl, 1 mM DTT (in a total volume of 1 ml). After a 20 min incubation at 4°C, the beads were washed 3 times with the same buffer, followed by electrophoretic analysis in a 15% SDS/polyacrylamide gel. Bound Nef protein as well as the GST-SH3 fusion protein were visualized by Coomassie blue staining. Peptide inhibition assays were done in an identical manner, except that, in

addition to Nef, a 2- to 800-fold molar excess of the peptide PVRPQVPLRPMT was also included in the binding reaction.

2.4 Structure Determination of HIV-1 Nef-FYN(I) SH3 Complex

2.4.1 Identification of the Nef Domain Structure and Optimization of Crystal Quality

Sequence comparison and secondary structure prediction of various Nef sequences suggested that Nef protein might contain two domains: a N-terminal anchoring domain (residues 1-57 in NL4-3) and a C-terminal core domain (residues 58-203 in NL4-3). Certain proteases, such as HIV protease, can cleave the Nef protein and separate these two domains *in vitro*. Apart from the myristoylation site at the N-terminus, the anchoring domains from different virus strains have divergent amino acid sequences and often contains large insertions/deletions or duplication. In contrast, except for one region (residues 148-179), the C-terminal domain is rather conserved among immunodeficiency viruses and has been predicted to contain a folded structure (Shugars *et al.*, 1993; Freund *et al.*, 1994).

Large quantities of the full-length Nef can be produced by first expressing it as a GST fusion protein and then cleaving with thrombin to remove GST (refer to Section 2.3.2.1). The full-length Nef was, however, proven to be resistant to crystallization —— in presence or absence of SH3 domain. I therefore sought a fragment of Nef that has a defined structure by a combination of limited proteolytic digestion, deletion/site-directed mutagenesis, and crystallization trials. The full-length Nef protein subjected to mild

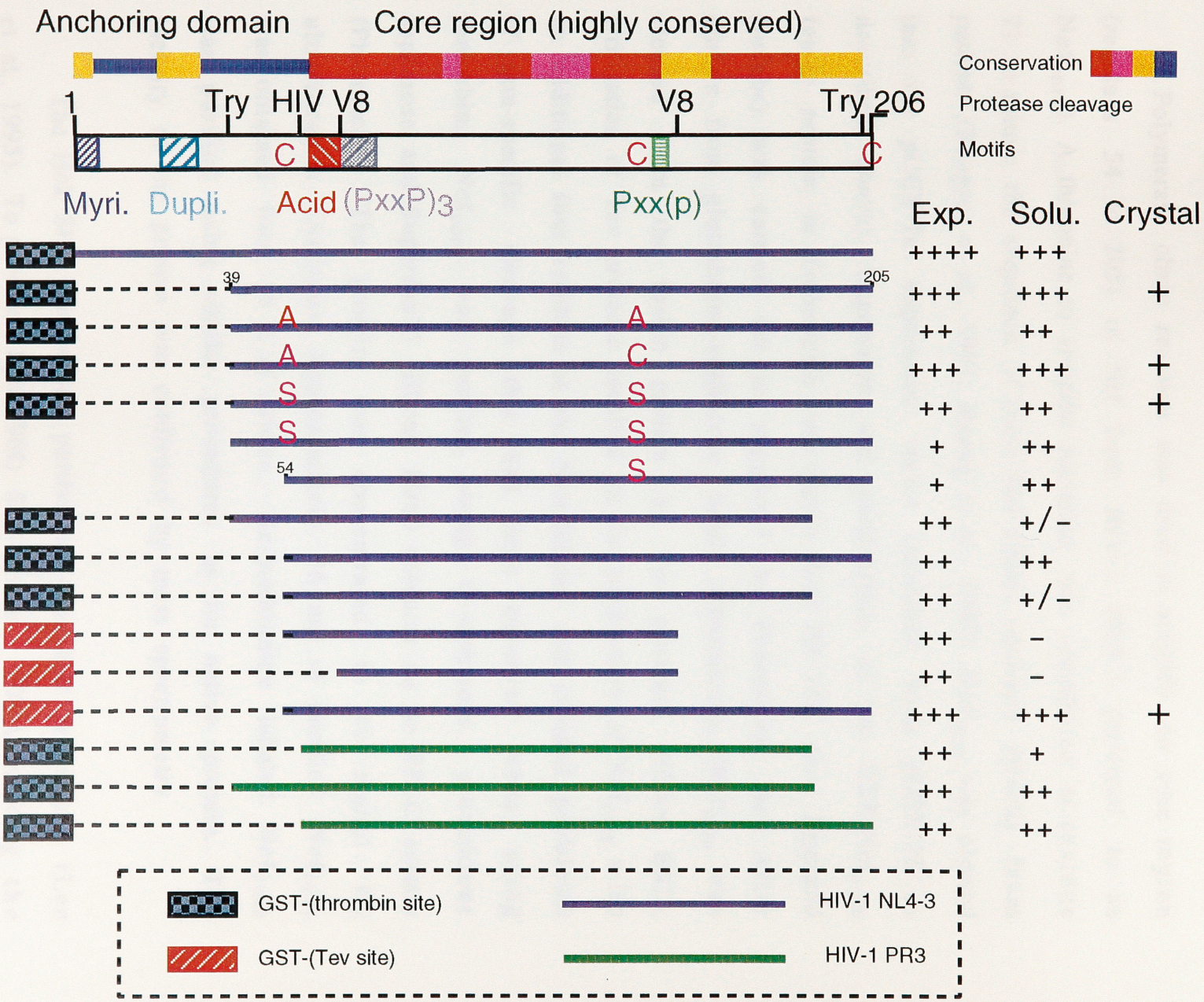
trypsin digestion resulted in a stable fragment corresponding to residues 39-205. This fragment of Nef (which still binds to the Hck SH3 domain with high affinity and specificity) is rather temperature sensitive; visible precipitation occurred after being exposed to moderate heat (< 37°C) during material handing. This fragment of Nef crystallized at 4°C in a solution containing MPD as precipitant 3 months after the setup. However, I was still not able to reproduce the crystals after exhaustive trials. Examining the protein preparation by SDS-PAGE at the non-reducing condition indicated around 10% of the protein formed dimers through intermolecular disulfide bonds. To avoid the formation of intermolecular disulfide bonds, each or both of the two cysteines (residues 55 and 142) in the Nef fragments was then mutated to alanine (or serine). The mutant Nef proteins were examined for dimer formation. The Cystine 55 appears to be solely responsible for the disulfide bond formation. Mutant proteins (C55A and C55S) were then set up for crystal trial. However, no crystal was yielded in the absence of the SH3 domain. Nevertheless, including the Hck SH3 domain readily yielded large crystals within a week. These crystals, however, only diffracted to around 6-8 Å resolution and can not be improved.

Examining the fragments of Nef generated by various low-stringency proteases (including chymotrypsin, subtilisin and V8) indicated that proteolytic cleavage sites concentrated in two regions: the N-terminal domain (residues 1-58 in NL4-2 sequence) and a region (residues 148-180) in the middle of the C-terminal core domain. According to the limited proteolysis results, a series of Nef

fragments with various lengths was generated (Tables 2.4.1.1). It was noted that the expression of Nef as a GST-fusion protein is necessary since fragments expressed using the pET system (without GST) were mostly insoluble. However, when expressing smaller fragments of Nef as GST-fusion protein, the thrombin proteolytic cleavage (to remove GST) became extremely inefficient. It was suspected that the thrombin cleavage sites become less accessible due to steric hindrance of two protein domains, GST and Nef core. Prolonged exposure of GST-Nef fragment with thrombin resulted in nonspecific cleavage and degradation of Nef fragment. To solve these problems, a specific vector, pGEX-Tev (modified form pGEX-2T, Pharmacia) was engineered so that a linker and a TEV (tobacco etch virus) protease cleavage site were placed after the GST structure gene. The TEV protease (used to remove GST from the fusion protein) was proven to be very specific and reliable. Furthermore, unlike thrombin, TEV protease is resistant to reducing reagent (such as DTT). Smaller fragments of Nef can be produced using this vector with little degradation and oxidation.

One Nef construct (residues 54-205), corresponding to the highly conserved core domain of Nef, produced crystals with the Hck SH3 domains. These crystals, however, still diffract poorly (around 5-6 Å). Nevertheless, cocrystallization with the Fyn(R96I) SH3 domain (to which Nef binds tightly) resulted in large hexagonal crystals that diffracted to 2.8 Å resolution using an in-house X-ray source. These crystals were used for the structure determination (described below).

Figure 2.4.1.1 Nef constructs for crystallization trials. (Top) The degree of conservation among different strains of HIV is indicated by solid boxes with red being most conserved and blue most divergent. Protease cleavage sites are also shown (Try, trypsin; HIV, HIV protease; V8, V8 protease). The conserved motif found in most immunodefficient viruses are indicated by striped boxes (Myri, myristolation site; Dupli., duplication region; Acid, acidic region; (PxxP)3, tetraPxxP motif; Pxx(p), a less conserved Pxxp motif). Three Cysteine in the NL4-3 Nef are shown in red letter "C" and point mutations (Cys-to-Ala or Cys-to-Ser) are also indicated. The expression level, solubility and crystallization for each construct are indicated by plus signs ("+"). Nef expressed as GST fusion proteins is indicated by a striped box. The Nef genes from two different strains of HIV-1, NL4-3 (colored in purple) and PR3 (in green) are used.


Anchoring domain
Core region (highly conserved)
Conservation
Protease cleavage
Motifs
1
Try
HIV V8
V8
Try 206
C
C
C
Myri.
Dupli.
Acid (PxxP)3
Pxx(p)
Exp.
Solu.
Crystal
39
205
54
A
A
A
C
S
S
S
S
S
S
++++ +++
+++ +++ +
++ ++
+++ +++ +
++ ++ +
+ ++
+ ++
++ +/−
++ ++
++ +/−
++ −
++ −
+++ +++ +
++ +
++ ++
++ ++
GST-(thrombin site)
HIV-1 NL4-3
GST-(Tev site)
HIV-1 PR3

2.4.1.1 Protein Expression and Purification

Polymerase chain reaction was used to amplify the core region (residues 54 to 205) of Nef from HIV-1 NL4-3 (referred to as Nef_{core}). A threonine to arginine mutation was introduced at residue 71 to mimic the sequence of most Nef alleles obtained directly from patients (Shugars et al., 1993; Huang et al., 1995). Nef_{core} was cloned into the pGEX-Tev expression vector (modified from pGEX-2T as described above). Expression and purification of the GST-Nef_{core} fusion protein in *Escherichia coli* (strain K12 PR 745, New England Biolabs) was carried out as suggested by Pharmacia, Inc. After elution from glutathione-sepharose beads (Pharmacia), Nef_{core} was cleaved from the fusion protein by Tev protease (Gibco BRL). Utilization of Tev protease resulted in the production of Nef_{core} with an additional four residues at the N-terminus, and avoided problems of non-specific cleavage that had been observed when using thrombin. Nef_{core} was purified using Q-sepharose, glutathione sepharose and Superdex75 HiLoad 16/60 columns on an FPLC system (Pharmacia). The protein was concentrated to 40 mg/ml by ultrafiltration (Amicon). Approximately 36 mg of purified Nef_{core} was obtained from 18 L of culture. Se-methionine labeled Nef_{core} was purified using similar procedures as for native protein. The identity of the protein was confirmed by mass spectrometry.

The Hck SH3 domain was purified as described previously (Lee et al., 1995). To produce Fyn(R96I) SH3 domain, cDNA encoding the SH3 domain of Fyn(R96I) was amplified by PCR using the

pGEX-Fyn(R96I) construct (refer to Section 2.3.2.1) as template. The amplified DNA fragment was subcloned into an expression vector, pET3a (Novagen), through NdeI and BamHI sites. Expression of the Fyn(R96I) SH3 domain using the resulting plasmid was performed according to the protocol provided by the supplier of the pET expression system (Novagen). Protein was purified using standard chromatography techniques, including a DEAE Fast Flow column, a Q-sepharose Fast Flow column, a Phenyl Sepharose (high performance) column and a Superdex75 HiLoad 16/60 column on an FPLC system (Pharmacia). The protein was concentrated to 20 mg/ml by ultrafiltration (Amicon). For each purification, 12 L of culture was typically used and the final recovery of purified Fyn(R96I) SH3 domain was ~24 mg.

Nef_{core} and Fyn(R96I) SH3 were mixed in a 1:1 molar ratio and dialyzed against a buffer containing 10 mM Tris (pH 7.5), 50 mM KCl and 2 mM of DTT. Crystallization conditions were scanned using the hanging drop method. Crystals of the complex (approximate dimensions 0.2 X 0.2 X 0.1 mm^3) were obtained overnight with ammonium sulfate as the precipitant (1 M) and Tris buffer (100 mM, pH 9.0), at 4° C. The crystals are hexagonal ($P6_522$; a=107.8, c=228.5) and diffract to Bragg spacings of 2.8 Å using an in-house X-ray source.

2.4.2 Crystallographic Methods

2.4.2.1 Data Collection

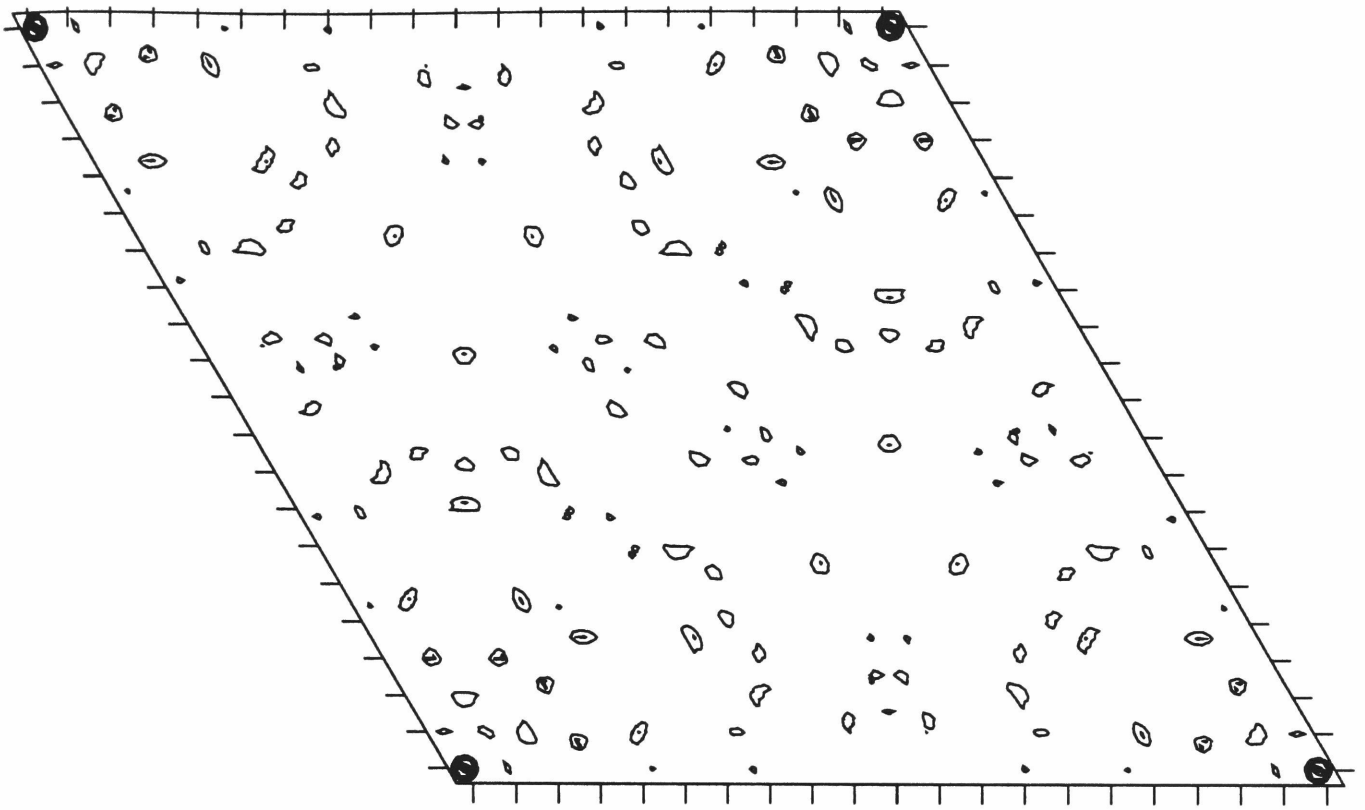
All diffraction data were measured from crystals cooled to 100 K. Crystals were flash-frozen in a stream of N_2 (for in-house data collection) or in liquid propane (for data collection at CHESS and BNL synchrotrons) after soaking serially in mother liquor (1.2 M ammonium sulfate, 100 mM Tris buffer, pH 9.0) containing 5% to 30% glycerol. X-ray data for MIR analysis were collected on a Rigaku RAXIS IIC area detector mounted on a Rigaku RU200 X-ray generator (Molecular Structure Corporation, USA) and the A1 beamline at Cornell High Energy Synchrotron Source (CHESS). MAD data were collected from a single frozen crystal derivatized with 30 mM of trimethyl lead acetate at the National Synchrotron Light Source, Brookhaven National Laboratory, using beamline X4A. All data were processed using program DENZO and SCALEPACK (Z. Otwinowski and W. Minor, unpublished results).

2.4.2.2 Heavy Atom Derivatives

Although several heavy atom derivatives were obtained, severe non-isomorphism between native and derivative crystals prevented the determination of accurate phases by multiple isomorphous replacement. Crystals soaked in solutions containing 30 mM of trimethyl lead acetate yielded a derivative with a strong anomalous diffraction signal (9 σ peaks in the Harker section of an

Figure 2.4.2.2.1 (a) Isomorphous difference Patterson map (b) Anomalous difference Patterson map. Heavy atom Patterson functions for the $(CH_3)_3Pb$ derivative of Nef_{core}:Fyn(I) complex using (a) isomorphous differences; (b) anomalous differences. The plot is drawn for the Harker section (w=0.5) with contour level starting at 2σ and increment of 1σ.

PB_isomorphous_Patterson.map w=0.5



PB_anomalous_Patterson.map w=0.5

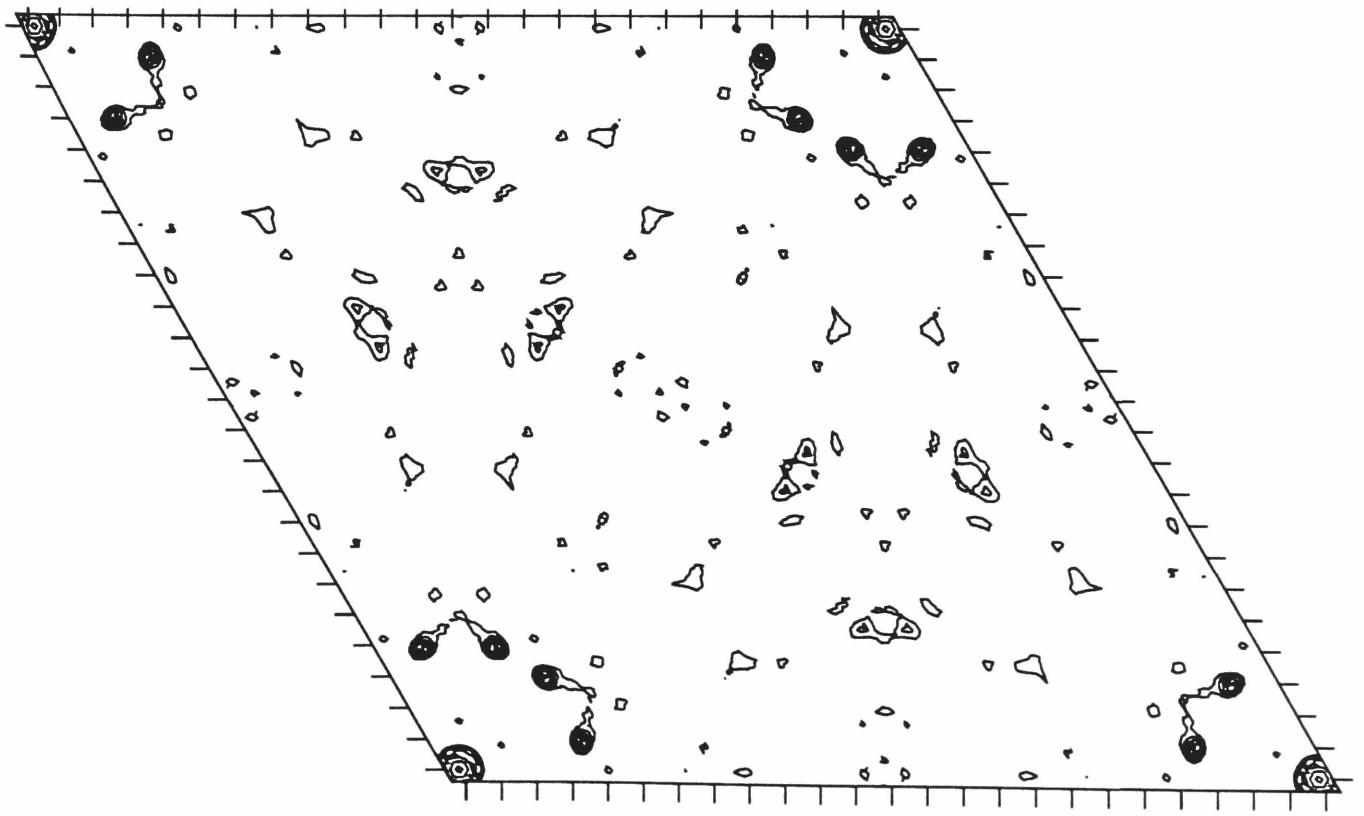
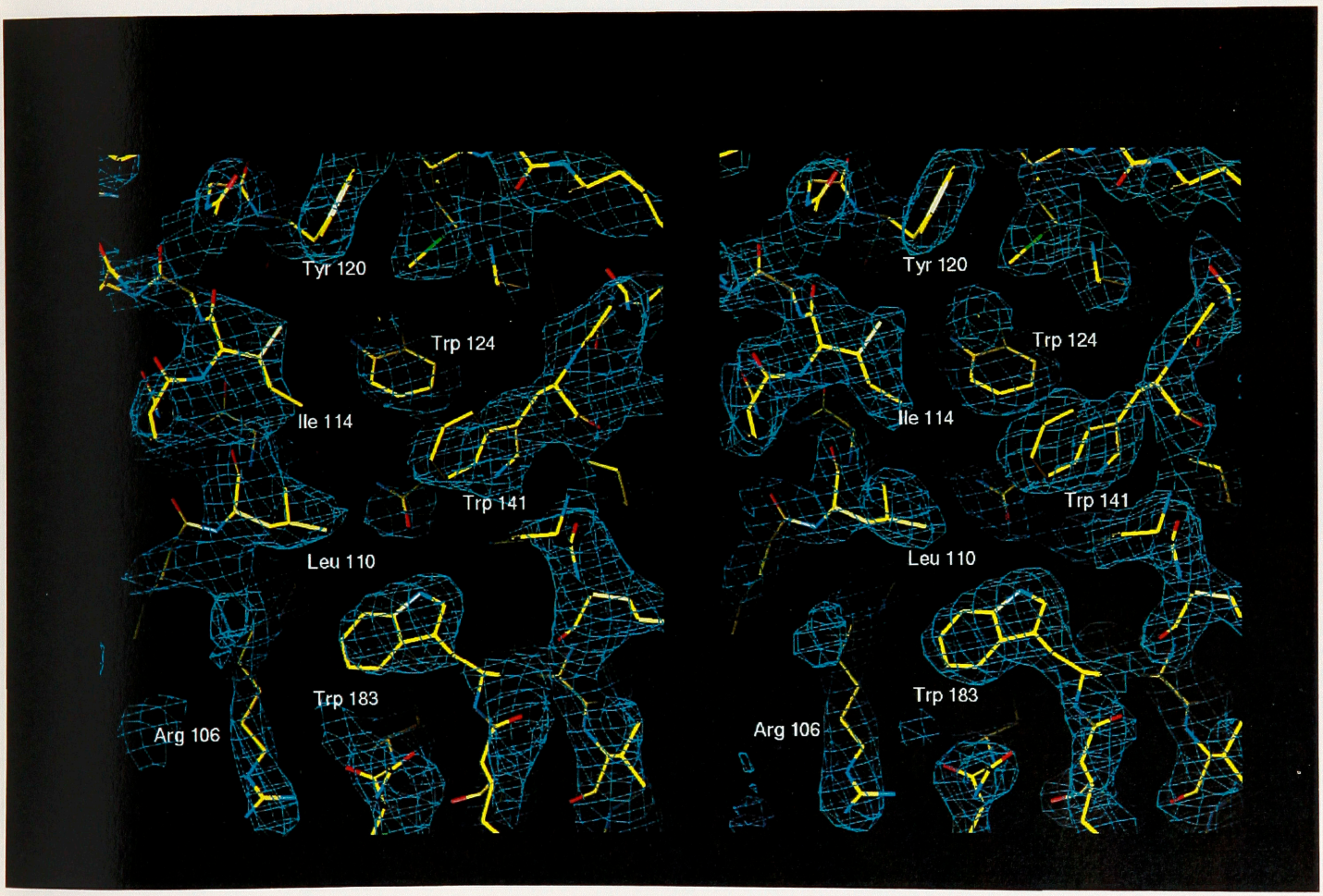


Figure 2.4.2.3.1 Electron density maps in a region of the hydrophobic core of Nef_{core}. (a) Electron density map at 3.2 Å resolution, calculated using experimental MAD phases. (b) Electron density map at 2.5 Å resolution, using phases calculated from refined model. In (a), the electron density map is calculated using experimentally observed amplitude and phases obtained from the MAD analysis and then modified by solvent leveling and two-fold real space density averaging. In (b), the electron density map was calculated using coefficients $(2|F_{obs}|-|F_{cal}|)\exp(-i\alpha_c)$, where F_{obs} is the observed structure factor amplitude and F_{cal} and α_c are the amplitude and phase calculated from the current model. Blue lines indicate electron density at the 1.5σ level. The protein atoms are shown in stick representation.


Tyr 120
Trp 124
Ile 114
Trp 141
Leu 110
Trp 183
Arg 106
Tyr 120
Trp 124
Ile 114
Trp 141
Leu 110
Trp 183
Arg 106

anomalous difference Patterson map at 3.0 Å resolution, calculated using data measured at the A1 beamline at CHESS, at a wavelength of 0.92 Å). Isomorphous difference Patterson maps contained no interpretable peaks, indicating non-isomorphism (Figure 2.4.4.2.1).

2.4.2.3 MAD phasing and Density Modification

Phases were determined by carrying out a multi-wavelength anomalous diffraction (MAD) experiment on the lead derivative. For MAD data collection, a single crystal was derivatized using trimethyl lead and was mounted, after flash freezing, with the c* axis aligned along the rotation axis and perpendicular to the X-ray beam to allow the near simultaneous recording of Bijvoet pairs of reflection. Data collection at four wavelengths near the Pb absorption edge (Table 2.4.2.4.1) were recorded on image plates, digitized with a Fuji scanner and processed with DENZO and SCALEPACK. Heavy-atom parameters were refined and phases were calculated using the program MLPHARE. (Z. Otwinowski). Phases calculated at 3.2 Å resolution using only MAD data yielded an interpretable electron density map which was further improved by solvent flattening and histogram matching, using SQUASH (Zhang and Main, 1990) (Figure 2.4.2.3.1). The handedness of the α helices provided definite confirmation of the space group assignment.

2.4.2.4 Model Building and Refinement

There are two Nef_{core}:SH3 complexes in the asymmetric unit of the crystal. A non-crystallographic two-fold symmetry operator that relates two complexes in the asymmetric unit was identified by the positions of equivalent helices and heavy atoms. It was evident from the original electron density map that the relative position of the SH3 domain with respect to Nef_{core} is slightly different between the two complexes. Two symmetry operators that relate the two Nef and SH3 molecules, respectively, were therefore used for all subsequent real space density averaging (RAVE, part of the "O" package) as well for the application of noncrystallographic symmetry restrains during refinement. The two-fold averaged electron density map (Figure 2.4.2.3.1) based on MAD phases was of good quality and allowed a model to be built unambiguously, using program "O" (Jones et al., 1991). The Fyn SH3 domain structure (Noble *et al.*, 1993) was fit into the averaged electron density map. The sequence assignment for Nef_{core} was quite straightforward since it has a relatively high aromatic content (12.5%, compared 8.4% on average). The positions of two methionine residues in the asymmetric unit (Met 79) were confirmed in a difference electron density map using X-ray data for a Se-methionine derivative (Met 173 is in a disordered region of the structure and is not observed).

Crystals that have been derivatized using trimethyl lead diffract somewhat better than unmodified crystals, and data collected for the lead derivative were used in all the analysis. The

trimethyl lead molecules are bound at crystal contact sites and are distant from the Nef:SH3 interface. The model was refined using X-PLOR (Brünger, 1992) initially using data collected during the MAD analysis. The free R value (Brünger, 1993) was used to monitor all stages of the refinement. Non-crystallographic symmetry restraints (separately for Nef_{core} and for the SH3 domain) were initially applied to the two complexes except for a small number of residues at the interface between Nef_{core} and SH3. The refinement was later extended to 2.5 Å using data collected at CHESS (Table 2.4.2.4.1). Finally, the non-crystallographic symmetry restraints were released and two complexes were allowed to refine independently. Well-ordered solvent molecules were included at this stage, and tightly restrained individual isotropic B-factors were refined. The refinement proceeded smoothly, without the utilization of simulated annealing refinement.

The statistics for data collection, phase determination and refinement are given in Table 2.4.4.4.1. The working R value is 21.5% using data between 6.0 to 2.5 Å (20684 reflections) and the free R value (10% of the data) is 28% for final model using reflections with $I/\sigma I$ >2.0. The current model includes two complexes of Nef_{core}: Fyn(R96I) SH3 domain (residues 71 to 148 and residue 179 to 203 of Nef and residues 85 to 131 of Fyn(R96I) SH3 domain), 99 water molecules and two trimethyl lead molecules. No electron density is present for 16 residues at the N-terminus of Nef_{core}, 29 residues in a loop on the side of Nef_{core} distal to the SH3 binding surface (residues 149 to 177), and for 2 residues at the C-terminus of Nef_{core}. Residues

85 to 131 of the Fyn (R96I) SH3 domain are included in the model. The average temperature factor for protein atoms is 34 $Å^2$. Although the two Nef_{core}:SH3 complexes are similar in general terms, their mutual orientations are slightly different (see main text). One of the SH3 domains exhibits higher temperature factors than the other components of the molecular model (the average temperature factor for this SH3 domain is 53 $Å^2$, compared to 30.0 $Å^2$ for the other SH3 domain and 31 and 28 $Å^2$ for the two Nef_{core} molecules). The complex with the lower temperature factors is more reliable, and is used for all analysis, unless stated otherwise. The final model is of excellent stereochemical quality and all residues lie in the most favored or additional allowed region (Figure 2.4.2.4.1).

Table 2.4.2.4.1 Statistics for data collection and refinement

*$R_{merge}=100 \times \Sigma_h \Sigma_i |I_{h,i} - \langle I_h \rangle| / \Sigma_h \Sigma_i I_{h,i}$

+$R_{iso}=\Sigma_h |F_{nat,h}-F_{deriv,h}|/\Sigma_h F_{nat,h}$ For MAD analysis, R_{iso} is calculated between data set at two wavelengths. Wavelength 2 is used as reference data set (F_{nat}) in this calculation, and F_{deriv} refers to data at another wavelength.

#Free R-factor was calculated with 10% of the data.

$Note that data with $|F| < 2\sigma F$ were not used in the refinement.

Table I Data collection and refinement statistics

Statistics of the crystallographic analysis

Data set	Resolution (Å)	Reflections measured/unique	Completeness(%) overall/outer shell	Rmerge(%)* overall/outer shell	Riso
MAD analysis					
MePbAc (trimethyl lead acetate, 2 sites, BNL synchrotron)					
λ1(0.9552Å)	15.0-3.1	153765/14703	99.6/100.	6.8/23.2	0.0377
λ2(0.9497Å)	15.0-3.1	153710/14701	99.6/100.	6.5/23.6	-
λ3(0.9473Å)	15.0-3.1	153506/14610	99.0/99.8	7.2/25.0	0.0356
λ4(0.9428Å)	15.0-3.1	153753/14696	99.6/100.	7.0/24.6	0.0369
Overall MAD figure of merit, 0.34 (15-3.2Å)					
Native	30-3.2	55790/10094	74.3/52.0	5.7/17.2	
SeMet (Seleno-methionine, 2 sites, CHESS synchrotron)	30-3.2	268848/13028	97.2/98.5	8.0/30.8	
MePbAc (trimethyl lead acetate, 2 sites, CHESS synchrotron)	20-2.5	565661/26464	94.2/84.9	6.8/35.8	

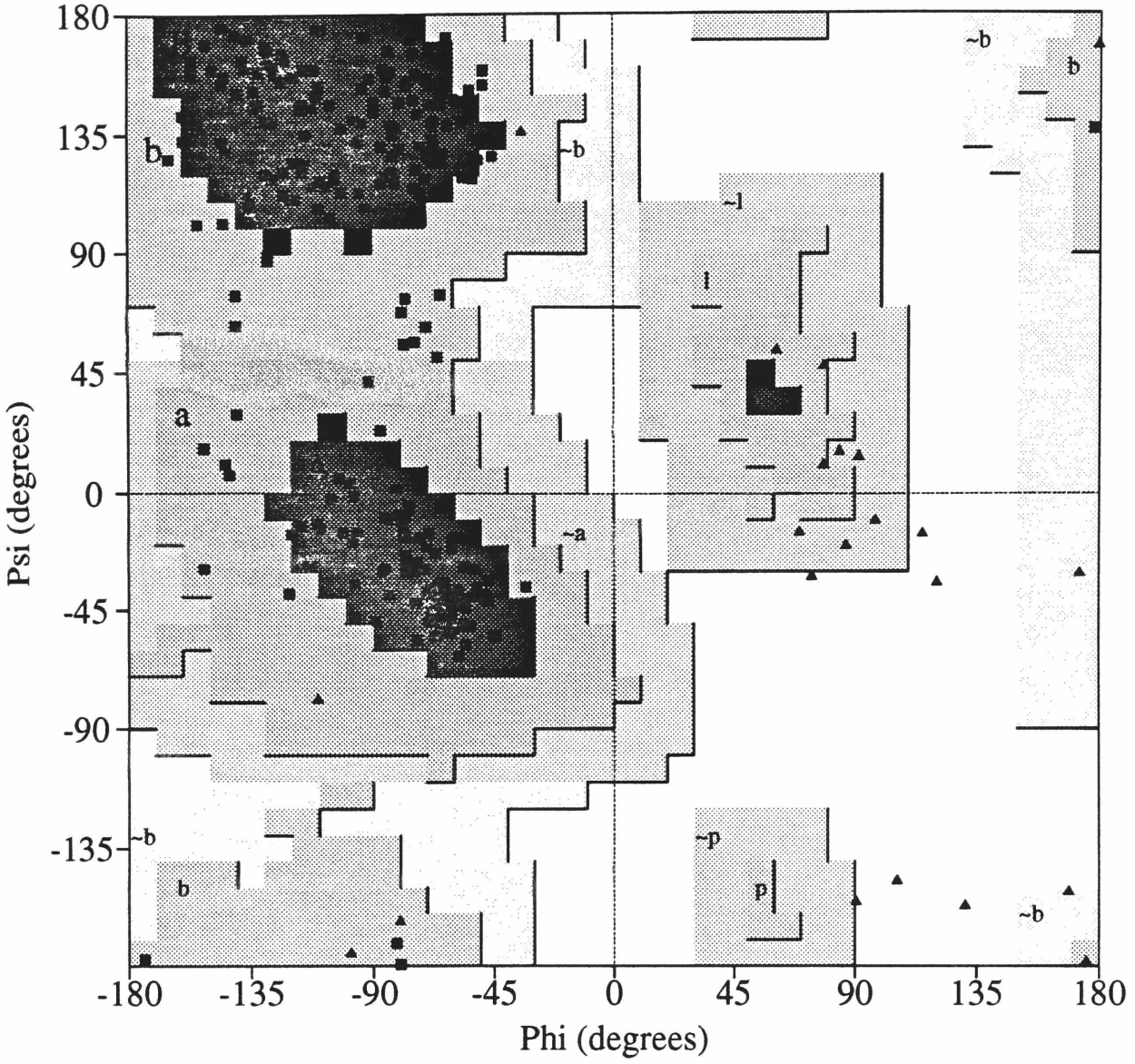
Refinement statistics

	Resolution(Å)	Completeness(%)	R-factor	Free R-factor	No. reflections
Data with F>2 σF	6.0-2.5	80.1	0.215	0.282	20684
Data with F>0 σF	6.0-2.5	90.2	0.245 (all data)	-	23299
R.m.s. deviations	bonds lengths, 0.010 Å		Bond angles, 1.5 Å	B-values, 1.4 Å^2	

Figure 2.4.2.4.1 Ramachandran plot of the 2.5 Å refined Nef$_{core}$/FYN(I) model. The glycine residues are indicated by triangles. Over 90% of the residues lie in the most favored regions. No residue is in the disfavored region of the plot. The plot is made with the PROCHECK program (Laskowski *et al.*, 1993).

ıamachandran Plot

nef



Plot statistics

Residues in most favoured regions [A,B,L]	240	90.2%
Residues in additional allowed regions [a,b,l,p]	26	9.8%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	266	100.0%
Number of end-residues (excl. Gly and Pro)	12	
Number of glycine residues (shown as triangles)	22	
Number of proline residues	22	
Total number of residues	322	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

3. Results

3.1 Phosphopeptide Recognition by SH-PTP2 SH2 domain

In order to further understand the peptide recognition mechanism of the SH2 domains, four structures of the amino-terminal SH2 domain of SH-PTP2 have been determined at the resolution between 2~3 Å. These structures include (i) two high-affinity peptide complexes (PDGFR-1009 and IRS1-895), (ii) a non-specific peptide (PDGFR-740) complex and (iii) the uncomplexed form of the SH-PTP2 SH2 domain. The structure of the PDGFR-1009 peptide complex has been refined to highest resolution (2 Å). In addition, the interaction of this peptide with the SH-PTP2 SH2 domain is likely to be physiologically relevant since the sequence surrounding pTyr1009 of PDGFR has been shown to mediate the interaction of PDGFR with SH-PTP2 *in vivo*. Because of its high resolution and biological relevance, this structure is used for the following analyses unless stated otherwise. Nevertheless, the SH2 domains in the four determined structures are essentially identical except for the phosphotyrosine binding and the BG loops which are disordered in the absence of peptide (see below).

3.1.1. General Architecture

The overall structural fold of the SH2 domains which was first described for Abl and p85 SH2 domains is a compact α/β structure comprised of around 100 residues (Figure 3.1.1.2 for sequence

alignment). This architecture is conserved among all SH2 domains. To facilitate the discussion, a notation of secondary structure which was first introduced by Eck *et. al.* for Lck SH2 domain is applied to the SH-PTP2 SH2 domain described here (see Figure 3.1.1.1 for a schematic diagram and the notation used). The central scaffold of SH2 domain is an antiparallel β-sheet formed by strands A, B, C, D and G with two α-helices, αA and αB, flanking the opposite sides of the sheet. The hydrophobic core of the SH2 domain, which holds the overall structure together, is formed between the central β-sheet and two helices.

The central β-sheet runs perpendicular to the peptide-binding surface and divides the domain into two functionally distinct regions. One region comprising helix αA, loop BC (the phosphate-binding loop), and the adjacent face of the central β sheet provides residues that interact with the phosphotyrosine. The other region includes helix αB, loops EF and BG, a small β-sheet (formed by strands βD', βE and βF), and the other face of the central β-sheet. This region interacts with peptide residues immediately following the phosphotyrosine and accounts for the sequence-specific recognition. With respect to the Src SH2 domain, the SH-PTP2 SH2 domain contains an insertion in BG loop which contains two short β-strands that form antiparallel backbone hydrogen-bonds with each other. This insertion, as also seen in the structures of the SH2 domains of p85 and PLC-γ1, provides important interactions with the peptide (see below).

Figure 3.1.1.1 Schematic diagram of the SH-PTP2 SH2 domain in complex with a peptide from the insulin receptor substrate-1 (IRS-1). Viewed from the peptide binding surface, illustrating the secondary structure elements and the notation used. The phosphotyrosine (at right), Val(+1), Asn(+2), Ile(+3), Glu(+4) and Phe(+5) (at left) of the peptide are shown in a ball and stick representation. α helices and β strands are shown as ribbons and arrows, respectively.

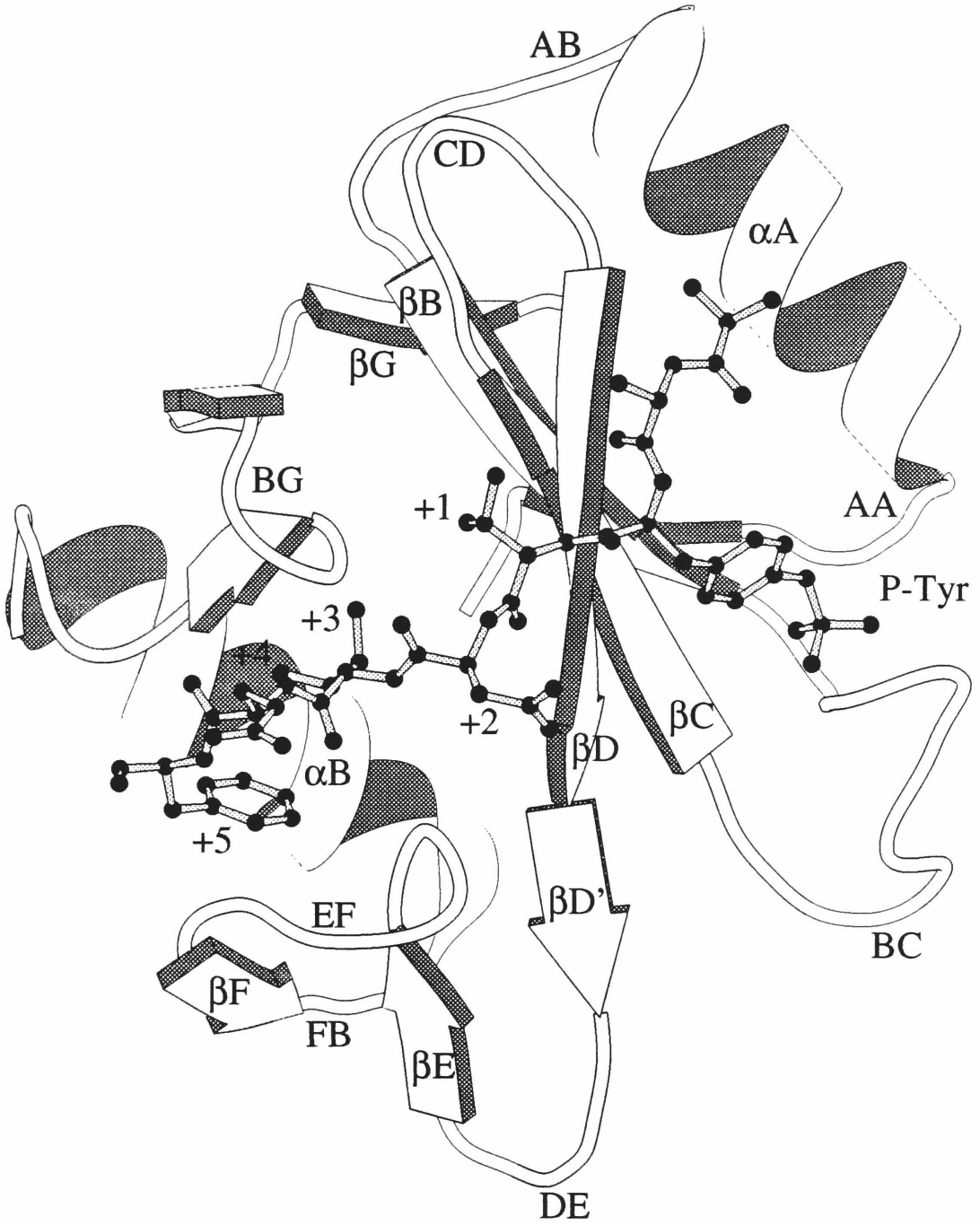
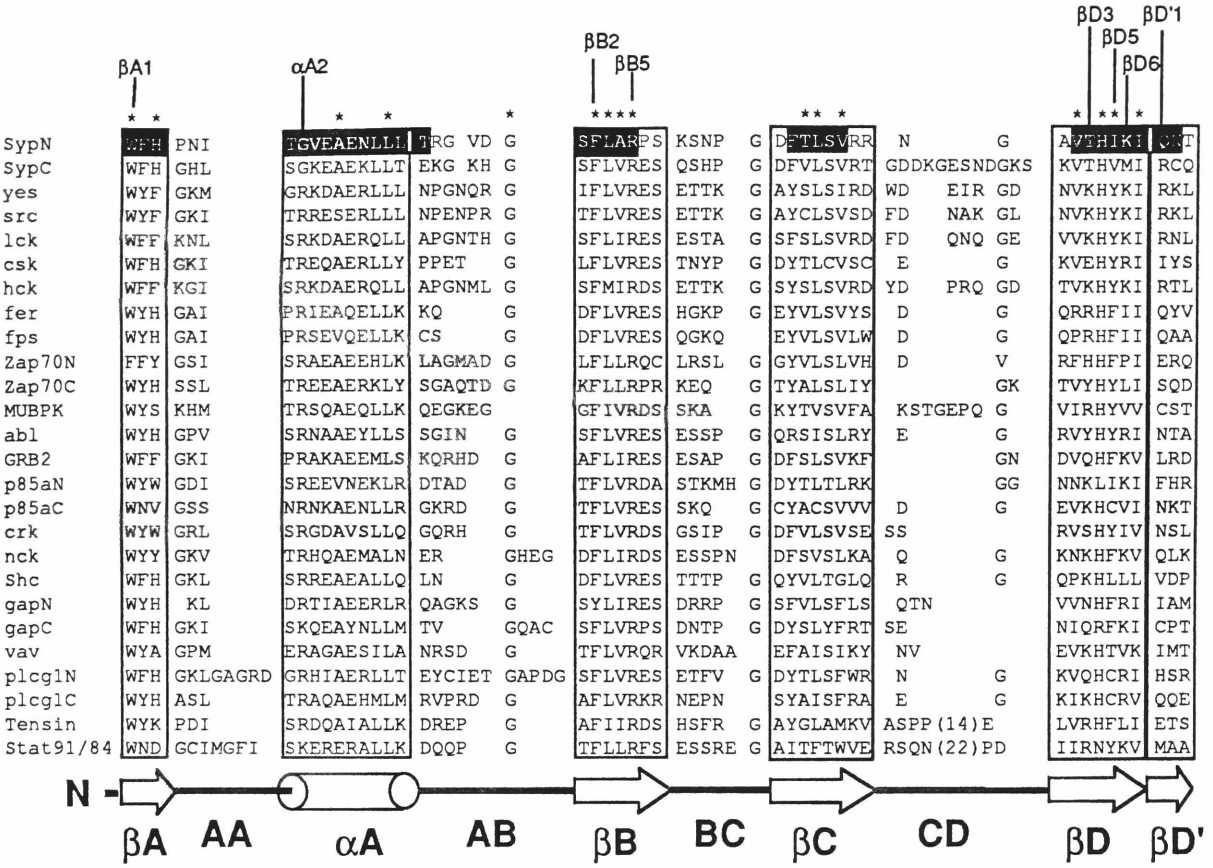
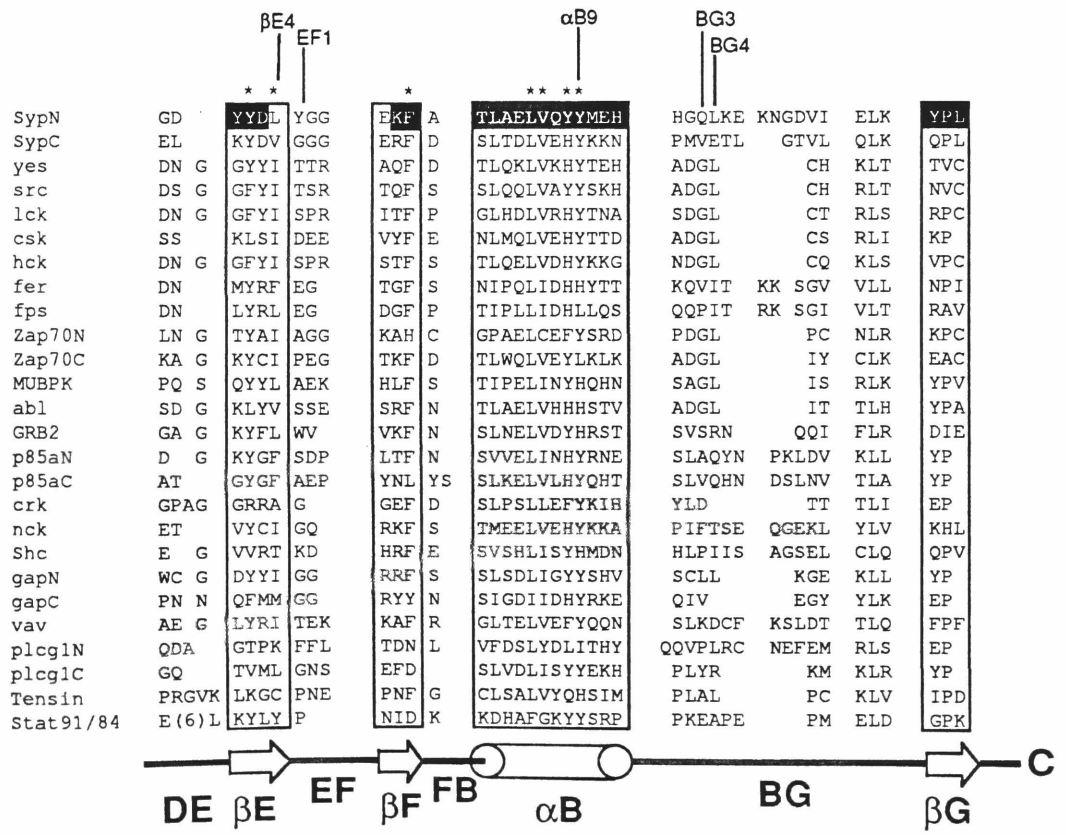

AB
CD
αA
βB
βG
BG
+1
AA
P-Tyr
+3
+4
+2
βC
βD
αB
+5
βD'
EF
BC
βF
FB
βE
DE

Figure 3.1.1.2 Alignment of SH2 sequences and definition of the residues notation. The sequences of different SH2 domains are aligned, based on the secondary structure definitions of Src used previously. The boundaries of the secondary structural elements of Src and shown by solid boxes, and the notation for these elements is shown schematically at the bottom. The important residues are indicated by vertical lines at the top. Not that the secondary structure elements of the SH-PTP2 domain are different from those of Src in some regions. The residues corresponding to the secondary structure elements of SH-PTP2 are shown as white letters on a black background. The residues comprising the common hydrophobic core of SH2 domains are marked with asterisks, as are certain conserved residues involved in phosphotyrosine binding, and which are sued for sequence and structure comparison.


βA1
αA2
βB2
βB5
βD3
βD5
βD6
βD'1
SypN WFH PNI TGVEAENLLL TRG VD G SFLARPS KSNP G DFTLSVRR N G AVTHIKI CST
SypC WFH GHL SGKEAEKLLT EKG KH G SFLVRES QSHP G DFVLSVRT GDDKGESNDGKS KVTHVMI RCQ
yes WYF GKM GRKDAERLLL NPGNQR G IFLVRES ETTK G AYSLSIRD WD EIR GD NVKHYKI RKL
src WYF GKI TRRESERLLL NPENPR G TFLVRES ETTK G AYCLSVSD FD NAK GL NVKHYKI RKL
lck WFF KNL SRKDAERQLL APGNTH G SFLIRES ESTA G SFSLSVRD FD QNQ GE VVKHYKI RNL
csk WFH GKI TREQAERLLY PPET G LFLVRES TNYP G DYTLCVSC E G KVEHYRI IYS
hck WFF KGI SRKDAERQLL APGNML G SFMIRDS ETTK G SYSLSVRD YD PRQ GD TVKHYKI RTL
fer WYH GAI PRIEAQELLK KQ G DFLVRES HGKP G EYVLSVYS D G QRRHFII QYV
fps WYH GAI PRSEVQELLK CS G DFLVRES QGKQ EYVLSVLW D G QPRHFII QAA
Zap70N FFY GSI SRAEAEEHLK LAGMAD G LFLLRQC LRSL G GYVLSLVH D V RFHHFPI ERQ
Zap70C WYH SSL TREEAERKLY SGAQTD G KFLLRPR KEQ G TYALSLIY GK TVYHYLI SQD
MUBPK WYS KHM TRSQAEQLLK QEGKEG GFIVRDS SKA G KYTVSVFA KSTGEPQ G VIRHYVV CST
abl WYH GPV SRNAAEYLLS SGIN G SFLVRES ESSP G QRSISLRY E G RVYHYRI NTA
GRB2 WFF GKI PRAKAEEMLS KQRHD G AFLIRES ESAP G DFSLSVKF GN DVQHFKV LRD
p85aN WYW GDI SREEVNEKLR DTAD G TFLVRDA STKMH G DYTLTLRK GG NNKLIKI FHR
p85aC WNV GSS NRNKAENLLR GKRD G TFLVRES SKQ G CYACSVVV D G EVKHCVI NKT
crk WYW GRL SRGDAVSLLQ GQRH G TFLVRDS GSIP G DFVLSVSE SS RVSHYIV NSL
nck WYY GKV TRHQAEMALN ER GHEG DFLIRDS ESSPN DFSVSLKA Q G KNKHFKV QLK
Shc WFH GKL SRREAEALLQ LN G DFLVRES TTTP G QYVLTGLQ R G QPKHLLL VDP
gapN WYH KL DRTIAEERLR QAGKS G SYLIRES DRRP G SFVLSFLS QTN VVNHFRI IAM
gapC WFH GKI SKQEAYNLLM TV GQAC SFLVRPS DNTP G DYSLYFRT SE NIQRFKI CPT
vav WYA GPM ERAGAESILA NRSD G TFLVRQR VKDAA EFAISIKY NV EVKHTVK IMT
plcg1N WFH GKLGAGRD GRHIAERLLT EYCIET GAPDG SFLVRES ETFV G DYTLSFWR N G KVQHCRI HSR
plcg1C WYH ASL TRAQAEHMLM RVPRD G AFLVRKR NEPN SYAISFRA E G KIKHCRV QQE
Tensin WYK PDI SRDQAIALLK DREP G AFIIRDS HSFR G AYGLAMKV ASPP(14)E LVRHFLI ETS
Stat91/84 WND GCIMGFI SKERERALLK DQQP G TFLLRFS ESSRE G AITFTWVE RSQN(22)PD IIRNYKV MAA
N
βA AA αA AB βB BC βC CD βD βD'


βE4
EF1
αB9
BG3
BG4
SypN GD YYDL YGG EKF A TLAELVQYYMEH HGQLKE KNGDVI ELK YPL
SypC EL KYDV GGG ERF D SLTDLVEHYKKN PMVETL GTVL QLK QPL
yes DN G GYYI TTR AQF D TLQKLVKHYTEH ADGL CH KLT TVC
src DS G GFYI TSR TQF S SLQQLVAYYSKH ADGL CH RLT NVC
lck DN G GFYI SPR ITF P GLHDLVRHYTNA SDGL CT RLS RPC
csk SS KLSI DEE VYF E NLMQLVEHYTTD ADGL CS RLI KP
hck DN G GFYI SPR STF S TLQELVDHYKKG NDGL CQ KLS VPC
fer DN MYRF EG TGF S NIPQLIDHHYTT KQVIT KK SGV VLL NPI
fps DN LYRL EG DGF P TIPLLIDHLLQS QQPIT RK SGI VLT RAV
Zap70N LN G TYAI AGG KAH C GPAELCEFYSRD PDGL PC NLR KPC
Zap70C KA G KYCI PEG TKF D TLWQLVEYLKLK ADGL IY CLK EAC
MUBPK PQ S QYYL AEK HLF S TIPELINYHQHN SAGL IS RLK YPV
abl SD G KLYV SSE SRF N TLAELVHHHSTV ADGL IT TLH YPA
GRB2 GA G KYFL WV VKF N SLNELVDYHRST SVSRN QQI FLR DIE
p85aN D G KYGF SDP LTF N SVVELINHYRNE SLAQYN PKLDV KLL YP
p85aC AT GYGF AEP YNL YS SLKELVLHYQHT SLVQHN DSLNV TLA YP
crk GPAG GRRA G GEF D SLPSLLEFYKIH YLD TT TLI EP
nck ET VYCI GQ RKF S TMEELVEHYKKA PIFTSE QGEKL YLV KHL
Shc E G VVRT KD HRF E SVSHLISYHMDN HLPIIS AGSEL CLQ QPV
gapN WC G DYYI GG RRF S SLSDLIGYYSHV SCLL KGE KLL YP
gapC PN N QFMM GG RYY N SIGDIIDHYRKE QIV EGY YLK EP
vav AE G LYRI TEK KAF R GLTELVEFYQQN SLKDCF KSLDT TLQ FPF
plcg1N QDA GTPK FFL TDN L VFDSLYDLITHY QQVPLRC NEFEM RLS EP
plcg1C GQ TVML GNS EFD SLVDLISYYEKH PLYR KM KLR YP
Tensin PRGVK LKGC PNE PNF G CLSALVYQHSIM PLAL PC KLV IPD
Stat91/84 E(6)L KYLY P NID K KDHAFGKYYSRP PKEAPE PM ELD GPK
C
DE βE EF βF FB αB BG βG

In the SH2-ligand structures, the bound peptide lies across the surface of the domain approximately orthogonal to the central β-sheet (Figure 3.1.1.1). The peptide ligands often adopt an extended conformation and do not participate in secondary structure formation with the domain. Upon binding, the phosphotyrosine "anchors" the domain, and thereby allows the specificity determining residues (the three to six residues immediately following the phosphotyrosine) to be read out. The peptide residues N-terminal to the phosphotyrosine make limited and non-specific interactions with the domains and therefore most likely contribute little to the binding specificity. The N-and C-termini of the SH2 domains are located on the side of the domain opposite to the peptide binding surface. For this reason, the domain can be readily inserted into different molecular contexts without affecting the peptide binding ability.

3.1.2. Recognition of Phosphotyrosine

The defining feature of the SH2 domain is its ability to recognize phosphotyrosine —— a capacity that allows SH2 domains to act as a tyrosine phosphorylation sensor. The recognition mode of phosphotyrosine, first captured in the crystal structure of the Src SH2/peptide complex, is conserved among all SH2 domains in general terms. Residues of the SH2 domain from αA, βB, βD, and the BC loop form the phosphotyrosine binding pocket which provides (i) hydrophobic interactions with the phenolic ring of phosphotyrosine and (ii) hydrogen bonding interactions with the phosphate group.

An invariant interaction with phosphotyrosine is provided by Arg βB5 which forms a bidentate ionic interaction with the phosphate group. This arginine is located at the bottom of the binding pocket; upon binding, it becomes completely buried and is thus inaccessible to solvent. As the backbone of the phosphotyrosine residue is often held in position by the outer strand of the central β sheet (βD), the ionic interaction between the phosphate group and Arg βB5 provides a stereochemical "ruler" that appears to be the key for discriminating between phosphotyrosine and other residues. The location of Arg βB5 is such that ——in a fully extended conformation ——this sidechain is just long enough to interact with the phosphate group of a fully extended phosphotyrosine sidechain. Such a geometric constraint provides a basis for excluding phosphoserine or phosphothreonine. Arg βB5 is strictly conserved in all SH2 domains and its presence in the context of sequence F-L-V-R (the "FLVR arginine, as it is often called) is the signature of an SH2 domain. Even a conservative mutation of this residue to lysine abolishes the binding to phosphotyrosine (Mayer, *et al.,* 1992).

In the SH-PTP2/PDGFR-1009 structure, the phosphotyrosine is well ordered and tightly bound with average temperature factors for the sidechain of 17.5 $Å^2$ (compared to the average temperature factor of 24.8 $Å^2$ for the entire structure). The sidechain of Arg βB5, which forms the salt bridge with the phosphotyrosine (as described above), is essentially unchanged in its position and conformation when compared to the Src structure: the rms deviation between the

atomic positions of this sidechain in Src and in SH-PTP2 is 0.42 Å — when the two structures are superimposed using all C^{α} atoms.

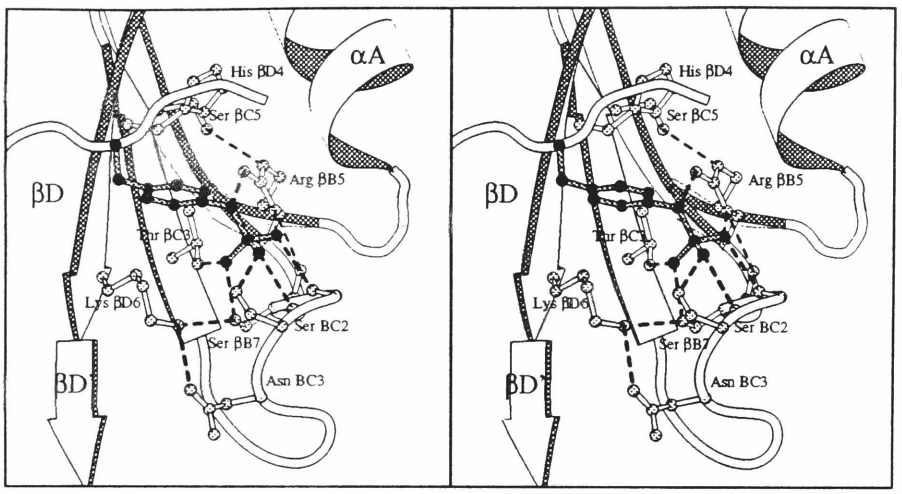
An unusual feature of the Src and Lck SH2-peptide complexes is the presence of an amino-aromatic interaction between an amino nitrogen of Arg αA2 and the phosphotyrosine ring (Waksman *et al.*, 1992; Eck *et al.*, 1993). In addition, the imino nitrogen of this arginine forms a hydrogen bond with one of the terminal oxygens of the phosphate and each of the amino nitrogens forms hydrogen bonds with the carbonyl oxygen of the peptide residue at position -1. Arg αA2 is well conserved (Figure 3.1.1.2) and the most conspicuous feature of the phosphotyrosine binding site in SH-PTP2 is the replacement of this arginine by a glycine. Despite the loss of the arginine mediated interactions, the phosphotyrosine residue is bound very similarly to that in Src or Lck (Figure 3.1.2.1). No significant structural rearrangement has occurred in the protein structure (in the phosphate group of pTyr, instead; see below) to compensate for the missing arginine, and no additional basic residues are found to interact with the phosphotyrosine at the binding site.

There is, however, a significant difference in the conformation of the phosphate group as a consequence of the loss of the interactions with Arg αA2. In the SH-PTP2 structure, the phosphate group is rotated ~180° about the bond linking it to the aromatic ring, and Arg βB5 in SH-PTP2 interacts with both the phenolic oxygen and a terminal oxygen of the phosphate. This is in contrast to Src in which the Arg βB5 interacts with two terminal oxygens (Figure

3.1.2.1). An important effect of the rotation of the phosphate group in SH-PTP2 is that an equivalent number of hydrogen bonds are formed between the phosphate group and the SH2 domain in SH-PTP2 and Src (Figure 3.1.2.2) —— despite the absence of Arg αA2.

As in the Src structure, one edge of the phosphotyrosine ring is sandwiched between the C^{α} and C^{β} atoms of the sidechain of His βD4 and the methylene groups of the sidechain of Lys βD6 (Figures 3.1.2.1 & 3.1.2.2). His βD4 is conserved in almost all SH2 domains. Replacement of this histidine in the N-terminal SH2 domain of GAP by a lysine or a glutamic acid completely abolishes binding (Marengere and Pawson, 1992), thereby demonstrating its critical role. The position of the histidine sidechain in SH-PTP2, like that of Arg βB5, is remarkably similar to that in Src; the rms deviation in the atomic positions of this sidechain is only 0.24 Å between the two structures. Comparison of the Src and SH-PTP2 structures emphasizes that the imidazole ring of His βD4 plays a structural role in forming the phosphotyrosine binding site and does not interact directly with the phosphate. The $N^{\delta 1}$ nitrogen of the histidine is hydrogen bonded to the sidechain oxygen of Ser βC5, which in turn hydrogen bonds to the imino nitrogen of Arg βB5, the key residue that coordinates the phosphate. Serine or threonine residues are found at position βC5 in almost all SH2 domains, indicating conservation of this hydrogen bonding network ——a network which appears to orient Arg βB5 for proper interaction with the phosphate.

Figure 3.1.2.1 Stereo views of the phosphotyrosine binding sites of (a) SH-PTP2 and (b) Src. The polypeptide backbone of the peptide is shown as a tube and the phosphotyrosine sidechain is shown in black. Hydrogen bonds are indicated by dashed lines.


αA
His βD4
Ser βC5
Arg βB5
βD
Thr βC3
Lys βD6
Ser βB7
Ser BC2
Asn BC3
βD'
αA
His βD4
Ser βC5
Arg βB5
βD
Thr βC3
Lys βD6
Ser βB7
Ser BC2
Asn BC3
βD'

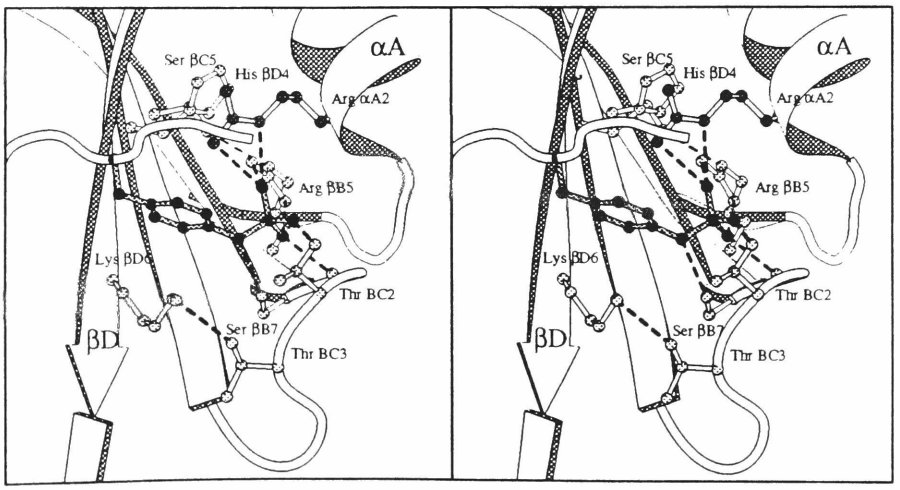
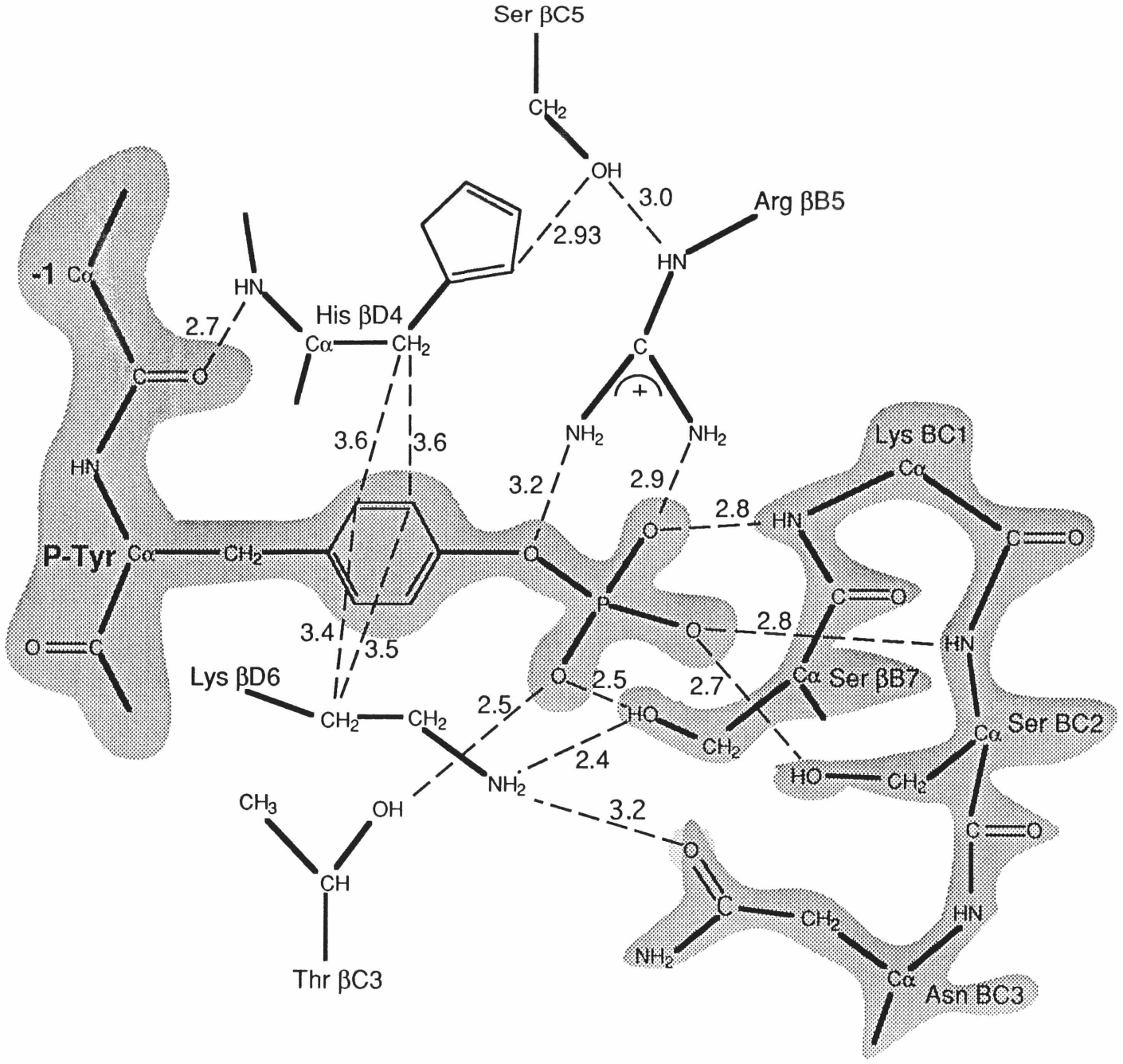
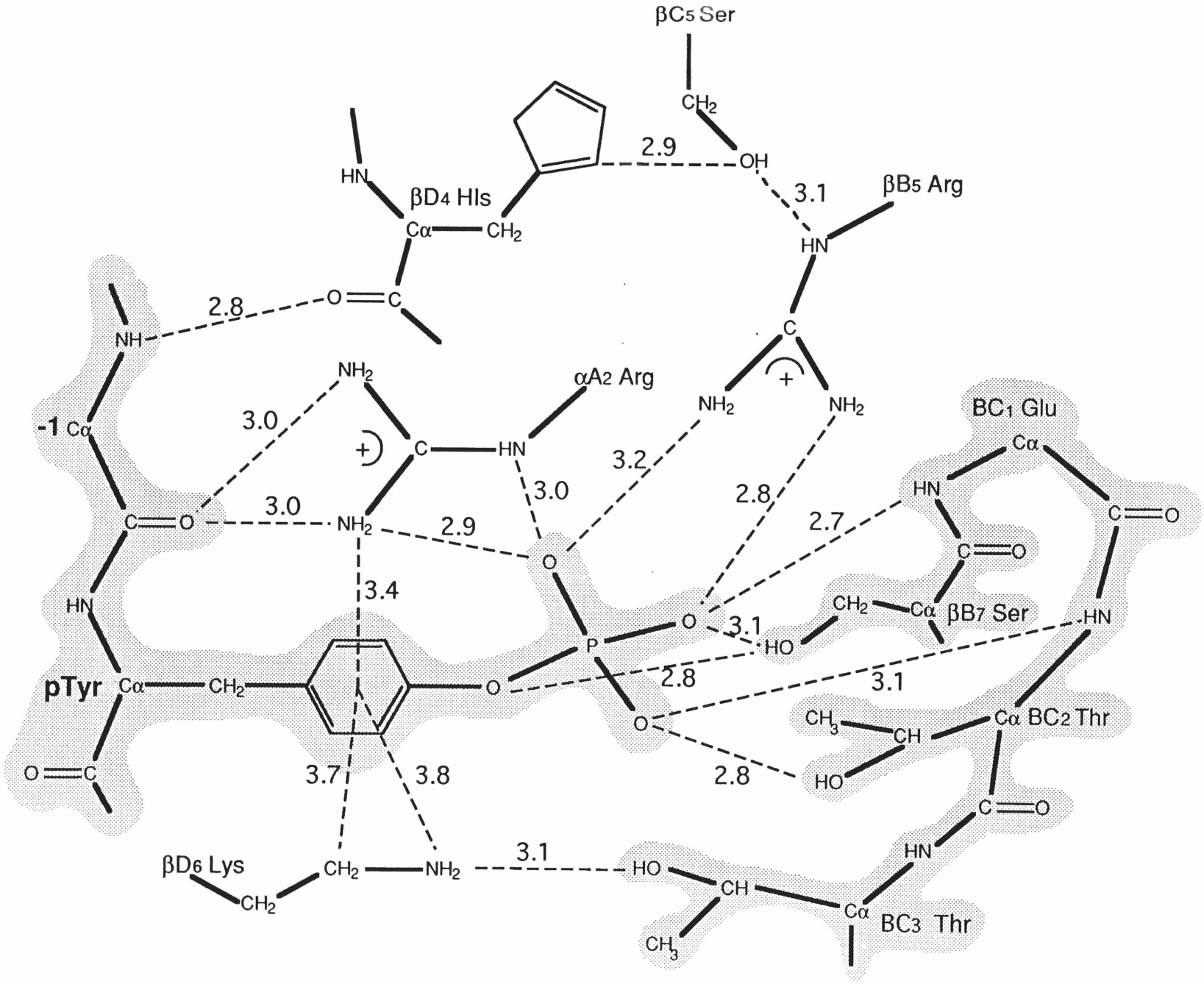

Ser βC5
His βD4
αA
Arg αA2
Arg βB5
Lys βD6
Thr BC2
Ser βB7
Thr BC3
βD
Ser βC5
His βD4
αA
Arg αA2
Arg βB5
Lys βD6
Thr BC2
Ser βB7
Thr BC3
βD

Figure 3.1.2.2 Schematic diagram showing interactions between the phosphotyrosine and the SH2 domain. (a) the SH-PTP2 SH2 domain, for the PDGFR-1009 complex, and (b) the Src SH2 domain for the pYEEI peptide complex. The peptide (including phosphotyrosine) and residues in the BC (phosphotyrosine binding loop) are indicated by shading. Distances (in Å) between non-hydrogen atoms are shown for the second molecule in the asymmetric unit of the PDGFR-1009 complex (a).


Ser βC5
CH2
OH
3.0
Arg βB5
2.93
HN
-1 Cα
HN
His βD4
2.7
Cα
CH2
C
O
+
NH2
NH2
3.6
3.6
3.2
2.9
Lys BC1
Cα
HN
2.8
O
C
O
HN
P-Tyr
Cα
CH2
O
P
C
O
2.8
3.4
3.5
O
C
HN
Lys βD6
O
2.5
2.7
Cα
Ser βB7
2.5
CH2
CH2
HO
CH2
Cα
Ser BC2
2.4
HO
CH2
NH2
3.2
CH3
OH
O
C
CH2
HN
CH
NH2
Cα
Asn BC3
Thr βC3


βC5 Ser
CH2
OH
2.9
3.1
βB5 Arg
HN
βD4 HIs
Cα
CH2
HN
2.8
O
C
NH
C
NH2
αA2 Arg
NH2
+
NH2
BC1 Glu
-1 Cα
3.0
C
HN
3.2
Cα
3.0
2.8
HN
C
O
3.0
NH2
2.9
2.7
C
O
O
3.4
HN
CH2
Cα
βB7 Ser
HN
O
3.1
P
HO
pTyr
Cα
CH2
O
2.8
3.1
O
CH3
CH
Cα
BC2 Thr
O
C
3.7
3.8
2.8
HO
C
O
HN
βD6 Lys
CH2
NH2
3.1
HO
CH
Cα
BC3 Thr
CH2
CH3

The conformation of the sidechain of Lys βD6 is not as well conserved (rms deviation between Src and SH-PTP2 of 1.7 Å) as those of His βD4 and Arg βB5. The lysine sidechain is very well ordered (average B values of 8.72 $Å^2$) and is hydrogen bonded to two sidechains in the BC loop, namely, Ser βB7 and Asn BC3. The amino nitrogen of the lysine is more than 5 Å from the nearest atom in the phosphotyrosine ring and is 3.7 Å from the nearest oxygen atom of the phosphate. Thus the lysine does not provide direct hydrogen bonding or amino-aromatic interactions with the phosphotyrosine. Rather, the main role of this lysine appears to be to stabilize the conformation of the phosphate binding loop and to provide hydrophobic interactions with the ring. This lysine is not well conserved, but is replaced in most cases by hydrophobic or basic groups that can provide similar interactions (Figure 3.1.1.2).

The orientation of the phosphotyrosine ring is very similar in Src and SH-PTP2/PDGFR-1009, although they are tipped at slightly different angles with respect to the C^{α}-C^{β} bond, so as to place the phosphate groups in roughly the same position despite the difference in the orientation of the phosphates (Figure 3.1.2.1). The valine sidechain at position -2 of the peptide covers the face of the phosphotyrosine ring in the SH-PTP2/PDGFR-1009 complex and occupies the position corresponding to the guanidium group of Arg αA2 in Src and Lck. In the SH-PTP2/IRS1-895 and SH-PTP2/PDGFR-740 complexes, the peptide does not cover the phosphotyrosine ring, which is rotated by approximately 70° about the C^{β}-C^{γ} bond, thereby leading to a closer interactions with the

sidechain of His βD4 and a reduction in the solvent accessibility of one face of the ring.

The ability of the phosphotyrosine binding site to compensate for the loss of the Arg αA2 is consistent with the results of mutagenesis studies on the N-terminal SH2 domain of GAP, which showed that replacement of this arginine by lysine, alanine, or even proline reduced ——but did not abolish ——binding to growth factor receptors (Marengere and Pawson 1992). This is in contrast to the essential role played by the strictly conserved Arg βB5. Unlike Arg βB5 which is inaccessible to solvent, Arg αA2 in Src is solvent accessible even in presence of bound phosphotyrosine residue. Thus, the binding energy contributed by the interactions between Arg αA2 sidechain and phosphotyrosine (in Src) would be offset by a considerable desolvation and sidechain entropy penalty. In addition, the amino aromatic interaction mediated by the Arg αA2 is not observed for some SH2 domains (such as p85) which contain an Arginine at αA2. Furthermore, several residues from the BC loop (also known as "phosphate binding loop") form a number of hydrogen bonds with the phosphate group. The exact bonding patterns, however, differs among SH2 domains. As observed for p85 SH2 domain, the bonding patterns even vary between two different peptide complex structures. Taken together, the differences in the detailed components for phosphotyrosine recognition suggests a certain degree of plasticity in the binding pocket.

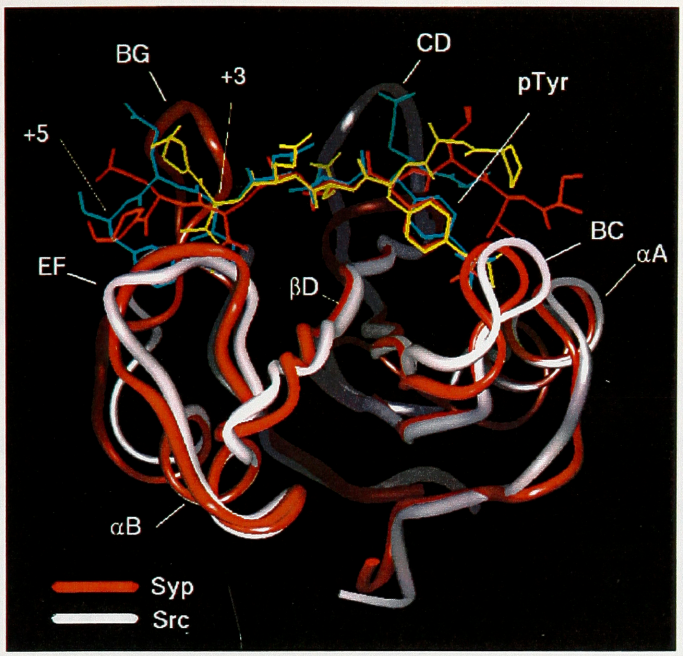
3.1.3. Sequence Specific Recognition

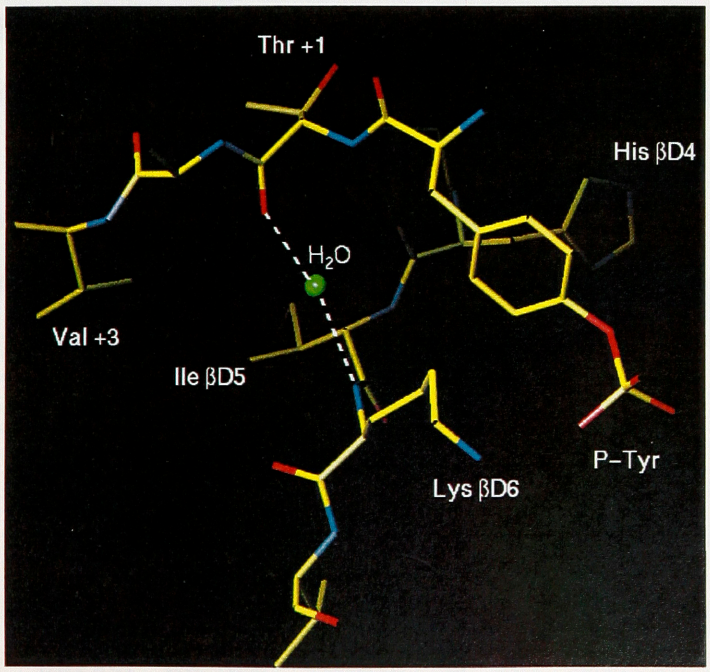
The structures of the closely-related Src and Lck SH2 domains in complex with high affinity pYEEI peptide provides the first piece of structural information on the sequence-specific recognition of SH2 domain. In these structures, the peptides adopt an extended conformation in binding to SH2 domains. The type of interaction, regarded as the "prototypical mode", resembles a two-pronged plug (the peptide) engaging a two-holed socket (the SH2 domain) (Waksman, *et al.*, 1993). The two prongs refer to the phosphotyrosine and the Ile +3 residue of the peptide which fit in the corresponding pockets independently. From this standpoint, the peptide recognition by SH-PTP2 SH2 domain is discussed below.

The most important conclusion obtained from comparing the structures of the two high affinity peptides bound to SH-PTP2 or Src and Lck is that the mode of interaction between the peptide and the SH2 domain is almost identical at the three positions immediately following the phosphotyrosine (Figure 3.1.3.1). The backbone conformation of the peptide is virtually superimposable when the SH2 structures are first overlaid on top of each other using C^{α} atoms (but excluding any atoms of the peptides). The rms deviation between the C^{α} atoms of the peptide residues at +1, +2, and +3 in Src/YEEI compared with either SH-PTP2 structure is 0.5 Å. Similar results are obtained for Lck/YEEI. This remarkable congruence in the positions of the residues that are most important for sequence specific recognition emphasizes that the original binding mode

Figure 3.1.3.1 Comparison of the SH-PTP2 and Src SH2 domains. The polypeptide backbones of the SH-PTP2 and Src SH2 domains are shown as red and white tubes, respectively. The YEEI peptide (Src) is shown in yellow and the IRS1-895 and PDGF-1009 peptides (SH-PTP2) are shown in orange and blue, respectively. The view is approximately perpendicular to the peptide binding surface. The two structures were first superimposed with the program O and displayed using Insight (Biosym Technologies).

Figure 3.1.3.2. The conserved water molecule that bridges the peptide and the SH2 domain in the SH-PTP2, Src and Lck structures. The water molecule (shown as a green sphere) forms hydrogen bonds (indicated by dashed lines) with the amide group of Lys βB6 and the carbonyl group of the peptide residue at +1 in all three SH2 domains. The structure shown is from the SH-PTP2/PDGFR-1009 complex, with carbon and phosphorus atoms are colored yellow, nitrogen and oxygen atoms colored blue and red, respectively.


BG
+3
CD
pTyr
+5
BC
αA
EF
βD
αB
Syp
Src


Thr +1
His βD4
H2O
Val +3
Ile βD5
P-Tyr
Lys βD6

proposed for Src and Lck (on the basis of the structure of the YEEI complex) captures the most important determinants of SH2-peptide interactions.

No such structural similarity is observed in the N-terminal regions of the peptides. In the SH-PTP2/PDGFR-1009 and Src/YEEI structures the two residues that are immediately N-terminal to the phosphotyrosine do not make significant interactions with the SH2 domain, but instead form a cap over the phosphotyrosine ring. Their conformations are significantly different, with an rms deviation of 2.5 Å in C^{α} positions between the Src and SH-PTP2 structures. In the case of the SH-PTP2/IRS1-895 structure, the N-terminal residues run parallel to strand βD of the central β sheet. These two N-terminal residues deviate from the corresponding residues in the SH-PTP2/PDGFR-1009 and Src/YEEI structures by 4.5 Å rms and 6.5 Å rms, respectively. However, N-terminal region contributions to specificity in other SH2 domains can not be ruled out by these results. Nevertheless, this pattern of structural divergence N-terminal to the phosphotyrosine and structural convergence C-terminal to the phosphotyrosine *reinforces* the idea that the most important determinants of sequence specificity are the peptide sequences immediately downstream of the phosphotyrosine.

There appear to be two important determinants of the direction of the peptide backbone. One is that the peptide is anchored at the +3 position by hydrophobic interactions between a hydrophobic pocket on the surface of the SH2 domains and the

sidechain of the +3 residue (Ile in SH-PTP2/IRS1-895, Src/YEEI and Lck/YEEI); Val in SH-PTP2/PDGFR-1009). Another conserved interaction that may set the direction of the peptide involves a water molecule that is found in virtually identical positions in SH-PTP2, Src, and Lck. This water molecule is highly ordered in SH-PTP2 with a temperature factor of 9.7 $Å^2$ and forms hydrogen bonds with the carbonyl group of the peptide residue at +1 position and the amide nitrogen of Lys βD6 (Figure 3.1.3.2).

There are, however, significant differences in the structure of the binding sites for peptide residues at +1 and +3. These changes are due to the differences in the sequence of SH-PTP2 with respect to Src and Lck. In Src and Lck, there is a physical partitioning of these binding sites due to the presence of Tyr βD5, which extends up from underneath the peptide and hydrogen bonds to the BG loop (Figure 3.1.3.3a). The +3 binding site in Src and Lck is further closed off by the approach of sidechains in the BC and EF loops, giving the appearance of a "hole" into which the isoleucine sidechain of the YEEI peptide fits (see Figure 3.1.3.3b). In SH-PTP2, Tyr βD5 is replaced by an isoleucine, and the absence of a bridge to the BG loop results in the formation of a shallow channel in which the sidechains at the +1 and +3 positions are bound (Figures 3.1.3.3a & 3.1.3.4). Rearrangements in the BG and EF loops result in this channel being open beyond the +3 site (Figures 3.1.3.4 & 3.1.3.5). As a result, there are interactions between the peptide and the SH2 domain at the +4 and +5 sites as well (see below).

Figure 3.1.3.3 Stereo view of the +1 and +3 peptide binding sites of (a) SH-PTP2/PDGFR-1009 complex and (b) Src/YEEI complex. View from the peptide binding surface, showing the interaction between residues of the SH2 domain and the peptide residues +1 and +3. The peptide residues are shown in white, except for the +3 residue, which is green. Important residues on the SH2 domain are labeled (**a** to **g** for SH-PTP2 and **a** to **e** for Src). In SH-PTP2 (A) the labeling is as follows: **a**, Ile βD5; **b**, Thr βD3; **c**, Ile BG12, **d**, Leu BG4, **e**, βE4, **f**, Arg βC8 and **g**, Glu BG6. The peptide residues +1 and +3 are Thr and Val, respectively. In Src (B), the labeling is as follows: **a**, Tyr βD5, **b**, Lys βD3, **c**, Tyr αB9, **d**, Leu BG4 and **e**, Ile βE4. The peptide residues at +1 and +3 are Glu and Ile, respectively.

Figure 3.1.3.4 The molecular surface of the SH-PTP2/PDGFR-1009 complex. The surface is calculated with the peptide removed (using GRASP) and is colored according to the local electrostatic potential calculated in the absence of peptide, ranging from deep blue (most positive region) to deep red (most negative). The peptide is shown in red and important residues in the SH2 domain are indicated.

f
c
d
g
b
+1
a
+3
e

d
b
c
+3
a
+1
e

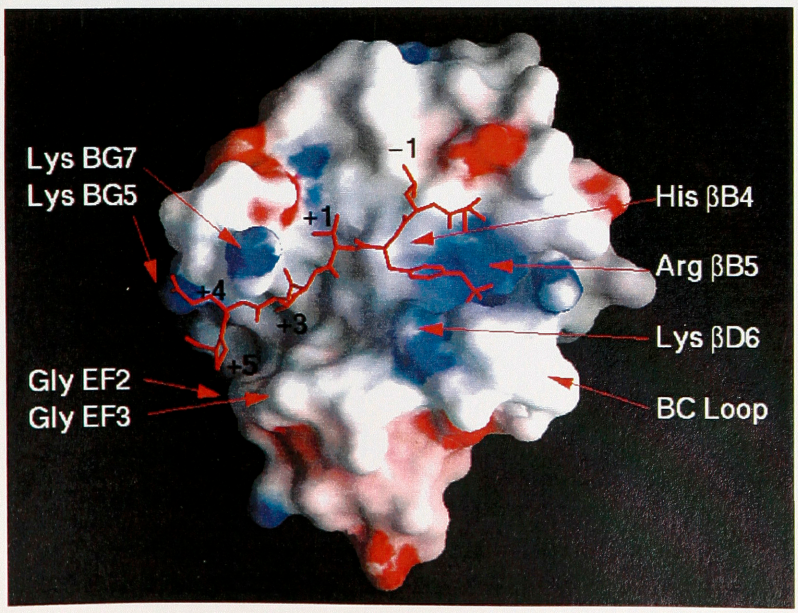
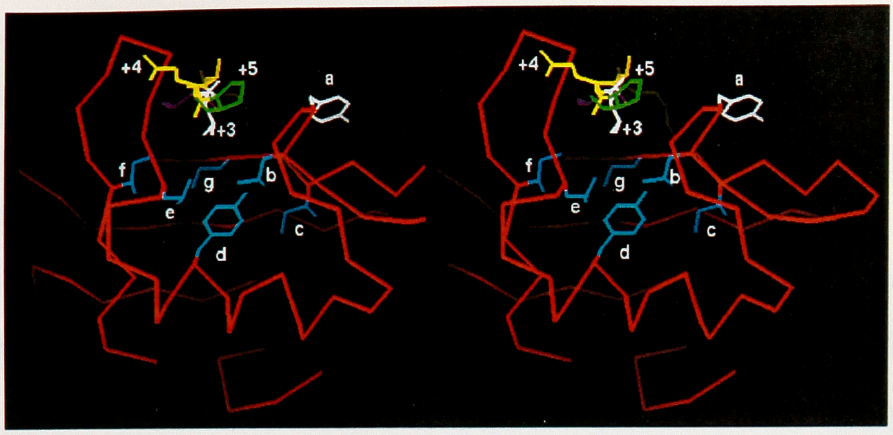
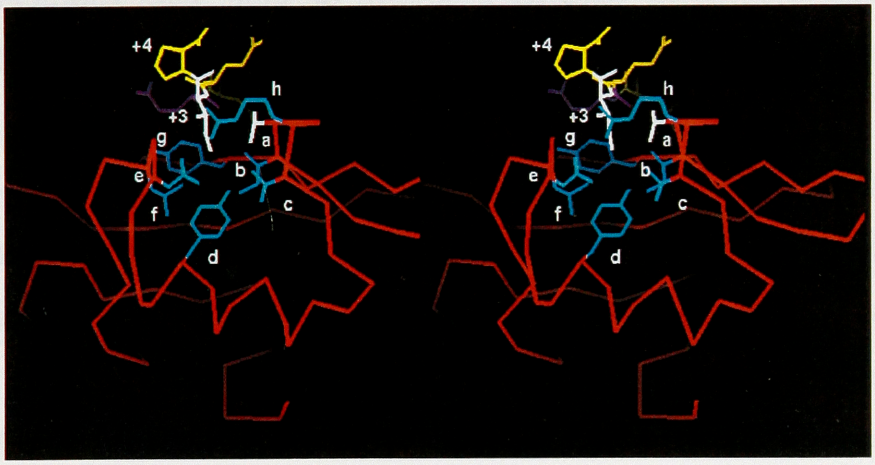
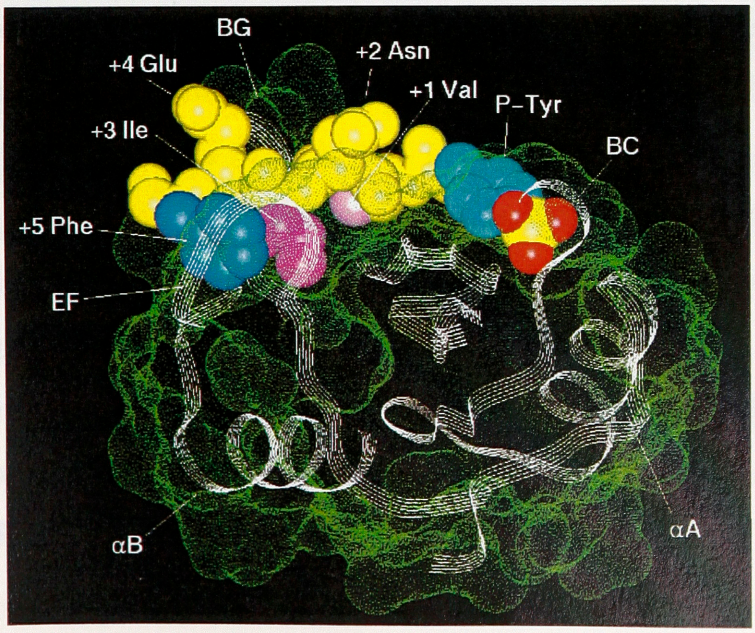

Lys BG7
Lys BG5
−1
+1
His βB4
Arg βB5
+4
+3
Lys βD6
+5
Gly EF2
Gly EF3
BC Loop

Figure 3.1.3.5 Stereo view of the +3 peptide binding site of (a) SH-PTP2 PDGFR-1009 complex and (b) Src YEEI complex. The view is from C-terminus of the peptide and is approximately perpendicular to the peptide binding surface. Helix αB is at the bottom and the BG and EF loops are at the left and the right, respectively. The C^{α} backbone of the SH2 domain is shown in red and the peptide residues at +1, +3, +5 are shown in purple, white and green, respectively, with the rest of the peptide in yellow. Important residues around the +3 peptide binding sites are colored blue (except for **a** which is white) and labeled. In SH-PTP2 (A), the labeling is as follows: **a**, Tyr EF1; **b**, Leu βE4; **c**, Ile βD7; **d**, Tyr αB9; **e**; Leu BG4; **f**, Ile BG12 and **g**, Ile βD5. The peptide residues +3, +4 and +5 are Val, Gln and Pro respectively. In Src (B), the labeling is as follows: **a**, Thr EF1; **b**, Ile βE4; **c**, Thr βF1; **d**, Tyr αB9; **e**, Asp BG2; **f**, Leu BG4; **g**, Tyr βD5 and **h**, Arg EF3. The peptide residues at +3 and +4 are Ile and Pro, respectively.

Figure 3.1.3.6 Cutaway view of the SH-PTP2 SH2 domain, showing the interactions with the IRS1-895 peptide. The side chain of Phe (+5) of the peptide and aromatic ring of the phosphotyrosine are shown in blue. Ile (+3) and Val (+1) are colored purple and pink, respectively.


+4
+5
+3
a
f
g
b
e
c
d
+4
+5
+3
a
f
g
b
e
c
d


+4
h
+3
g
a
e
b
c
f
d
+4
h
+3
g
a
e
b
c
f
d


BG
+4 Glu
+2 Asn
+1 Val
P-Tyr
+3 Ile
BC
+5 Phe
EF
αB
αA

The selectivity at the +1 site is different for SH-PTP2 which prefers hydrophobic residues at this position (as compared to Src or Lck which prefers glutamate or polar non-basic residues) (Songyang *et al.*, 1993). Important determinants of the change in selectivity are likely to be the replacement of Tyr βD5 by isoleucine (a less bulky sidechain) in SH-PTP2. As mentioned earlier, this results in the +1 and +3 sites being contiguous. Also, the replacement of Lys βD3 by threonine in SH-PTP2 removes positive charge from the vicinity of this site. Peptides that bind to SH-PTP2 contain isoleucine (found in IRS1-1172, the peptide with highest affinity for SH-PTP2), valine, threonine, or leucine at the +1 position (Case *et al.*, 1994) (Table 3.1.4.1) with valine and threonine represented in the peptides used in this study. Nevertheless, the conformation of the +1 residue is very similar in both cases with the sidechain lying across the surface of the domain and not enclosed in a pocket. In the SH-PTP2/PDGFR-1009 structure, the oxygen atom of the threonine is hydrogen bonded to a water molecule that bridges this atom and the backbone amide of His βD4. The methyl group of the threonine is in van der Waals contact with a methyl group of Thr βD3 and with the C^{γ} methylene group of Glu BG6, which is part of the BG insertion in SH-PTP2. The SH2 domains of the p85 subunit of PI 3-kinase also have insertions in the BG loop (Figure 3.1.1.2). Chemical cross-linking studies and subsequent structural analysis have indicated a similar interaction between this loop and the +1 site for p85 SH2 domain (Williams *et al.*, 1993). This is consistent with the observed contact between Glu BG6 and the +1 sidechain in SH-PTP2.

Neither sidechain at +1 makes extensive interactions with SH2 sidechains. Thus selectivity at this site is likely to be weak, consistent with the *in vitro* selection results (Songyang *et al.*, 1993). The C^{α}-C^{β} bond of the peptide sidechain is oriented approximately parallel to the peptide binding surface so that the sidechain is not engulfed by the SH2 domain. However, the peptide backbone is twisted beyond the +1 position, such that the C^{α}-C^{β} bond of the sidechain at +3 is rotated almost by 90° and points directly into the surface of the SH2 domain. The sidechain at this position (valine in PDGFR-1009 and isoleucine in IRS1-895) is almost completely buried due to interactions with hydrophobic sidechains of the SH2 domain. These interactions are reminiscent of the anchoring of the isoleucine sidechain in the Src/YEEI (or Lck/YEEI) complex, although the specific residues that interact are different in the SH-PTP2 (see Figure 3.1.3.5).

The SH-PTP2 structures explain the selection for hydrophobic residues at +1 and +3 in terms of the nature of the SH2 sidechains that are present at each binding site. However, it is very difficult to predict the relative binding affinities for different hydrophobic sidechains in the peptide, such as the differences between β-branched sidechains compared to methionine (as reported by Case *et al.*, 1993). A complete understanding of the gradations in selectivity between different high affinity peptides will require binding measurements on systematically varied peptide sequences. Interpretation of these experiments will be aided greatly by the availability of the present structures.

Perhaps the clearest correlation between changes in sequence and changes in peptide binding specificity is seen at the outer edge of the peptide binding surface (the +5 binding site). This region has at its base helix αB and is framed on two sides by the EF and BG loops. In Src, the sidechains of Thr EF1, Arg EF3, Thr βF1, and Asp BG2 contribute towards the closure of the outer edge of the +3 site (Figure 3.1.3.5). In SH-PTP2, Thr EF1 is replaced by tyrosine, and the backbone of the polypeptide chain is rearranged such that the large tyrosine sidechain now points in the opposite direction —— away from the peptide binding region (Figure 3.1.3.5, residue colored white) ——a conformation that is also adopted in the uncomplexed form of SH-PTP2. The residue immediately following ——which is Ser EF2 in Src ——is mutated to glycine in SH-PTP2, as are Arg EF3 and Asp BG2. The result is a more open binding site in SH-PTP2 (Figure 3.1.3.4). Consequently, the +5 sidechains of both the PDGFR-1009 and IRS1-895 peptides interact with the SH2 domain in this region. Particularly striking is the conformation of Phe +5 (IRS1-895), which is well localized between the BG and EF loops (Figure 3.1.2.1); this Phe appears to cap the isoleucine at +3 (Figure 3.1.3.6). The proline residue at +5 in the PDGFR-1009 peptide does not interact as extensively with the SH2 domain, but the peptide conformation is very similar. The BG loop ——6 residues longer in SH-PTP2 than in Src —— is seen to partially fold over the peptide, which may contribute towards the interactions at +4 and +5. Two direct hydrogen bonds are formed in this region: (i) the carbonyl oxygen of Lys BG5 is hydrogen bonded to the amide nitrogen of peptide residue

at +4 and (ii) the amide nitrogen of Lys BG7 is hydrogen bonded to the carbonyl oxygen of the peptide residue at +1.

3.1.4. Peptide Binding Affinity and Specificity

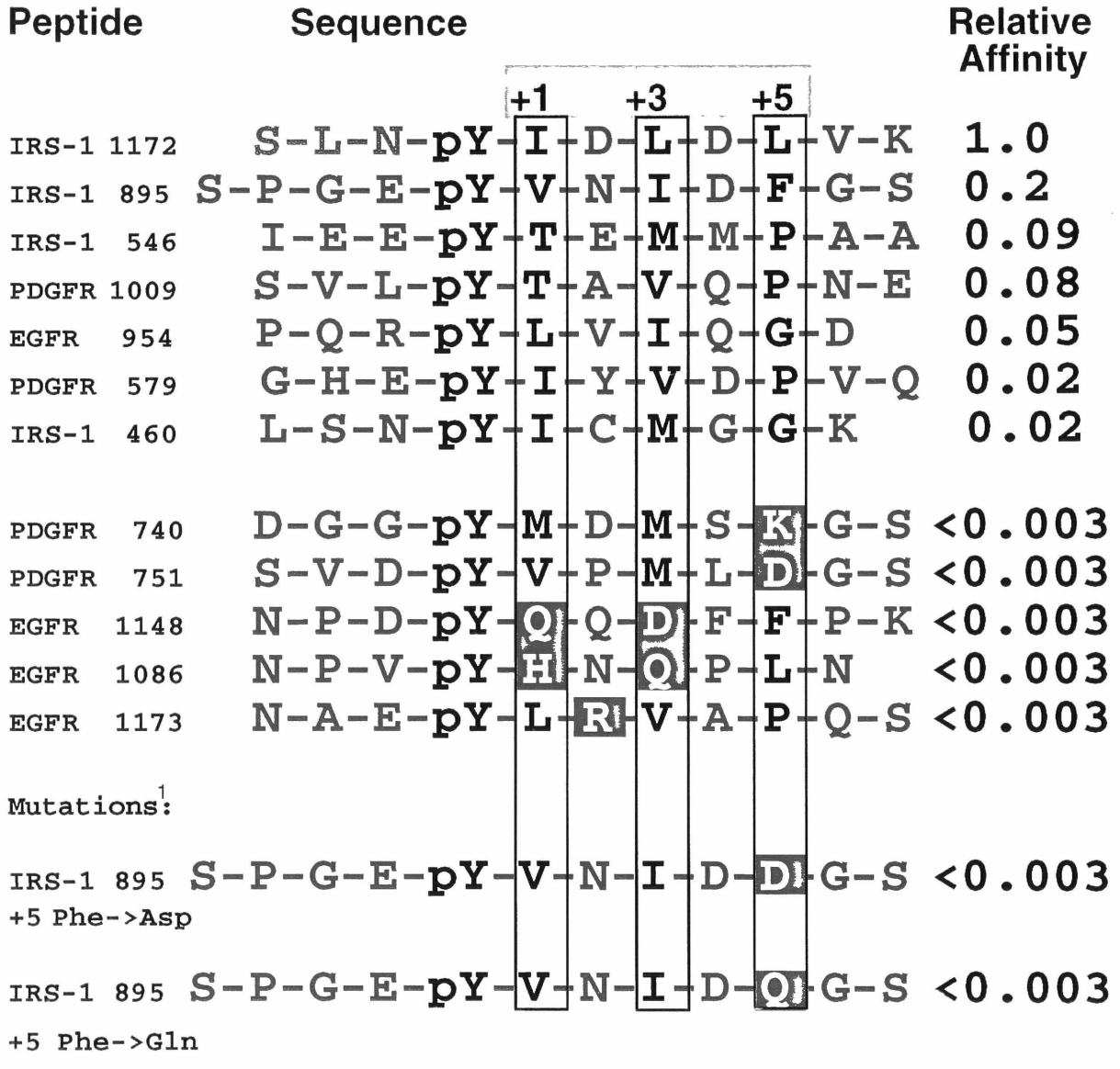
As described, the linear sequence surrounding the specific phosphotyrosine appears to be the major determinant of specificity in binding to the SH2 domains. Previous studies using a random phosphopeptide library demonstrated that individual SH2 domain shows distinct binding specificity in the three residues immediately C-terminal to phosphotyrosine in their peptide ligands (Songyang, *et al.*, 1992). The overall selectivity measured by this study is rather modest, around 40~100. An obvious assumption made in the studies of this type is that the interaction mediated by each peptide residue is not interdependent, an assumption that is only true in general terms. Perhaps, it is not surprising that quantitative analyses of optimal peptides ligand and their mutants gave slightly different results. Although the affinity for the SH2-peptide interaction is generally moderate (K_D~0.1-3.0 μM), considerable selectivity up to a few hundred-fold has been observed for a single peptide residue — particularly at +1 or +3 position. Nevertheless, the results obtained by these two different approaches (i.e. random peptide library and mutational analysis) are generally consistent with each other qualitatively. The derived consensus peptide sequence often agrees well with the sequences of *in vivo* binding sites.

However, a general understanding of the binding specificity of the SH-PTP2 SH2 domain was not reached by the biochemical studies. The selectivity evaluated by the random library study is surprisingly weak (7 vs. an average of 57) and the results of the mutational analysis were difficult to interpret. In these studies, only three residues following the phosphotyrosine were analyzed. The structural findings that the interaction mediated by the SH-PTP2 SH2 domain involves five peptide residues following the phosphotyrosine is therefore unexpected. In the light of the hydrophobic interaction mediated by the peptide residue at the +5 position, it has become apparent that all the peptides that exhibit high affinity binding also contain hydrophobic sidechains at the +5 position (Leu, Phe, Pro or Gly, see Table 3.1.4.1). Shoelson and co-workers have tested this selectivity (for a hydrophobic residue) explicitly by changing the phenylalanine residue at the +5 position in the IRS1-895 peptide (S-P-G-E-pY-V-N-I-D-F-G-S) to either asparagine or to aspartic acid. In both cases, the affinity of the peptide for the SH-PTP2 SH2 domain dropped to less than the sensitivity of the competition assay technique, thus corresponding to a 300-fold less affinity than the tightest binding peptide. In addition, it was noted that truncation of the peptide to less than 5 residues following the phosphotyrosine also abolished high affinity binding (Case *et al.*, 1994).

Peptides that have polar sidechains at the +1 and +3 positions do not bind tightly to the SH-PTP2 SH2 domain, thus consistent with the hydrophobic nature of the binding site for these residues. This result is also in agreement with the peptide library study which

demonstrates noticeable selectivity for hydrophobic residues (particularly valine and isoleucine) at these two positions. In addition to the selectivity for hydrophobic residues at +1, +3 and +5, it is likely that the SH-PTP2 SH2 domain may exert some selectivity — albeit weak —— at positions +2 and +4. The presence of two lysines in the BG loop (BG5 and BG7) places positive charge in the vicinity of the sidechains at these positions (see the electrostatic potential in Figure 3.1.3.4) thereby possibly disfavoring basic resides at these sites. Indeed, the EGFR-1173 peptide —— a peptide which has hydrophobic residues at +1, +3 and +5 (Leu, Val and Pro) but does not bind to SH-PTP2 SH2 with high affinity (Table 3.1.4.1) ——has an arginine at position +2. Although the peptides studied so far do not have basic residues at the +4 site, the prediction that such residues would impair binding can be readily tested.

Table 3.1.4.1 Relative binding affinities of phosphopeptides to the N-terminal SH2 domain of SH-PTP2. Data for the IRS-1, PDGF receptor and EGF receptor peptides have been taken from reference (Case *et al.*, 1994) Note that not all of the phosphopeptides correspond to actual phosphorylation sites *in vivo*. Data on the mutations to the IRS-1 895 peptide were obtained as a consequence of the present structural analysis (S.E. Shoelson *et al.*, unpublished). The phosphotyrosine residue is indicated by "pY", and the residues that are important for specificity are in bold. Polar residues at the +1, +3 or +5 positions that correlate with reduced binding affinity are shown as white on black, as is an arginine residue in EGFR-1173 that might reflect negative selectivity at position +2.


Peptide
Sequence
Relative Affinity
+1
+3
+5
IRS-1 1172 S-L-N-pY-I-D-L-D-L-V-K 1.0
IRS-1 895 S-P-G-E-pY-V-N-I-D-F-G-S 0.2
IRS-1 546 I-E-E-pY-T-E-M-M-P-A-A 0.09
PDGFR 1009 S-V-L-pY-T-A-V-Q-P-N-E 0.08
EGFR 954 P-Q-R-pY-L-V-I-Q-G-D 0.05
PDGFR 579 G-H-E-pY-I-Y-V-D-P-V-Q 0.02
IRS-1 460 L-S-N-pY-I-C-M-G-G-K 0.02
PDGFR 740 D-G-G-pY-M-D-M-S-K-G-S <0.003
PDGFR 751 S-V-D-pY-V-P-M-L-D-G-S <0.003
EGFR 1148 N-P-D-pY-Q-Q-D-F-F-P-K <0.003
EGFR 1086 N-P-V-pY-H-N-Q-P-L-N <0.003
EGFR 1173 N-A-E-pY-L-R-V-A-P-Q-S <0.003
Mutations:[1]
IRS-1 895 S-P-G-E-pY-V-N-I-D-D-G-S <0.003
+5 Phe->Asp
IRS-1 895 S-P-G-E-pY-V-N-I-D-Q-G-S <0.003
+5 Phe->Gln

3.1.5 Summary

Crystal structures of the amino-terminal SH2 domain of SH-PTP2 in separate complexes with two high affinity peptides, in complex with a non-specific peptide and in the uncomplexed form have been determined at between 2Å and 3Å. The structure of the SH2 domain and the mode of high-affinity peptide binding is essentially similar to that seen in the Src and Lck structures. However the binding interface is more extensive in SH-PTP2. Most SH2 target peptides have hydrophobic residues at the third position following the phosphotyrosine, and the SH-PTP2 structure confirms that the peptide is anchored to the SH2 surface by this residue and by the phosphotyrosine. In addition, the SH-PTP2 structure has revealed that sequence specificity can extend across the five residues following the phosphotyrosine, and has shown how the SH2 domain's surface topography can be altered with resulting changes in specificity, while conserving the structure of the central core of the domain.

3.2 Crystal Structure of the X-11 PTB Domain in Complex with APP Peptide

The X-11 project was carried out through a close collaboration with two postdoctors, A. Zhang and V. Mandiyan in J. Schlessinger's laboratory at NYU Medical Center. All the biochemistry experiments (including protein purification and binding studies) were performed at NYU while my involvement in this project was limited to structure determination and analysis. Nevertheless, the binding study is an integrated part of this project and some of the results are thus included here in order to facilitate discussion.

3.2.1 The Binding of the X11 PTB Domain to APP Peptide by Biosensor Analysis

Using the Biosensor assay, it was found that X-11 PTB domain binds strongly to both phosphorylated and unphosphorylated APP peptides with less than a two-fold difference in their binding affinities (Table 3.2.1.1). The highest affinity was observed for a 14mer unphosphorylated APP peptide whose equilibrium dissociation constant (K_D) for binding to the X11 PTB domain is 0.32 μM (as determined by competition assay). This affinity is comparable with those of high-affinity phosphopeptide interactions mediated by the Shc PTB (K_D=42 nM, for Trk peptide, (Mandiyan *et al.*, 1996)) and the IRS-1 PTB domains (K_D=6 μM, for IL-4 peptide, (Zhou *et al.*, 1996)).

Table 3.2.1.1. Binding Affinity of APP Peptides for the X11 PTB Domain

Peptide	Sequence	K_D (μM)	r^2	Relative Affinity
14mer	QNGYENPTYKFFEQ	0.32 ± 0.03	0.981	1.0
E(-4)A	QNGY**A**NPTYKFFEQ	1.73 ± 0.14	0.993	0.19
F(+2)A	QNGYENPTYK**A**FEQ	3.44 ± 0.48	0.976	0.10
F(+3)A	QNGYENPTYKF**A**EQ	3.11 ± 0.30	0.989	0.11
E(+4)A	QNGYENPTYKFF**A**Q	0.59 ± 0.05	0.992	0.56
10mer	GYENPTYKFF	4.56 ± 0.70	0.963	0.07
10mer+P	GYENPTY*KFF	8.26 ± 1.71	0.932	0.04

1. The equilibrium dissociate constants, K_D are determined by SPR competition assay (see Materials and Methods) and are averages of three independent experiments.
1. r^2 is the coefficient of determinants.
2. * indicates that the tyrosine is phosphorylated.
3. the 10 residue (10mer) and 14 residue (14mer) APP peptides were also used in the crystallographic study
4. peptide residues are numbered as following :

Q	N	G	Y	E	N	P	T	Y	K	F	F	E	Q
-8	-7	-6	-5	-4	-3	-2	-1	+0	+1	+2	+3	+4	+5

The kinetic parameters for the 14mer unphosphorylated APP peptide were also measured which is comparable with those of the Shc PTB domains (Laminet *et al.*, 1996; Zhou *et al.*, 1995). The determined dissociation rate constant (k_{off}) is 8.5 X 10^4 $M^{-1}s^{-1}$ and the derived association rate constant (k_{on}) is 2.75 X 10^{-2} s^{-1}. Together they gave a K_D of 0.33 μM that is almost identical to the value obtained by competition assay. It was noted that both dissociation and association rate constants are around 10-fold lower than those of SH2/high affinity peptide interactions (k_{on}, 1 X 10^5 to 2 X 10^6 $M^{-1}s^{-1}$ and k_{off}, ~0.1 s^{-1}). Whether or not the relatively slow kinetics of the PTB/peptide interaction reflect distinct biological functions of the SH2 domains and the PTB domains awaits further investigation. In order to study the binding specificity of the X11 PTB domain, alanine scanning mutagenesis on the APP peptides were performed and the results will be described below in conjunction with structural information.

3.2.2 Overall Structure and Peptide Binding Mode

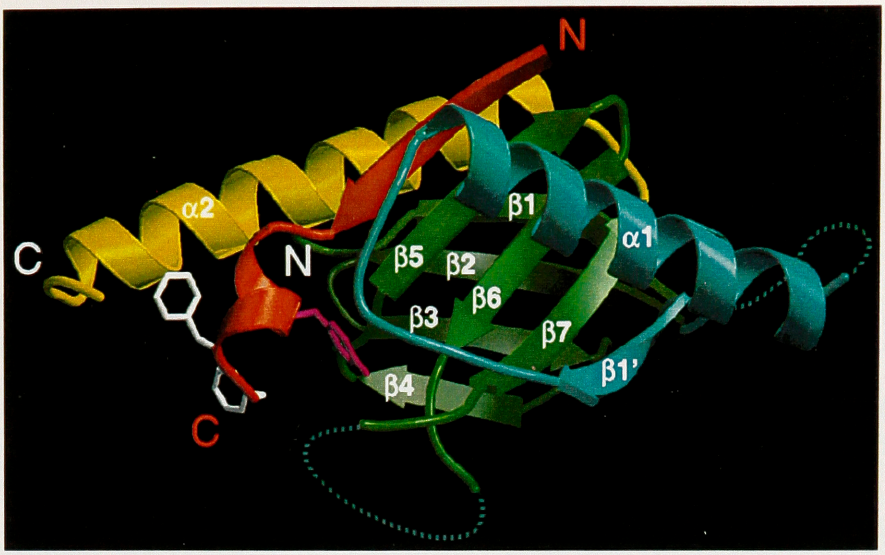
The crystal structures of X11 in separate complexes with two APP peptides (10mer and 14mer) have been determined to nominal resolutions of 2.3 and 2.5 Å, respectively. SIRAS (single isomorphous replacement with anomalous scattering) in combination with MAD (multiple anomalous dispersion) methodology (Hendrickson, 1991) was used to provide independent phase information and allowed an unbiased model to be built. The two crystallographically independent X11/APP peptide complexes in the asymmetrical unit are very

similar, with rms deviation of 0.94Å for all Cα atoms. However, there are subtle differences in terms of the detailed peptide interactions between these two complexes, indicating a certain degree of flexibility in the interacting interface (see below).

The overall structure of the X-11/APP peptide is shown in Figure 3.2.2.1 Like the Shc and the IRS-1 PTB domains, the X-11 PTB domain contains a *bona fide* "PH-fold" (pleckstrin homology domain) that is composed of (i) two anti-parallel β-sheets forming a β-sandwich structure (Figure 3.2.2.1, colored green), and (ii) a C-terminal α-helix (α2; colored yellow) packing on one edge of the β-sheets (colored light green) (Lemmon *et al.*, 1996). With respect to the basic PH-fold, X11 (as well as other Shc-like PTB domains) contains a large insertion that forms the β1' strand and the α1 helix (colored in blue, Figure 3.2.2.1; also see Figure 3.2.2.3) (Zhou *et al.*, 1995). While the β1' strand joins the β sandwich on the opposite edge of peptide binding site, the α1 helix packs against one face of the β-sheets. The N-terminal tip of helix α1 together with the β1'/β1 loop flanks the bound peptide on one side with helix α2 on the other side. Furthermore, the X11 PTB domain contains a unique insertion of 20 residues in the β6/β7 loop in contrast to the Shc and other similar PTB domains. This large loop is located near the tyrosine residue of the NPxY motif of the APP peptides. However, it mediates few interactions with the bound peptide and is mostly disordered in the structure.

Figure 3.2.2.1. Ribbon representation of the structure of the X11 PTB domain in complex with a 14 residue APP peptide. The core structure of the X11 PTB domain resembles the "PH" fold that consists of a β sandwich (colored in light and dark green) and the C-terminal helix (α2; in yellow). With respect to PH fold and IRS-1 PTB domains, the X11 PTB contains a insertion that forms helix α1 and strand β1' (colored in blue). The APP peptide (colored in red) forms an anti-parallel β-strand with strands β5 of the X11 PTB. Also shown are the peptide sidechains of the tyrosine (+0; in pink) and the two C-terminal phenylalanine (+2 and +3; in white) of the peptide. Two missing loops (α1/β2 and β6/β7) are indicated by dash lines.

Figure 3.2.2.2. Stereodiagram of the interactions between the X11 PTB domain and the APP peptide. The X11 PTB domain is in ribbons with sidechains colored in white. The peptide (in yellow) is shown in ball-and-stick representation with oxygen atom colored in red, and nitrogen in blue. The labeling is as following: a=Phe 487 ; b=Tyr 484; c=Phe 480; d=Gly 477; e=Ala 473; f=Gln 474; g=Asp 421; h=Ile 420; i=Tyr 419, j=Ser 418; k=359; l=Met 355; m=Gln 353; n=Arg 354; o=Ser 345. The sidechain of a peptide residue, Lys(+1) is removed for clarity. Figures are generated using MOLSCRIPT (Kraulis, 1991) and Raster3D (Bacon and Anderson, 1988).


N
α2
β1
α1
C
N
β5
β2
β6
β3
β7
β1'
β4
C

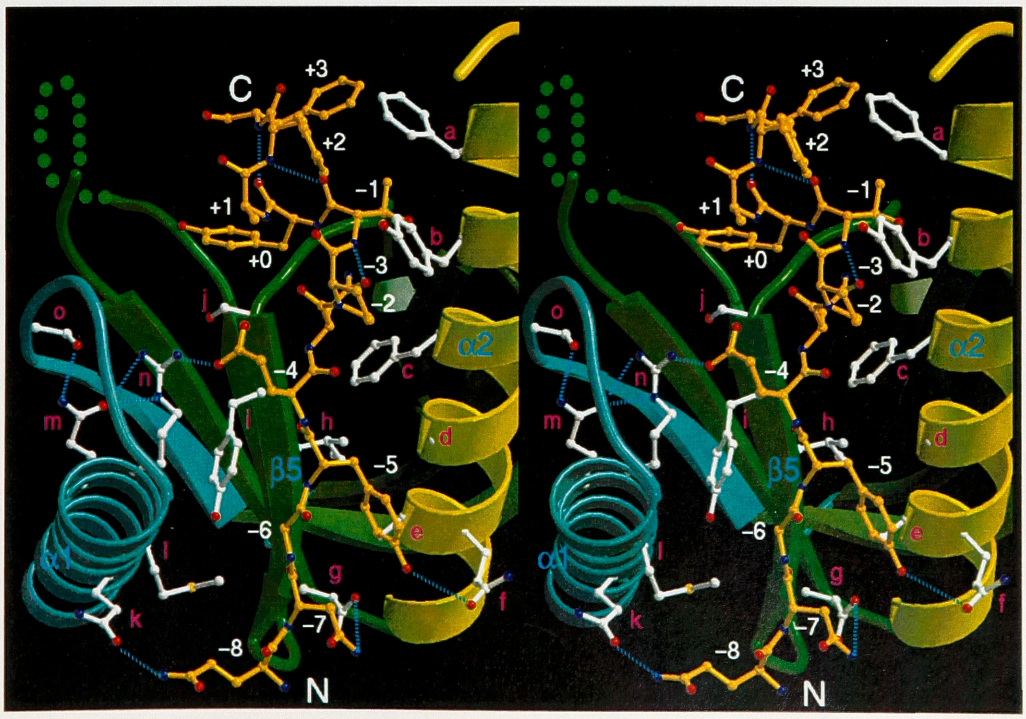
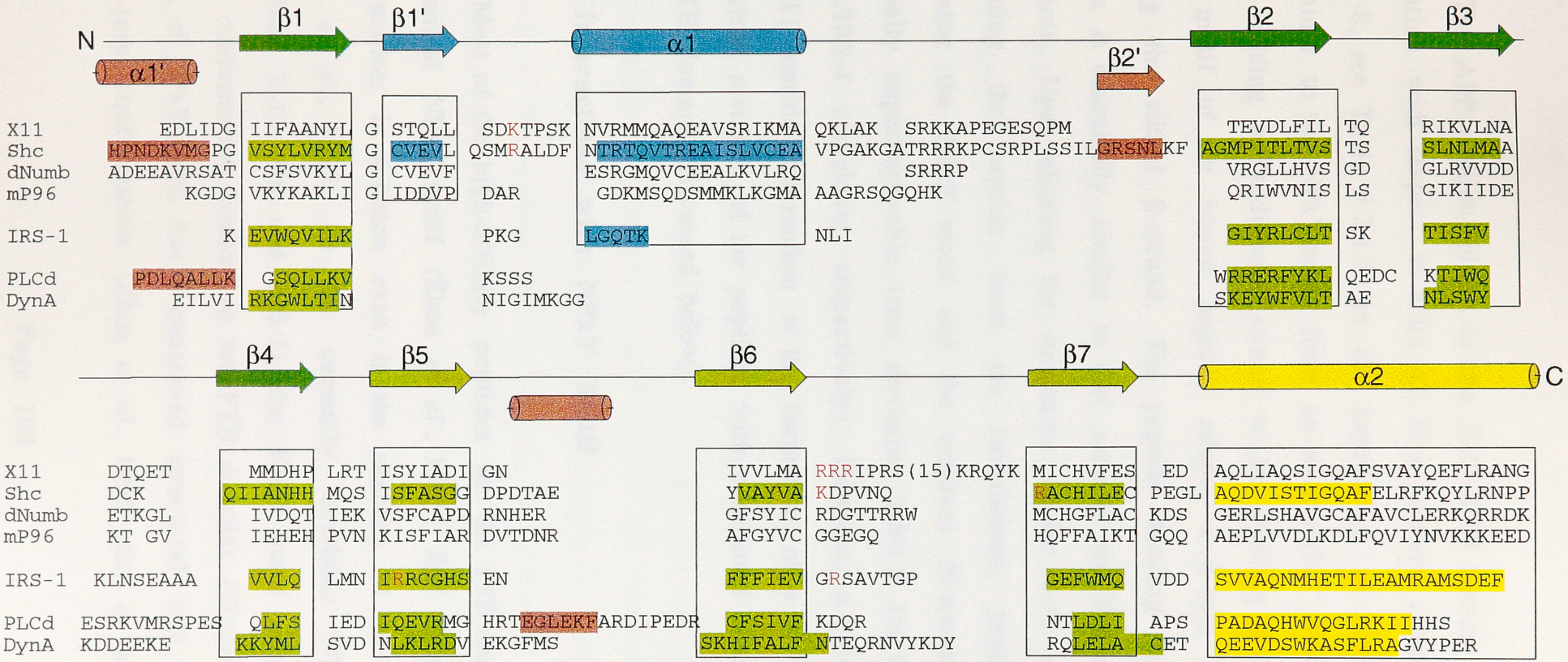

C
+3
+2
+1
+0
−1
−2
−3
−4
−5
−6
−7
−8
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
α2
β5
α1
N

Figure 3.2.2.3 Sequence alignment of the PTB and PH domains. The secondary structure elements of the X11 PTB are boxed and are shown schematically as colored arrows (for β-strand) and cylinders (for α-helix) above the sequence. Noted that the secondary structure elements for Shc, IRS-1 and two PH domains are different form those of X11 and are colored accordingly.


N
β1
β1'
α1
β2
β3
α1'
β2'
X11 EDLIDG IIFAANYL G STQLL SDKTPSK NVRMMQAQEAVSRIKMA QKLAK SRKKAPEGESQPM TEVDLFIL TQ RIKVLNA
Shc HPNDKVMGPG VSYLVRYM G CVEVL QSMRALDF NTRTQVTREAISLVCEA VPGAKGATRRRKPCSRPLSSILGRSNLKF AGMPITLTVS TS SLNLMAA
dNumb ADEEAVRSAT CSFSVKYL G CVEVF ESRGMQVCEEALKVLRQ SRRRP VRGLLHVS GD GLRVVDD
mP96 KGDG VKYKAKLI G IDDVP DAR GDKMSQDSMMKLKGMA AAGRSQGQHK QRIWVNIS LS GIKIIDE
IRS-1 K EVWQVILK PKG LGQTK NLI GIYRLCLT SK TISFV
PLCd PDLQALLK GSQLLKV KSSS WRRERFYKL QEDC KTIWQ
DynA EILVI RKGWLTIN NIGIMKGG SKEYWFVLT AE NLSWY
β4
β5
β6
β7
α2
C
X11 DTQET MMDHP LRT ISYIADI GN IVVLMA RRRIPRS(15)KRQYK MICHVFES ED AQLIAQSIGQAFSVAYQEFLRANG
Shc DCK QIIANHH MQS ISFASGG DPDTAE YVAYVA KDPVNQ RACHILEC PEGL AQDVISTIGQAFELRFKQYLRNPP
dNumb ETKGL IVDQT IEK VSFCAPD RNHER GFSYIC RDGTTRRW MCHGFLAC KDS GERLSHAVGCAFAVCLERKQRRDK
mP96 KT GV IEHEH PVN KISFIAR DVTDNR AFGYVC GGEGQ HQFFAIKT GQQ AEPLVVDLKDLFQVIYNVKKKEED
IRS-1 KLNSEAAA VVLQ LMN IRRCGHS EN FFFIEV GRSAVTGP GEFWMQ VDD SVVAQNMHETILEAMRAMSDEF
PLCd ESRKVMRSPES QLFS IED IQEVRMG HRTEGLEKFARDIPEDR CFSIVF KDQR NTLDLI APS PADAQHWVQGLRKIIHHS
DynA KDDEEKE KKYML SVD NLKLRDV EKGFMS SKHIFALF NTEQRNVYKDY RQLELA CET QEEVDSWKASFLRAGVYPER

The APP peptide binds to the X-11 PTB domain in a parallel orientation with respect to helix α2. Five peptide residues (residues -4 to -8, see Table 3.2.1.1 for the numbering of peptide residues) N-terminal to the NPxY motif form an antiparallel β-strand with β5 strand, adding an additional strand to the existing β-sandwich. The NPTY motif of the bound peptide adopts a β-turn conformation, capping the added β-strand. The peptide binding mode in these regions is essentially similar to that seen in the Shc and the IRS-1 structures. Upon aligning the structures of the X-11 and the IRS-1 (excluding the peptide from the calculation), peptide residues containing the NPxY motif and the residues N-terminal to it are essentially superimposable (rms deviation 1.5/1.9 for the backbones of Shc/IRS-1 complexes, respectively). However, a unique aspect of the X11/peptide interaction is the formation of a 3^{10} helix by the C-terminal residues of the peptide, with additional interactions with the PTB domain (discussed below).

3.2.3 Interactions with NPxY Motif

Most of the high-affinity peptides for the PTB domains contain a canonical NPxY motif (Zhou *et al.*, 1995) that adopts a β-turn conformation in solution even in the absence of the PTB domains (Trub *et al.*, 1995) In the currently available PTB structures (including X-11, Shc and IRS-1), the NPxY motif adopts a type I β-turn conformation binding to the PTB domains (Figure 3.2.3.1). The Asn_{-3} of NPxY motif has a conserved structural role for stabilizing the β-turn conformation (Zhou *et al.*, 1995; Eck *et al.*, 1996). The

carboxamide oxygen of the Asn-3 sidechain forms an intramolecular hydrogen bond with the backbone amide group of the peptide residue at -1 position and the nitrogen atom of the Asn sidechain is hydrogen bonded to the carbonyl groups of Leu 91 and Ile 94 (of X11 PTB). The selection for Asn at -3 position is due to the formation of these hydrogen bonds which are specific for an asparagine. The Asn-3 and hence the β-turn conformation appears to be the most important determinant for interacting with the PTB domains. A similar hydrogen bonding pattern was observed in the structure of the IRS-1/peptide complex (Eck *et al.*, 1996) and is expected to be conserved in other PTB/NPxY-peptide interactions as well. The Pro-2 also plays a role in stabilizing the β-turn conformation of the peptide ligand; however, its importance varies. Replacing Pro-2 with Ala decreases the binding affinity to the Shc PTB domain up to 50-fold depending on the sequence contexts of the peptide (Wolf *et al.*, 1995; Laminet *et al.*, 1996; Mandiyan *et al.*, 1996). In the X-11 structure, the pyrrolidine ring of Pro-2 packs against the benzene ring of Tyr 161 (from helix α2) while similar hydrophobic interactions are provided by the aliphatic portion of Arg 258 sidechain in the IRS-1 PTB domain (Eck *et al.*, 1996).

In the X11/APP peptide structure, the peptide is unphosphorylated. Despite the lack of a phosphate group, the Tyr+0 at the tip of the β-turn resumes a position that is almost identical to that of the bound phosphotyrosine residue observed in the IRS-1 and the Shc structures. The phenolic ring of the Tyr+0 forms hydrophobic interactions with the Cβ of Ser 95 in the same fashion as in the

Shc/phosphopeptide structure (Zhou *et al.*, 1995) while similar interactions are mediated by Arg 212 in the IRS-1/peptide structure (Eck *et al.*, 1996). Apart from this, no other specific interactions for this tyrosine residue is observed. As previous biochemical study suggested, the Tyr_{+0} is not important for the interaction as replacing the Tyr_{+0} with Ala results in no significant loss of binding affinity (Borg *et al.*, 1996).

Crystals of the X11 PTB in complex with tyrosyl-phosphorylated APP peptide have been obtained. However, a small contamination of unphosphorylated peptides in the crystals (<25%, as estimated by mass spectroscopy) and the resulting deterioration of electron density map for phosphotyrosine precluded detailed structural analysis. Nevertheless, it was found that the phosphorylated APP peptide binds to the X11 PTB domain in almost the same conformation as the unphosphorylated one. In addition, several basic residues including two of those that have been implicated in phosphotyrosine recognition of the Shc PTB domain (Zhou *et al.*, 1995) are located near the phosphotyrosine although they seem to be in suboptimized conformations needed for the interaction with phosphotyrosine (as seen in the unphosphorylated peptide complex). The high affinity interaction with unphosphorylated peptides is not an unique feature for the X11 PTB domain. The PTB domains of FE65 and Shc have been shown to mediate phosphotyrosine-independent interactions. Even in the cases of the Shc and IRS-1 PTB domains in which tyrosyl phosphorylation

Figure 3.2.3.1 Schematic depiction of contacts between the APP peptide and the X11 PTB domain. Residues of the X11 PTB domain are shaded green and the APP peptide residues are shaded red. The antiparallel packing of β-stands between peptide and PTB domain is indicated by the two shaded arrows in green (X11 PTB) and in brown (APP peptide).

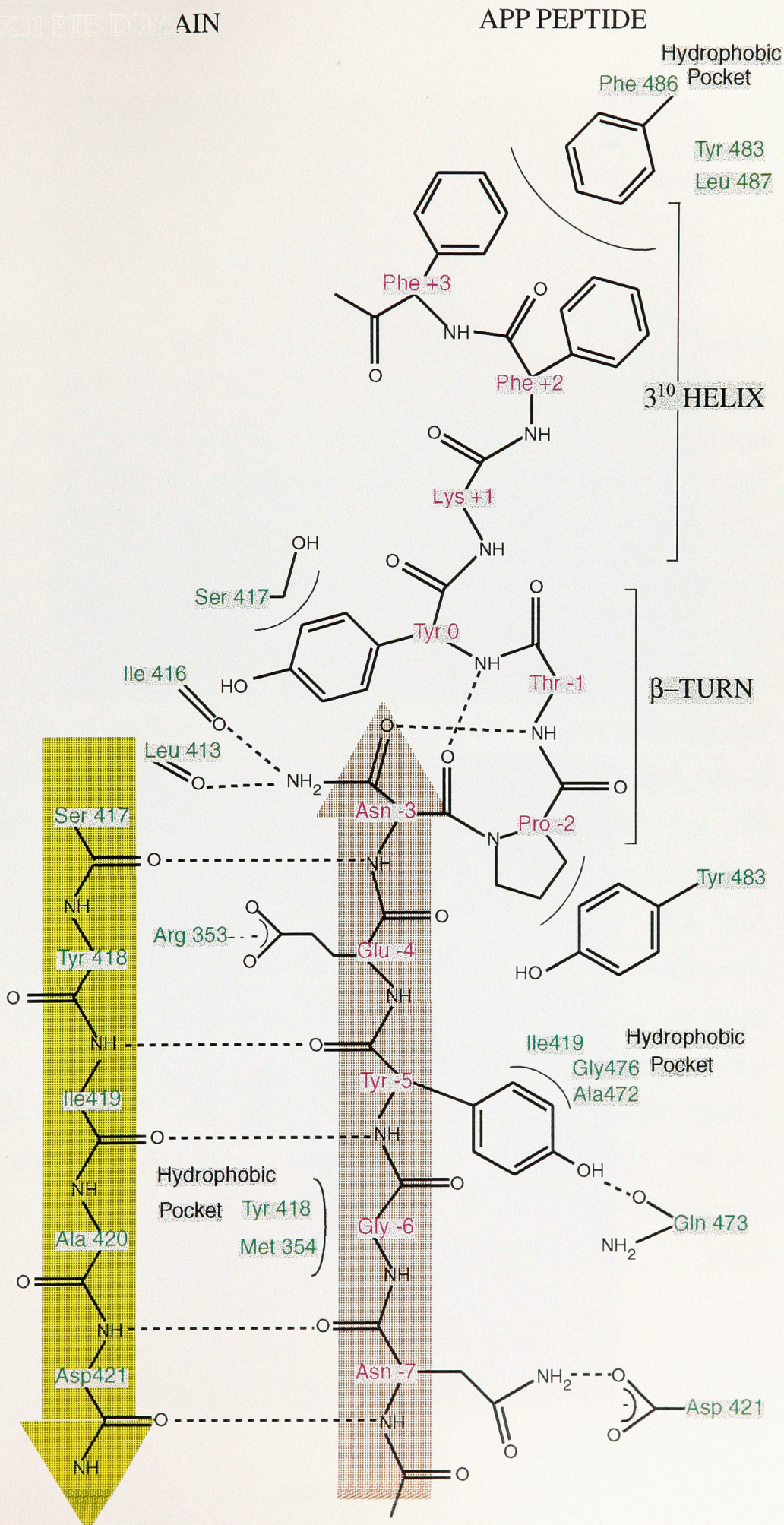
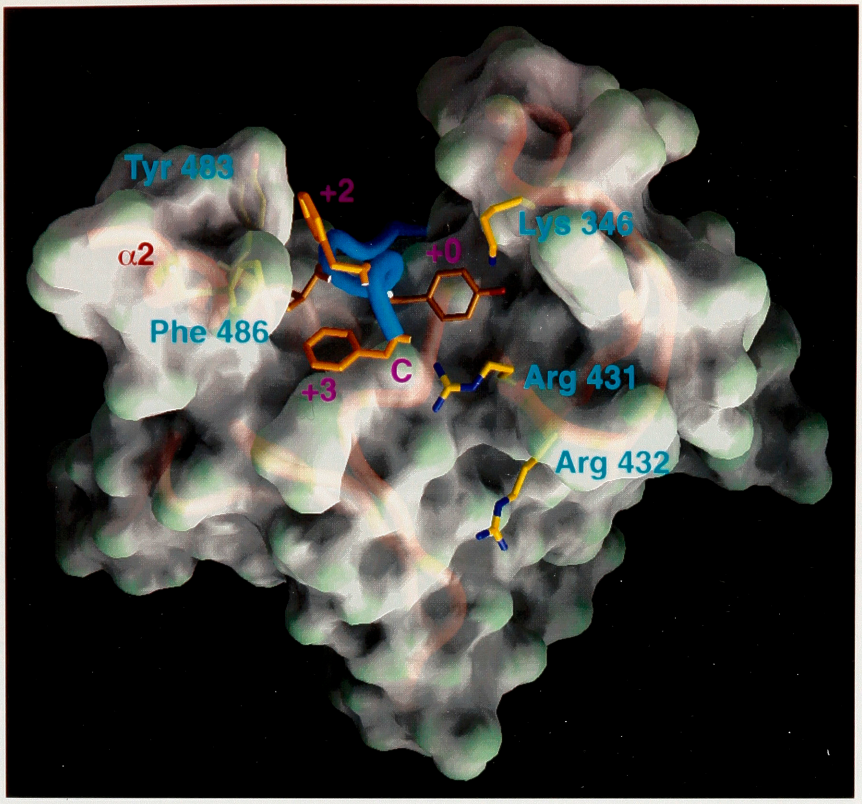
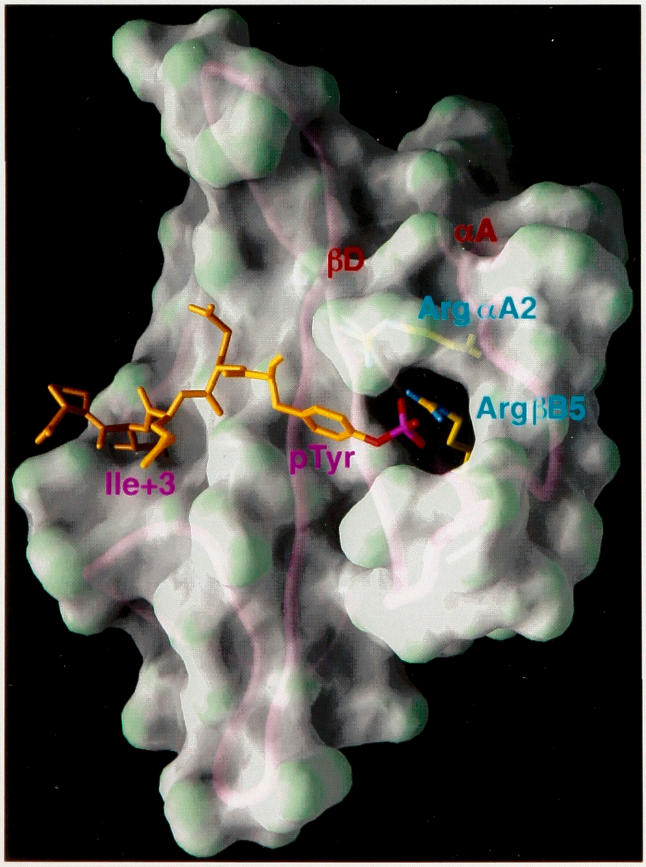

AIN
APP PEPTIDE
Hydrophobic
Pocket
Phe 486
Tyr 483
Leu 487
Phe +3
Phe +2
3[10] HELIX
Lys +1
Ser 417
Tyr 0
Thr -1
β–TURN
Ile 416
Leu 413
Asn -3
Pro -2
Tyr 483
Ser 417
Arg 353
Glu -4
Tyr 418
Ile419
Hydrophobic
Pocket
Gly476
Ala472
Tyr -5
Gln 473
Hydrophobic
Pocket
Tyr 418
Met 354
Gly -6
Ala 420
Asn -7
Asp421
Asp 421

Figure 3.2.3.2 Comparison of the phosphotyrosine/tyrosine binding sites of PTB domain and SH2 domain. (a) Molecular surface of the X11 PTB domain, showing the binding site for the tyrosine (+0) and the C-terminal region of the peptide. The polypeptide backbones of the bound peptide and the X11 PTB are shown in red and blue tubes, respectively. Also shown are the three basic residues (Lys 347, Arg 432, and Arg 433; removed from surface calculation) of the X11 PTB that are potentially involved in the phosphotyrosine binding. The sidechains of Phe(+2) and Phe(+3) of the peptide (colored in yellow) pack against the sidechains of Tyr 484 and Phe 487 of the X11 PTB (colored in gold). (b) Molecular surface of the unliganded Src SH2 domain showing the deep buried Arg βB5 sidechain (removed from surface calculation). For comparison, also shown is the sidechain of Arg αA2 that is also involved in phosphotyrosine recognition but is solvent exposed. To indicate the position of the phosphotyrosine binding site, the high affinity phosphopeptide (colored in gold), taken from the Src/Y*EEI complex structure, is placed in the approximate position. The polypeptide backbone of the Src SH2 domain is shown in purple tubes.


Tyr 483
+2
Lys 346
α2
+0
Phe 486
C
+3
Arg 431
Arg 432


αA
βD
Arg αA2
Arg βB5
pTyr
Ile+3

is often critical for the interaction, two distinct sets of residues are used for phosphotyrosine binding. As suggested, the Shc and IRS-1 PTB domains might independently acquire the ability of binding to phosphotyrosine (Eck *et al.*, 1996). Taken together, it appears that the most critical element for interacting with the PTB domains is the secondary structure elements: β-turn conferred by the NPxx sequence and β-strand formed by the N-terminal residues. Despite the fact that the PTB domain was originally identified as an alternative for the SH2 domain, the phosphotyrosine-binding is not a conserved feature for all PTB domains.

3.2.4 The C-terminal Specificity

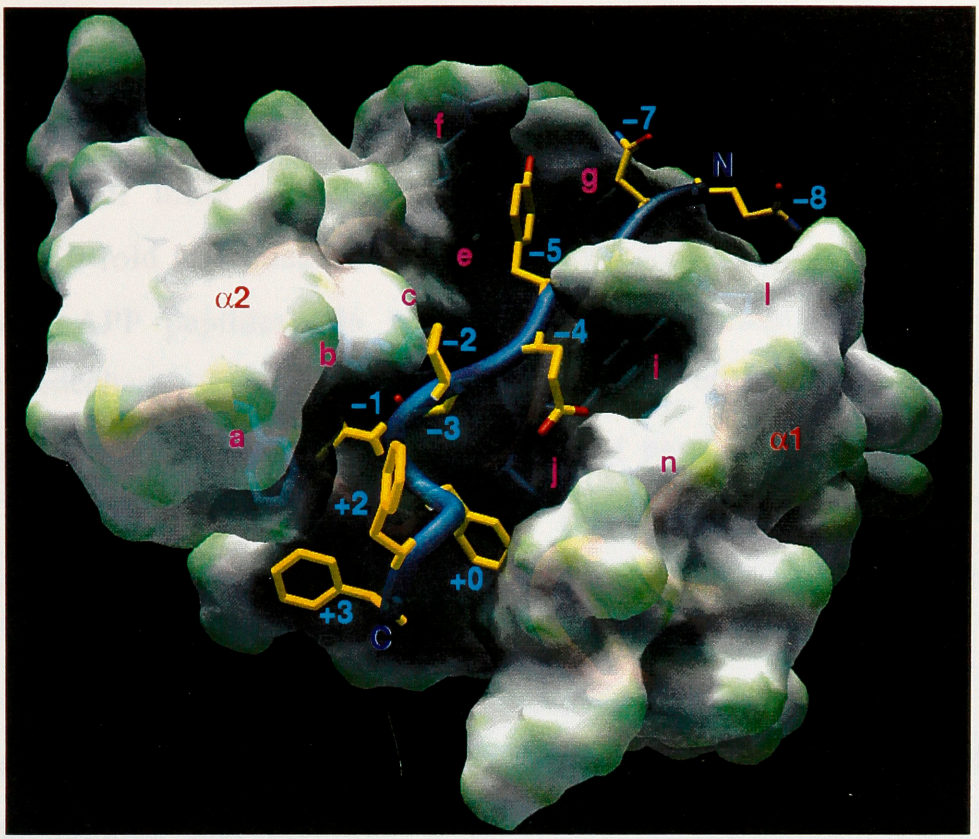
A striking feature of the X-11 structure is a 3^{10} helix that is formed by C-terminal 5 peptide residues (residues +1 to +5), capping the C-terminal end of the peptide. This finding is unexpected since no secondary structure has been observed for the peptide residues C-terminal to NPxY motif in previously determined PTB structures. In contrast to previous notion that the peptide residues N-terminal to NPxY* motif determine the binding specificity, the C-terminal residues (Phe +2 and +3) of the APP peptide contribute positively to binding affinity. Mutation of either Phe_{+2} or Phe_{+3} to Ala decreases the binding affinity by 10 fold. In the X11/APP peptide structure, the phenolic rings of Phe +2 and +3 of APP peptide pack against a hydrophobic surface that is formed by a number of residues arose from helix α2 ——particularly, Phe 464 (Figures 3.2.2.2 & 3.2.3.2a). As a result of these additional interactions, the total interaction

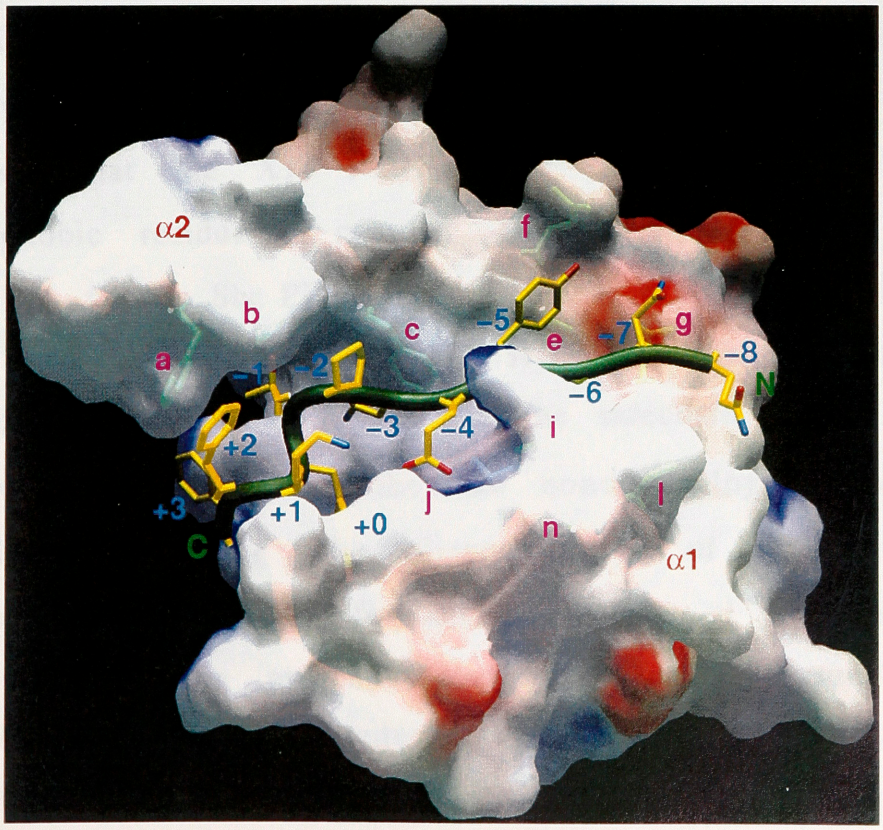
interface between the APP peptide and the X11 PTB domain is 2,000 Å^2 which is larger than those for Shc (~1,800 Å^2) and IRS-1 (~1,300 Å^2) structures. The C-terminal α-helix (α2) is three turns longer than that of the Shc PTB domains although they share considerable sequence homology (63%, 15 out of 24 residues) in this region. The Phe 464 is located at the extended region of helix α2 and its interaction with the peptide residues at +2 and +3 positions might have stabilized the helical conformation. Interestingly, the extension of the helix α2 induced by ligand binding has been documented in the case of the IRS-1 PTB domain for which both liganded and apo structures are available (Eck *et al.*, 1996). No specific interaction mediated by other C-terminal peptide resides, including residues +1, +4 and +5, is observed. This suggests that these residues are important for neither binding specificity nor affinity.

3.2.5 The N-terminal Specificity

The bound peptide fits snugly into a groove on the peptide binding surface of the X11 PTB domain. The floor of this cleft is created by the strand β5 which forms an antiparallel β-strand interactions with the N-terminal peptide (residues -4 to -8) through backbone hydrogen bonds. The importance of the peptide residues -4 to -8 for the sequence specific recognition has been described. On one side of the bound peptide, the sidechains of odd numbered residues (-3, -5, and -7) point toward the C-terminal helix (α2) and form interactions with residues from this helix. On the other side, the sidechains of even numbered residues (4, -6 and -8) point toward

Figure 3.2.5.1 Molecular surface of the X11 PTB/APP peptide complex. The surface is calculated with the peptide removed (using GRASP, Nicholls, *et. al.*, 1991) and is colored according to surface curvature (a) or local electrostatic potential (b) (Gilson, *et. al.*, 1988) calculated in the absence of peptide, ranging from deep blue (most positive region) to deep red (most negative). The peptide (backbone shown as green tubes) binds to a deep groove of the X11 PTB domain. Important sidechains in the X11 PTB domain are shown in green sticks and labeled as in Figure 3.2.2.2.


f
g
−7
N
−8
e
−5
α2
c
l
b
−2
−4
i
−1
−3
a
α1
n
j
+2
+0
+3
C


α2
f
b
c
−5
e
−7
g
a
−1
−2
−8
N
−6
i
−3
−4
+2
+3
+1
+0
j
l
n
C
α1

strand β5 and helix α1. Two residues at the N-terminus of the APP peptide appear to be important for optimal binding since the removal of these two residues decreases the binding affinity by more than 10-fold (as judged by comparing the affinities of 14mer and 10mer APP peptides). In addition to the backbone hydrogen bonding, the sidechain of Asn-7 also forms hydrogen bonds with sidechains of Asp 99 and Gln 147. The polar nature of peptide residues at position -7 is in contrast to the high affinity peptide for Shc and IRS-1 which favor hydrophobic residues at this position (Wolf *et al.*, 1995).

In addition, X11 shows preference in binding to two peptide residues: Tyr-5 and Glu-4. Previous biochemical study has demonstrated that mutation of the Tyr-5 to Gly abolishes the binding (Borg *et al.*, 1996). The phenolic ring of the Tyr-5 sidechain pack against Cα of Gly 154 and the sidechains of Ala 150 and Ile 97 (Figure 3.2.2.2). The hydrophobic nature of the Tyr-5 binding site suggests that the X11 PTB domain might select for a large hydrophobic residue at position -5. Such a preference has been described for the Shc PTB domain (Trub *et al.*, 1995; Laminet *et al.*, 1996; van der Geer *et al.*, 1996). Hydrogen bonding between the hydroxyl group of the Tyr-5 and the sidechain of Gln 151 was observed only in one of the two noncrystallographically-related complexes, implying the dispensability of this interaction. Furthermore, selectivity for acidic residues at -4 position was suggested by the present biochemical and structural studies. Replacing Glu-4 with Ala reduced binding affinity by 5-fold. In the structure, Glu-4 sidechain formed a salt bridge with the guanidio

group of Arg 31 sidechain which is held in the position by salt bridges with the sidechains of Gln 34 and Ser 22 and is mostly buried upon ligand binding. The salt bridge mediated by the sidechain of Glu$_{-4}$ was observed in the original experimental electron density map and remained throughout the refinement processes. Selectivity for an acidic residue at position -4 has been described for the Shc PTB domain also (Laminet *et al.*, 1996). Finally, the selectivity for a glycine at position -6 is suggested by the present structure. Met 354 and Tyr 418 form hydrophobic interactions with the Cα of Gly$_{-6}$. Given these steric constraints imposed, this position is expected to accommodate only small hydrophobic residue (Figures 3.2.2.2 & 3.2.5.1).

In conclusion, the X11 PTB domain shows N-terminal selectivity rather different from those of Shc and IRS-1 PTB domains. In addition to the NPxY* motif, the high-affinity peptides for Shc and IRS-1 often contain hydrophobic residues at position -4 to -8 (although Shc seems to prefer an aromatic residue at position +1 (Laminet *et al.*, 1996)). In contrast, the X11 PTB domain appears to select for polar residues at position -4 and -7, large hydrophobic residues at -5, and small hydrophobic residue at position -6. Furthermore, selectivity of C-terminal residues, +2 and +3 for phenylalanine are unique to the X11 PTB domain. It was noted that each peptide residue contributes rather modest selectivity (5-10); however, collectively, they would give rise to significant binding specificity in binding to APP peptide. As the region encompassing the NPTY motif accounts for the binding of β-APP to the X11 protein, the

interactions described above are likely responsible for the remarkable specificity of the X11 PTB domain in binding to β-APP.

3.2.6 Summary

The crystal structure of the phosphotyrosine binding domain (PTB) of the X11 protein has been determined, in complex with unphosphorylated peptides corresponding to a region of β-amyloid precursor protein (βAPP) that is required for receptor internalization. The mode of binding to X11 of the unphosphorylated peptides, which contain an NPxY motif, resembles that of phosphorylated peptides bound to the Shc and IRS-1 PTB domains. Eight peptide residues make specific contacts with the X11 PTB domain, and they collectively achieve high affinity (K_D = 0.32 μM) and specificity. These results suggest that, in contrast to the SH2 domains, the PTB domains are primarily peptide-binding domains that have, in some cases, acquired specificity for phosphorylated tyrosines.

Figure 3.3.1.1. Determination of the equilibrium dissociation constant for the Nef/Hck-SH3 interaction using surface plasmon resonance (SPR). (a) The response functions (sensorgrams) shown illustrate specific binding and dissociation profiles obtained after an injection of Nef protein at various concentrations was added to the buffer flowing over a biosensor chip coated with GST-Hck-SH3. The data in these sensorgrams and **(b)** in the Scatchard analysis are expressed as corrected response units (cRU) as the contribution by the bulk refractive index (background) has been subtracted. For clarity, only the sensorgrams obtained using 0.08, 0.2, 0.4, 0.8, 2 and 6 μM Nef are shown, whereas the SPR response values from all ten concentrations of Nef tested (ranging 0.06 to 6 μM) were included in the Scatchard analysis. The data is best-fit by a line with equation y=-0.255x+173.6, the coefficient of determination (R^2) being 0.99.

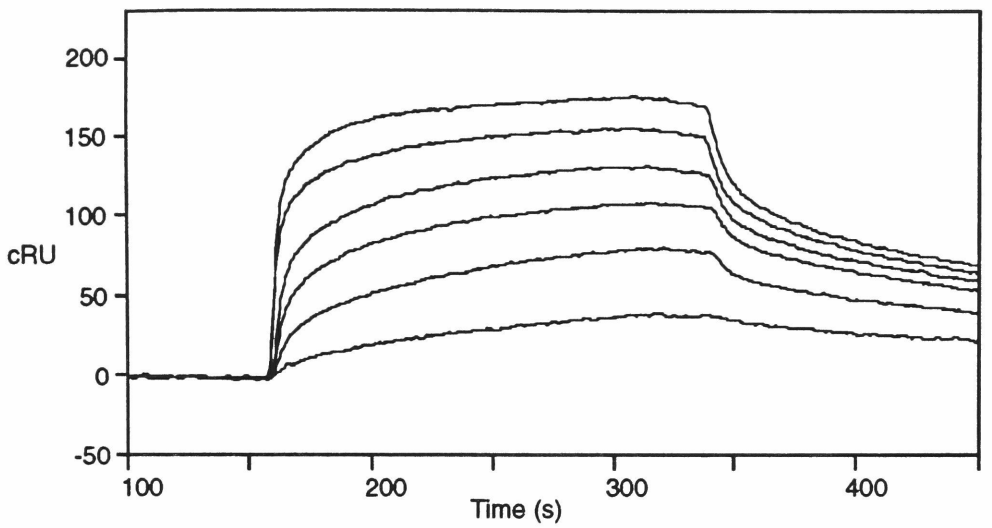
3.3 Biochemical Characterization of HIV-1 Nef - SH3 Domain Interaction

Previous studies by Saksella and co-workers indicated that HIV-1 Nef protein specifically interacts with the SH3 domains of certain Src family tyrosine kinases (such as Hck and Lyn, but not closely related Fyn and Src). The ability of Nef to interact with the SH3 domains is required for optimal virus replication *in vivo*, suggesting its biological relevance. In order to further characterize the interaction between Nef and Src family SH3 domains, I have used quantitative assays to examine the structural determinants of the differential affinity of Hck and Fyn SH3 domains.

3.3.1 High-Affinity Binding of Nef to Hck-SH3 Revealed by SPR Analysis

Two different techniques were used to measure the affinity for the HIV-1 Nef/Hck-SH3 domain interaction: surface plasmon resonance (SPR) measurements and isothermal titration calorimetry (ITC; Ladbury *et al.*, 1995; O'Shannessy *et al.*, 1993; Wiseman *et al.*, 1989). The binding of wild-type and mutant Nef proteins to various SH3 domains was also examined. These SH3 domains were expressed as glutathione-S-transferase (GST) fusion proteins, later biotinylated and immobilized onto streptavidin-coated chips of a BIAcore

A.


200
150
100
50
0
-50
cRU
100
200
300
400
Time (s)

B.

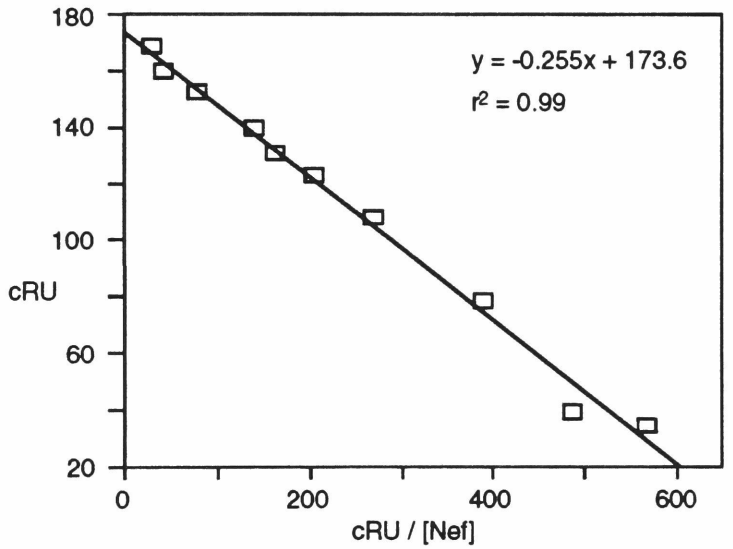
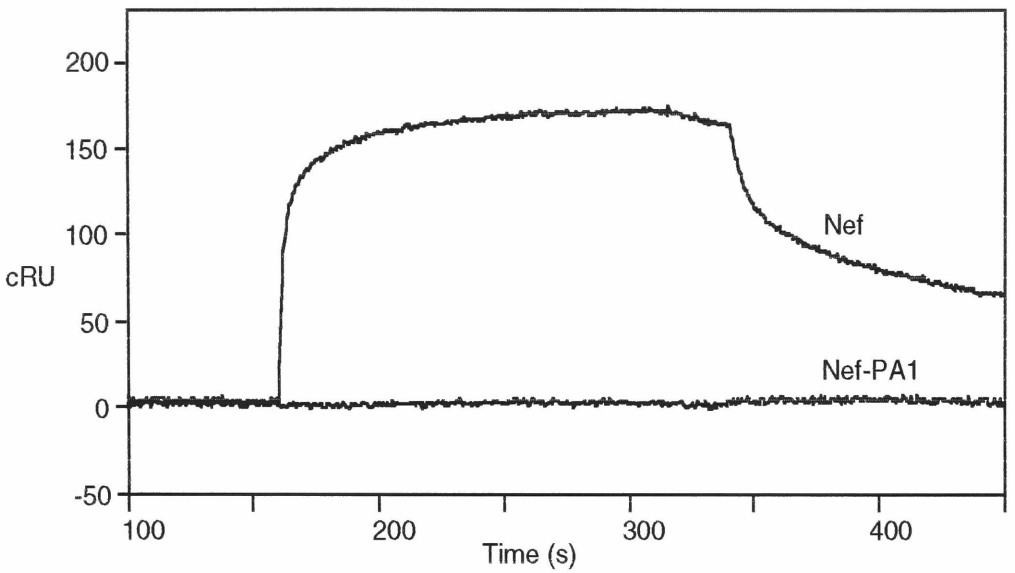
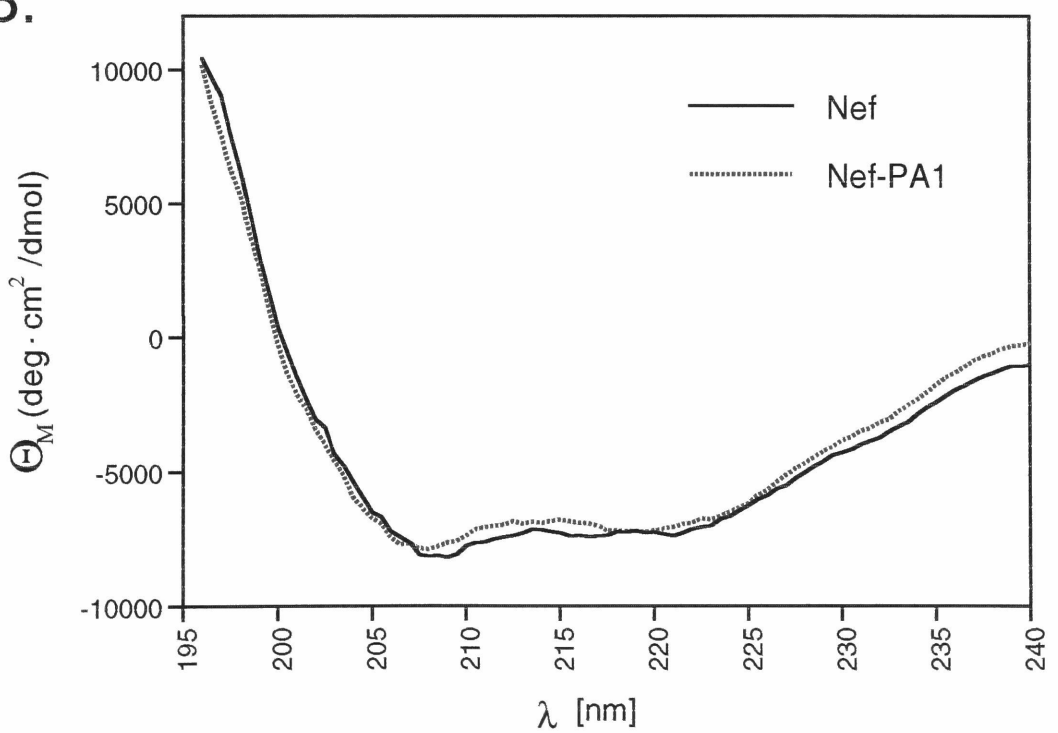

y = -0.255x + 173.6
r2 = 0.99
180
140
100
60
20
cRU
0
200
400
600
cRU / [Nef]

Figure. 3.3.1.2 Comparison of binding of wild-type Nef and the PxxP defective mutant Nef-PA1 to Hck-SH3. (**a**) SPR profiles using Nef (upper line) at a concentration close to saturation (8 μM) and Nef-PA1 (lower line) at the same concentration are shown. Both sensorgrams were derived and corrected for the bulk refractive index effect as in Figure 3.3.1.1. (**b**) Circular dichroism spectra of Nef (solid line) and Nef-PA1 (dashed line). Spectra of Nef and Nef-PA1 were recorded at 15°C using 50 μM protein concentration in the same buffer as in SPR experiments shown in Figure 3A. 5 individual measurements were averaged and expressed in molar mean residue ellipticity (Θ_M) value.

A.


200
150
100
50
0
-50
cRU
Nef
Nef-PA1
100
200
300
400
Time (s)

B.


10000
5000
0
-5000
-10000
Θ_M (deg · cm^2 /dmol)
Nef
Nef-PA1
195
200
205
210
215
220
225
230
235
240
λ [nm]

biosensor apparatus. The full-length wild-type Nef protein produced for these experiments corresponds to the HIV-1 NL4-3 Nef, but has a threonine-to-arginine mutation at position 71 in order to mimic the amino acid sequence that occurs in this region of most Nef alleles that have been obtained directly from patients (Huang *et al.*, 1995; Shugars *et al.*, 1993). Arginine 71 is adjacent to amino acid residues of Nef that are involved in SH3-binding and may influence this interaction (Saksela *et al.*, 1995). As a negative control, I also produced and tested a mutant HIV-1 NL4-3-derived Nef protein (Nef-PA1) which carries a double proline-to-alanine substitution that disrupts the two internal prolines of the conserved Nef tetraproline repeat. In a filter binding assay, Nef-PA1 is unable to bind Hck-SH3 (Saksela *et al.*, 1995).

Figure 3.3.1.1a shows SPR profiles obtained after flowing buffer containing various concentrations of Nef over a biosensor chip coated with Hck-SH3. Corresponding SPR signals obtained in parallel by testing identical concentrations of Nef on biosensor chips without SH3 proteins (to control for bulk refractive index effects) have been subtracted from these values. Therefore these profiles, as well as those shown later, reflect specific association between Nef and the immobilized SH3 protein. Based on these measurements, the equilibrium dissociation constant (K_D) for the interaction between Nef and Hck-SH3 was determined by Scatchard analysis of the SPR response at equilibrium (Figure. 3.3.1.1b), and was found to 250 nM. The linearity of the Scatchard plot over a wide range of tested Nef concentrations (0.06~6 μM) indicates that Nef/Hck-SH3-binding

represents a simple bimolecular interaction. Judged by the sensorgram (Figure 3.3.1.1a), the on-rate for Nef binding to Hck-SH3 was very rapid. Due to this rapid on-rate, significant re-binding of dissociated Nef probably also occurred under these conditions. Therefore, it was not possible to measure reliably the kinetic parameters (off- and on-rate) with current experimental setup, since both would be artificially influenced by experimental variables of the system (such as the flow-rate, concentration of Nef, and the surface density of Hck-SH3). By contrast, the Scatchard analysis used to determine the K_D values is not sensitive to these effects since it relies only on the SPR response at equilibrium.

The low K_D value determined for binding of Nef to Hck-SH3 (250 nM) indicates an affinity that is significantly greater than previously reported SH3/PxxP interaction (K_D 1~50 μM). To confirm this unusually high affinity, an independent methodology was used. I have collaborated with Lemmon (at NYU) to examine the Hck/Nef interaction by isothermal titration calorimetry (ITC; Lemmon and Ladbury, 1994; Wiseman *et al.*, 1989). In ITC experiments, free Hck-SH3, rather than the GST-Hck-SH3 fusion protein was used, while the other ITC experimental conditions, including buffer composition and temperature, were essentially identical to those used in SPR studies. The heat released upon each of the sixteen sequential injections of Hck-SH3 into a Nef solution was integrated and normalized per mol of injectant. The plot showing the heat released per mole of Hck-SH3 injected against the SH3:Nef molar ratio takes the form of a sigmoidal titration curve with its midpoint

occurring at a ratio of 1:1. The data were fit to the curve shown using a non-linear least squares algorithm (Wiseman *et al.*, 1989), and the enthalpy per mole of ligand added (ΔH), the association constant K_B ($=1/K_D$) and stoichiometry were derived. The K_D value (188 $\pm$34 nM), obtained using ITC is comparable to that observed by SPR, and the stoichiometry for this interaction (1:0.89) was found to be very close to a ratio of 1:1. The ΔH for this interaction at 25 ^oC was -12.75 Kcal/mol, which together with the experimentally derived K_B value gives an entropy change ΔS^o= -12.0 cal/K/mol. The large negative enthalpy change and the small unfavorable entropy change observed for the Nef/Hck-SH3 interaction indicate that under the experimental conditions used this interaction is mainly driven by exothermic enthalpy.

In contrast to wild type Nef protein, no binding to Hck-SH3 was observed when the mutant Nef-PA1 protein was tested using SPR (Figure 3.3.1.2a). This indicates that the high-affinity of the Nef/Hck-SH3 interaction is critically dependent on an intact Nef PxxP motif. To confirm that this lack of binding was due to the defective PxxP motif of Nef-PA1, rather than to a non-specific structural alteration of the protein caused by the introduced proline-to-alanine changes, the CD (circular dichroism) spectra of the wild-type Nef and Nef-PA1 were measured and compared. As can be seen in Figure 3.3.1.2b, under experimental conditions similar to those used for the SPR measurements, the CD spectra of Nef and Nef-PA1 were virtually identical, excluding the possibility of gross structural abnormalities in Nef-PA1. It should be noted that Nef proteins used in all

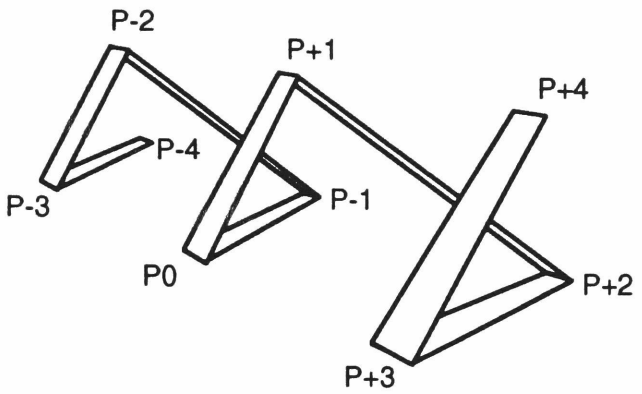
experiments were highly purified and free of GST. The lack of apparent structural alterations in Nef-PA1 by CD is in agreement with our previous observations, which indicate that, despite the loss of SH3 binding and hence the ability to increase HIV replication, Nef-PA1 can still mediate down-modulation of surface expression of the CD4 molecule (Saksela *et al.*, 1995). Moreover, the CD spectra obtained for these Nef proteins appear identical to those published by Wittinghofer and colleagues in their biochemical characterization of the HIV-1 NL4-3 Nef protein (Wolber *et al.*, 1992).

In summary, these data show that native Nef protein can bind in solution to Hck-SH3 domain with very high affinity, and that this binding is absolutely dependent on a functional Nef PxxP site.

3.3.2 Sequence Comparison of SH3 Domains and Molecular Modeling

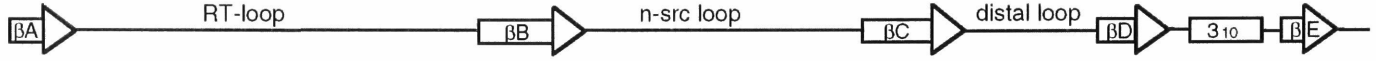
In an effort to understand the structural basis of specific SH3-binding by Nef, the PxxP motif of Nef was aligned with other PxxP motifs (Figure 3.3.2.1a), as well as the sequence of Hck-SH3 with those of other SH3 domains (Figure 3.3.2.1b). Figure 3.3.2.1a shows an alignment of various SH3 binding ligands adapted from Lim *et al* (1994). On the basis of these alignments, a model of the Nef-PxxP peptide in complex with Hck-SH3 was generated (Figure 3.3.2.1c). Based on the molecular modeling and experimental data, my attention was drawn to the two middle prolines of the HIV-1 Nef tetraproline repeat (which define the PxxP motif) as well as the

Figure. 3.3.2.1 (a) Alignment of a consensus Nef PxxP motif (RPxVPLR; amino acids 71-77 in HIV-1 NL4-3 Nef) with other peptides binding to different SH3 domains in "plus" or "minus" orientation. This alignment is generated according to Lim *et al.* (1994) and Feng *et al.* (1994). The Nef residues which are highly conserved among primate immunodeficiency viruses are indicated in boldface, whereas the poorly conserved residue in position P_1 is indicated by "x". The positions P_{-3}, P_{-1}, P_0, P_2, and P_3 contain the ligand residues which interact with the conserved hydrophobic residues forming the binding interface of SH3 domains. The spacing of positions are also shown in the ribbon diagram representing a left-handed PP-II helix.



Orientation		P-3	P-2	P-1	P0	P1	P2	P3		Origin	SH3 domain
+	N-	R	K	L	P	P	R	P	-C	*Screen	PI3K
	N-	R	A	L	P	P	L	P	-C	*Screen	Src
	N-	M	P	P	P	L	P	P	-C	3BP1	Abl
-	C-	R	P	P	V	P	P	P	-N	mSos1-1	Grb2/Sem5
	C-	R	R	P	L	P	P	H	-N	*Screen	Src
	C-	R	S	P	V	P	P	A	-N	Dynamin	p85 PI3K
	C-	**R**	L	**P**	**V**	x	**P**	R	-N	Nef	Hck

Figure. 3.3.2.1 (b) Alignment of SH3 domain amino acid sequences. The indicated secondary structure is based on the crystal structure of the Fyn-SH3 domain (Noble *et al.*, 1993). The five antiparallel β-sheets (A-E) forming the SH3 structure, as well the loops between them, are indicated. The amino acid identity to Hck in this region is indicated as a percentage, and by purple coloring in the alignment. Amino acid similarity (D/E, K/R, S/T, W/F/Y, and A/L/I/V were considered similar) is indicated by light blue coloring. The conserved acidic (b, d, f, g) and hydrophobic (a, c, h, i, j, k, l) residues implicated in ligand binding are indicated on top of the amino acid alignment.



	a b c	d	e	f	g	h	i	j	k	l	Sequence identity (%)
Hck	IVVA	LYDY	EAIHHE	DLSFQKGD	QMVVLE	ES	GE	WWKARS	LATRK	EGYI	PSNY VAR VD	100
Fyn	LFVA	LYDY	EARTED	DLSFHKGE	KFQILN	SS	EGD	WWEARS	LTTGE	TGYI	PSNY VAP VD	62
Lyn	IVVA	LYPY	DGIHPD	DLSFKKGE	KMKVLE	EH	GE	WWKAKS	LLTKK	EGFI	PSNY VAK LN	71
Lck	LVIA	LHSY	EPSHDG	DLGFEKGE	QLRILE	QS	GE	WWKAQS	LTTGQ	EGFE	PFNF VAK AN	55
Src	TFVA	LYDY	ESRTET	DLSFKKGE	RLQIVN	NT	EGD	WWLAHS	LSTGQ	TGYI	PSNY VAP SD	53
Yes	IFVA	LYDY	EARTTE	DLSFKKGE	RFQIIN	NT	EGD	WWEARS	IATGK	NGYI	PSNY VAP AD	62
Fgr	LFIA	LYDY	EARTED	DLTFTKGE	KFHILN	NT	EGD	WWEARS	LSSGK	TGCI	PSNY VAP VD	55
Blk	FVVA	LFDY	AAVNDR	DLQVLKGE	KLQVLR	ST	GD	WWLARS	LVTGR	EGYV	PSNF VAP VE	55
Csk	ECIA	KYNF	HGTAEQ	DLPFCKGD	VLTIVA	VTK	DPD	WYKAKN	KV GR	EGII	PANY VQK RE	33
Abl	LFVA	LYDF	VASGDN	TLSITKGE	KLRVLG	YNH	NGE	WCEAQT	KN G	QGWV	PSNY ITP VN	40
CrkN	YVRA	LFDF	NGNDEE	DLPFKKGD	ILRIRD	KPE	EQ	WWNAED	SE GK	RGMI	PVPY VEK YR	36
Sem5	EAVA	EHDF	QAGSPD	ELSFKRGN	TLKVLN	KDE	DPH	WYKAEL	D GN	EGFI	PSNY IRM TE	36
PI3K	QYRA	LYDY	KKEREE	DIDLHLGD	ILTVNK	GSL(14)EIG		WLNGYN	ETTGE	RGDF	PGTY VEY IG	31

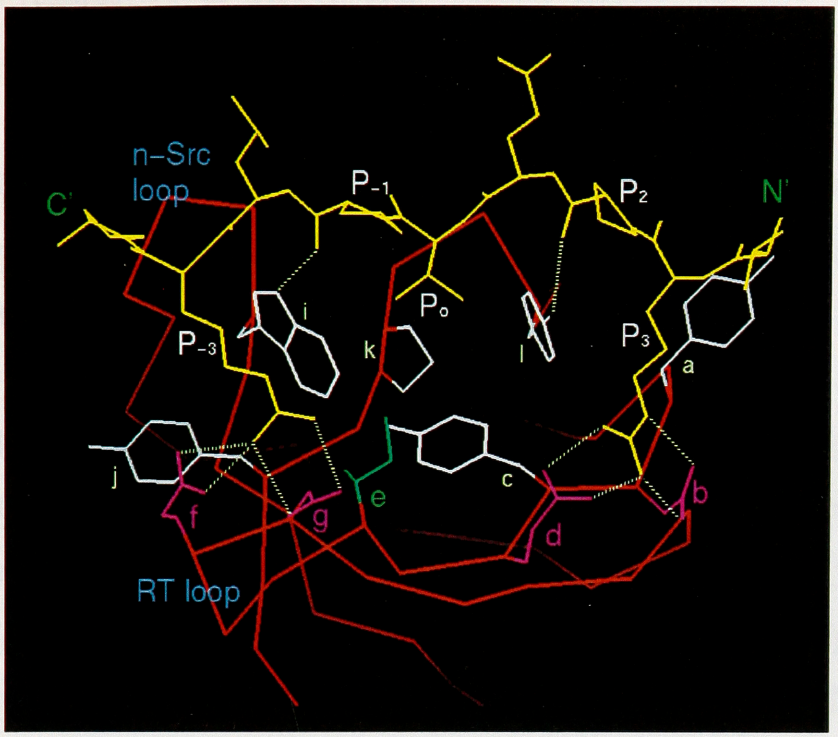


Figure 3.3.2.1 (C) Molecular model of interactions between Hck-SH3 and the Nef PxxP motif. The C^{α} backbone of the Hck-SH3 is shown in red and the Nef PxxP motif (sequence VRPQVPLRP, modeled in the "minus" orientation, right to left) is shown in yellow. The conserved hydrophobic residues (a, c, h, i, j, k, and l) forming the binding surface of Hck-SH3 domain are shown in white, except for the Phe residue 77 (h) which is omitted for clarity. The acidic residues (f and g) in RT-loop which are shown to interact with the arginine residue at the P_{-3} position of Nef PxxP motif and the acidic residues (b and d) which are modeled to interact with the arginine residue at the P_3 position are shown in purple. The isoleucine residue (e) in the RT-loop which is implicated in the interaction with Nef is shown in blue.

arginine residue 77 of HIV-1 Nef (strictly conserved among different primate immunodeficiency viruses). Together, they would define the binding of the Nef-PxxP peptide in the "minus"-orientation. The Arg-77, modeled to fit in the P_{-3} binding pocket, would thus be involved in interactions with acidic residues in the RT-loop of Hck-SH3 (residues f and g in Figures. 3.3.2.1b and c).

The critical role of the two internal prolines is supported by data from filter-binding experiments indicating that mutation of either one of them alone (PVRAQVPLRP or PVRPQVALRP) abolishes binding to Hck-SH3. In addition, a peptide overlapping the corresponding region in SIVmac239 Nef (LVGISVRPKVPLRTMSYK) can bind to Hck-SH3, despite its lack of both of the external prolines (data not shown). Moreover, a hydrophobic residue (V74) assigned to the binding pocket P_0 is also extremely well conserved among Nef proteins encoded by different viruses (including SIVmac239). Thus, this valine and the arginine in the P_{-3} pocket form the two asymmetric interactions defining the "minus"-orientation (Feng *et al.*, 1994; Lim *et al.*, 1994).

The structural model (Figure 3.3.2.1c) described here was generated in a simplistic manner and relies on the strong conservation of three dimensional structure in SH3 domains (Wu *et al.*, 1995). In particular, the model is based on the structures of the highly homologous Fyn and Lck SH3 domains, as well as the Crk-SH3/peptide complex (Noble *et al.*,. 1993; Eck *et al.*, 1994; Wu *et al.*, 1995). The other arginine residue in this region of Nef (R71) was

modeled in position P_3 so that the aliphatic region of the side chain packs against the face of two conserved tyrosines residues 66 and 111 (a and l) of the Hck-SH3 domain. The guanido group of R71 forms hydrogen bonds with aspartic acid 67 (b) and glutamic acid 69 (d). Unlike the lab-adapted strain HIV-1 NL4-3, almost all Nef genes isolated from patients encode an arginine or a lysine at position 71 (Huang *et al.*, 1995; Shugars *et al.*, 1993), and an arginine residue in this position of Nef (as compared to Thr) increases its binding to Hck-SH3 (Saksela *et al.*, 1995).

3.3.3 RT-loop Determines the Differential Affinity of Hck and Fyn SH3 Domains

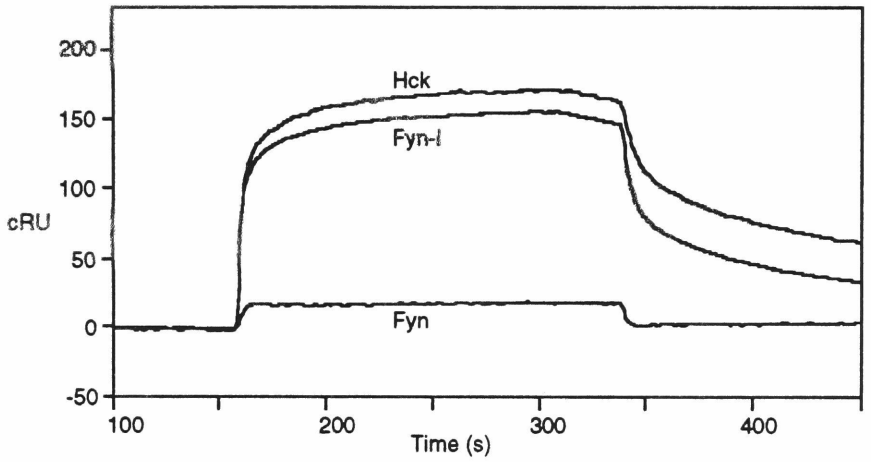
The SH3 domains of Hck and Lyn can bind to Nef whereas the highly homologous SH3 domains of Fyn or Lck do not (Saksela *et al.*, 1995). Based on (i) the molecular model of the Hck-SH3/Nef-PxxP complex (Figures 3.3.2.1a and c), (ii) conservation analysis (Figure 2.3.1.1), and (iii) the alignment of the amino acid sequences of various SH3 domains (Figure 3.3.2.1b), my attention was drawn towards adjacent isoleucine ("e" in Figure 3.3.2.1b) and histidine residues that are present in the SH3 RT-loops of Hck and Lyn, but not in other SH3 domains. To test the importance of this part of the RT-loop in differential binding to Nef-PxxP, a reciprocal change involving the three divergent amino acids in this region of Hck and Fyn SH3's was made. This resulted in a mutant Hck-SH3 domain which has a Fyn-like RT-loop (Hck-RTE), and a mutant Fyn-SH3 which has a Hck-like RT-loop (Fyn-IHH) (see Table 3.3.3.1).

The Affinities of Nef and Nef-derived peptide to the SH3 domains of Tyrosine kinases Hck, Fyn, and RT-Loop-targeted mutants

		RT-Loop		Nef protein Kd(μM)	Nef peptide Kd(μM)
Hck-RTE	...YDY	EARTEE	DLS..	>20	N.D.
Hck-wt	...YDY	EAIHHE	DLS..	0.25±0.05	91±5
Fyn-wt	...YDY	EARTED	DLS..	>20	202±7
Fyn-IHH	...YDY	EAIHHD	DLS..	2.0±0.5	N.D.
Fyn-IH	...YDY	EAIHED	DLS..	3.1±0.2	N.D.
Fyn-H	...YDY	EARHED	DLS..	>20	N.D.
Fyn-I	...YDY	EAITED	DLS..	0.38±0.04	91±4
Fyn-A	...YDY	EAATED	DLS..	4.4±0.8	N.D.
Fyn-D	...YDY	EADTED	DLS..	>20	N.D.
Fyn-K	...YDY	EAKTED	DLS..	>20	N.D.

Table 3.3.3.1. Summary of the quantitative Nef/SH3-binding data. Amino acid sequences in the region of the SH3 domain RT-loop that was targeted for a mutagenesis analysis are shown for wild-type (wt) Hck and Fyn and the derived altered SH3 proteins. The introduced amino acid changes are underlined. The equilibrium dissociation constants (K_D) for binding of these SH3 domains to Nef protein, or a 12-mer peptide overlapping the Nef PxxP motif (PVRPQVPLRPMT), measured based on surface plasmon resonance or tryptophan fluorescence, respectively, are shown on their right.

Figure 3.3.3.1 Differential binding of Nef to Hck, Fyn-I and Fyn SH3 domains. Shown are sensorgrams obtained by testing the binding of 8 μM Nef to wild-type Hck (upper line), Fyn-I (middle line) and wild-type Fyn (lower line) SH3 domains. The amounts of Hck, Fyn-I and Fyn SH3 proteins immobilized on the biosensor chips were similar, corresponding to 430, 400 and 490 RU, respectively. All sensorgrams have been corrected for the bulk refractive index effect.


200
150
100
50
0
-50
cRU
Hck
Fyn-l
Fyn
100
200
300
400
Time (s)

The affinity of these mutant SH3 domains in binding to Nef was then compared with wild-type Hck and Fyn SH3 domains by SPR. As expected, in contrast to the strong interaction between Nef and Hck-SH3, binding of Nef to Fyn-SH3 was very weak (Figure 3.3.3.1). At high concentrations of Nef (8 μM), some binding over the background was observed but these sensorgrams were not sufficiently well-defined to determine accurately a K_D value for this interaction (estimated to be weaker than K_D 20 μM) by Scatchard analysis. In contrast, good binding to Nef (K_D 2.0 μM) was observed (Table 3.3.3.1) when the Fyn-SH3 domain with an Hck-like RT-loop (Fyn-IHH) was tested. Conversely, substituting the RT-loop of Hck-SH3 with that of Fyn (Hck-RTE) resulted in loss of quantifiable binding to Nef (K_D >20 μM).

A number of additional mutant Fyn-SH3 domains was then constructed and tested to examine the individual contribution of each of the three Hck-specific amino acids (IHH) to Nef's affinity for Fyn-SH3 (see Table 3.3.3.1). It was found that the affinity of Fyn-IH was similar (K_D 3.1 μM) to that of Fyn-IHH (K_D 2.0 μM), suggesting that the second histidine residue in Fyn-IHH did not contribute to its increased affinity for Nef. Like wild-type Fyn-SH3, binding of Nef to Fyn-H was too weak to be accurately quantitated by the current assay (K_D >20 μM), indicating that the first histidine in the Hck-like RT-loop of Fyn-IHH does not increase binding to Nef when introduced alone into Fyn-SH3. In contrast, Fyn-I bound to Nef even better that Fyn-IHH (or Fyn-IH), showing an affinity for Nef (K_D 380 nM) almost as great as that of the Hck-SH3 (K_D 250 nM). This

suggests that the increased affinity of Fyn-IHH for Nef resulted primarily from replacement of the Fyn arginine residue 96 with an isoleucine (position "c" in Figure 3.3.2.1b). The fact that Fyn-I binds Nef more strongly than Fyn-IH or Fyn-IHH seems counterintuitive, given that both Hck and Lyn have a histidine residue following this isoleucine. The reason for this is unclear in the absence of Hck SH3 structure. One possibility is that in Fyn-IH and Fyn-IHH, the presence of this histidine may interfere with the interaction between the arginine residue of the Nef-PxxP motif and Asp 100 in Fyn.

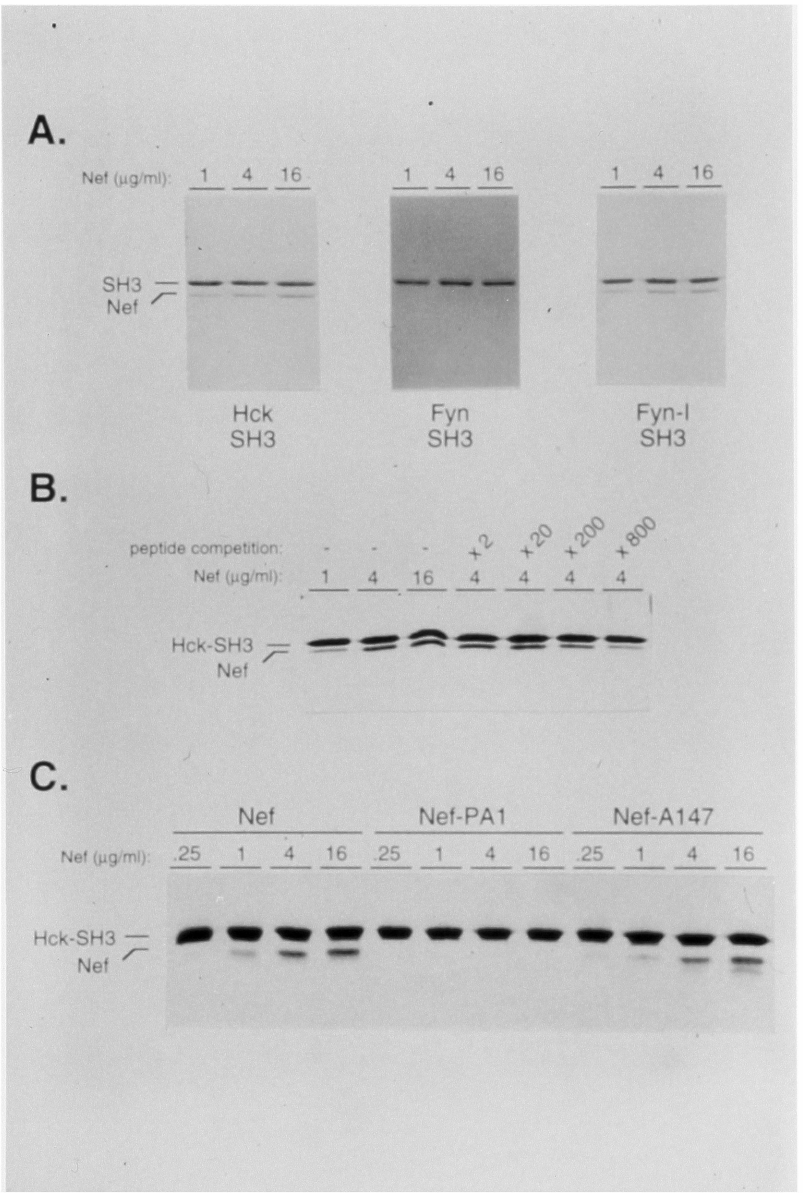
To explore the basis for the increased affinity of the mutant Fyn-SH3 domain (Fyn-I) for Nef, the naturally occurring arginine residue (R96) in this position of Fyn was replaced with other amino acids (Table 3.3.3.1). These substitutions included a basic residue (lysine; Fyn-K), an acidic residue (aspartic acid; Fyn-D), and a non-polar residue with reduced hydrophobic surface (alanine; Fyn-A). None of these Fyn-SH3 mutants bound to Nef as strongly as Fyn-I. However, in contrast to Fyn-K and Fyn-D which did not bind to Nef any better than wild-type Fyn-SH3, Fyn-A showed a relatively high affinity for Nef (K_D 4.4 μM). The fact that lysine and aspartate substitutions at this position lead to equally poor binding indicates that repulsion by the positive charge of the Fyn arginine residue 96 ("e" in Figure 3.3.2.1b) is not the major determinant of the low affinity of Fyn. Rather the interaction involves specific selection of Hck because of the isoleucine sidechain at this position.

3.3.4 Nef PxxP Peptide Alone Has Low Affinity and Limited Specificity in SH3 Binding

The loss of binding to Hck-SH3 by the Nef mutant with a double proline-to-alanine substitution within its tetraproline repeat region indicated that this PxxP motif is essential for the high affinity SH3-binding by Nef. To study the binding properties of an isolated peptide containing the Nef PxxP motif, a 12-mer peptide PVRPQVPLRPMT spanning this region was synthesized, and its interaction with different SH3 domains was studied by tryptophan fluorescence (Knudsen *et al.*, 1995). This method is well-suited for detection of interactions involving small polypeptides, and has the capacity to measure affinities of interactions with K_D values in the high micromolar range.

The affinity of the 12-mer Nef peptide in binding to Hck-SH3 was 91 μM, indicating that this interaction is more than 300-fold weaker than that with full-length Nef protein (Table 3.3.3.1). Furthermore, a K_D value of 202 μM was determined for the interaction between Fyn-SH3 and the Nef peptide, indicating that the differential specificity of Hck-SH3, as compared to Fyn-SH3, in binding to native Nef protein is much less evident when the Nef-PxxP peptide is separated from its natural context.

Figure 3.3.5.1 SH3/Nef co-precipitation assay (a) One μg of the bead-immobilized GST-fusion proteins containing Hck, Fyn, or Fyn-I SH3 domains were incubated with increasing concentrations of Nef as indicated. The amount of Nef protein retained with beads after washes was analyzed by SDS gel electrophoresis. The positions of the GST-SH3 and Nef proteins in the Coomassie blue-stained gels are indicated on the left. **(b)** Beads containing two μg of GST-Hck-SH3 protein were incubated with increasing concentrations of Nef, or with 4 μg/ml Nef in the presence of increasing molar excess of the Nef-PxxP peptide PVRPQVPLRPMT, as indicated, followed by analysis of co-precipitated Nef protein as in (a). **(c)** Beads containing four μg of GST-Hck-SH3 protein were incubated with increasing concentrations of Nef, Nef-PA1, or Nef-A147, followed by analysis of co-precipitated Nef proteins as in (a). The Nef-A147 protein preparation contained traces of residual free GST protein, which binds to the glutathione-beads, and can be seen migrating below the Nef protein in these lanes.


A.
Nef (μg/ml): 1 4 16 1 4 16 1 4 16
SH3
Nef
Hck SH3
Fyn SH3
Fyn-I SH3
B.
peptide competition: - - - x2 x20 x200 x800
Nef (μg/ml): 1 4 16 4 4 4 4
Hck-SH3
Nef
C.
Nef
Nef-PA1
Nef-A147
Nef (μg/ml): .25 1 4 16 .25 1 4 16 .25 1 4 16
Hck-SH3
Nef

3.3.5 Nef-PxxP Motif-mediated SH3-binding Examined by a Co-precipitation Assay

In order to confirm the results on the specificity and affinity of different SH3 domains for Nef and the Nef-PxxP peptide by another technique, I developed a robust and semi-quantitative co-precipitation assay. In this assay GST-SH3 proteins were used to coat glutathione-agarose beads which were then incubated with various concentrations of Nef in the presence or absence of increasing amounts of the competing Nef PxxP peptide PVRPQVPLRPMT. After stringent washes, the amount of Nef protein retained with the SH3-containing beads was analyzed in Coomassie blue-stained SDS-gels. As seen, a dose-dependent increase in co-precipitating Nef protein was observed using beads coated with SH3-domains of Hck or Fyn-I, showing apparent saturation between 4 and 16 µg/ml of Nef (Figure 3.3.5.1, panels Hck and Fyn-I). In contrast, no Nef could be co-precipitated using beads coated with Fyn-SH3 —— even when the highest Nef concentration (16 µg/ml) was tested (Figure 3.3.5.1a, right panel). Also in agreement with the BIAcore experiments (Table 3.3.3.1), some Nef could be co-precipitated with Fyn-A, but not with Fyn-K or Fyn-D (data not shown).

In order to compare the relative affinities of native Nef protein and the Nef-PxxP peptide PVRPQVPLRPMT in binding to the Hck-SH3 domain, increasing concentrations of this peptide were added into the co-precipitation assay to compete with a constant amount of Nef

(4 μg/ml). As seen in Figure 3.3.5.1b, an 800-fold molar excess of the peptide was required to significantly compete with the full-length Nef protein in binding to Hck-SH3, thereby supporting the magnitude of the differences in their measured affinities based on SPR (Nef) or tryptophan fluorescence (Nef-PxxP).

Also in agreement with data obtained by the BIAcore system, the mutant Nef-PA1 protein with a disrupted tetraproline repeat could not be co-precipitated with Hck-SH3 (Figure 3.3.5.1c), again confirming the critical dependence of the Nef/Hck-SH3 interaction on this motif. In contrast, another mutant form of Nef (Nef-A147) that contains a proline-to-alanine substitution in position 147 bound to Hck-SH3 with an apparently similar affinity to the wild-type Nef (Figure 3.3.5.1c) This result indicates that in the context of native soluble Nef protein, the proline residue 147 is not involved in binding to Hck-SH3.

In conclusion, it was found that the full-length Nef protein bound to Hck SH3 with the highest affinity reported for an SH3-mediated interaction (K_d = 250 nM). This is in contrast to a closely-related Fyn SH3 domain whose interaction with Nef was too weak (K_d > 20 μM) to be accurately quantified using the current experimental setup. The distinct specificity lies in a variable loop, the "RT-loop", which is located near the conserved SH3 residues implicated in the binding of PxxP motif. A single amino substitution (Arg to Ile) at residue 96 in the RT loop converted the low affinity binder Fyn into a high affinity one (K_d = 380 nM). Based on

additional mutagenesis studies, it is likely that the selective recognition of Nef by Hck SH3 is determined by hydrophobic interactions involving an isoleucine residue in its RT loop. Since the high affinity and high specificity binding is only observed for the full-length Nef protein (but not for the PxxP peptide), I proposed that the interaction between Nef and Hck SH3 domain is likely to involve a region of Nef other than the PxxP motif. To test this hypothesis, I have determined the structure of Nef in complex with Fyn-I SH3 domain. The structure described below is in excellent agreement with the biochemistry data.

3.2.6 Summary

The differential binding of Hck and Fyn to HIV-1 Nef has been examined to elucidate the structural basis of SH3 binding affinity and specificity. Full-length Nef bound to Hck-SH3 with the highest affinity reported for a SH3-mediated interaction (K_D 250 nM). In contrast to Hck, affinity of the highly homologous Fyn-SH3 to Nef was too weak (K_D >20 μM) to be accurately determined. It was shown that this distinct specificity lies in a variable loop, "the RT-loop", positioned close to conserved SH3 residues implicated in the binding of proline-rich (PxxP) motifs. A mutant Fyn-SH3 with a single amino acid substitution (R96I) in its RT-loop had an affinity (K_D 380 nM) for Nef comparable to that of Hck-SH3. Based on additional mutagenesis, we propose that the selective recognition of Nef by Hck-SH3 is determined by hydrophobic interactions involving an isoleucine residue in its RT-loop. Although Nef contains a PxxP motif which is necessary for the interaction with Hck-SH3, high affinity binding was only observed for the intact Nef protein. The binding of a peptide containing the Nef PxxP motif showed >300-fold weaker affinity to Hck-SH3 than full-length Nef.

3.4 Crystal Structure of HIV-1 Nef-FYN(I) SH3 complex

The biochemical analysis of the Nef-SH3 interaction indicated that the tertiary structure of Nef is critical for its high affinity and high specificity in binding to the Hck SH3 domain. In order to understand the molecular basis of the Nef-SH3 recognition, efforts have been made to obtain crystals suitable for structural analysis. The full-length Nef have been proven to be refractory to crystallization. Therefore, I applied deletion mutagenesis to define the structured region of the Nef protein (for more details, see the Methods section 2.4.1).

Data quality crystals were obtained for the core domain of Nef (Nef_{core}; residues 54-205) in complex with a mutant Fyn SH3 domain (Fyn-I, which binds tightly to Nef). The structure was determined by multiwavelength anomalous diffraction (MAD) using a lead derivative and was refined at 2.5 Å resolution (see Method section). The asymmetric unit contains two Nef_{core}-SH3 complexes. The two molecules of Nef in the asymmetric unit are very similar (rms deviation in Cα position of 0.7Å), as are the two SH3 domains (rmsd in Cα of 0.6 Å). There is, however, a small but significant difference in the relative orientation of the SH3 domain with respect to Nef in the two complexes (see section 3.4.4) . This difference in the orientation appears to be due to differences in crystal packing interactions. As a result, the SH3 domain from one complex (with more crystal contacts) has a lower average temperature factor (~30 $Å^2$) than the other (53 $Å^2$), while the average temperature factors for

both Nef_{core} are about the same (31 and 28 $Å^2$). Therefore, the complex that has lower average temperature factor is used for most of the analyses described below, while the differences in the structures of the two complexes is discussed separately (section 3.4.4). Two regions composed of a significant portion of the Nef_{core} are disordered: the N-terminal 16 residues and 29 residues in a large internal loop (residue 149-177). These regions of Nef, however, are distant from the SH3 domain binding surface. Therefore, their omission from the model may bear little relevance to our discussion below.

3.4.1 Overall Structure of Nef

The architecture of Nef_{core} consists of three layers. The N-terminal region forms an outer layer that consists of a polyproline type II (PP-II) helix (Arg 71 to Arg 77), which contains the PxxP motif, followed by two anti-parallel α-helices (αA and αB), that pack against a middle layer of 4 anti-parallel β strands (Figures 3.4.1.1a & b). The C-terminal region consists of 2 short α-helices, and packs on the other side of the β strands. The first helical layer forms a contiguous region on the surface of Nef_{core} that is responsible for the entire interaction with SH3 (Figure 3.4.1.2). The interaction surface is formed by the closed arrangement of the helices. The PP-II helix leads into helix αA, and also packs against the C-terminal edge of helix αB. The two α-helices are connected by a relatively long linker (10 residues), and although they are in an anti-parallel arrangement, they do not pack closely against each other. Instead, they are

separated by approximately 10.5 Å (distance of closest approach of Cα atoms), with their axes inclined by ~70°, so as to create a hydrophobic and solvent accessible crevice between them (Figures 3.4.1.2 & 3.4.1.3a). The major portion of this crevice between the helices is unoccupied by the SH3 domain, and is separated from the SH3 binding region by two aromatic residues (Trp 113 and Phe 90, see Figure 3.4.1.3a). Conserved sidechains that line this crevice are available for potential interactions with other molecules (Figure 3.4.1.3a). The general features of the crystal structure are consistent with the demonstration by NMR of the presence of two anti-parallel α-helices corresponding to αA and αB and the anti-parallel nature of the strand network (Grzesiek *et al.*, 1995).

The four anti-parallel strands that make up the second layer of Nef_{core} are irregular in their architecture. They do not form a contiguous 4-stranded β sheet, but are instead separated into two distinct anti-parallel pairs of strands (Figure 3.4.1.3b). Strand βA is part of an extended chain that runs from the C-terminus of helix αB, at Gly 119, to a turn at Val 133. The chain is kinked at a Pro-Asp-Trp sequence (residues 122 to 123), with the tryptophan sidechain being buried in the hydrophobic core of the molecule. The short βA strand consists of residues 126 to 128 that each form at least one anti-parallel β-sheet hydrogen bonding interaction with residues in βB, and is terminated by a Pro-Gly-Pro-Gly-Val-Arg-Tyr-Pro sequence that induces a wide bend and irregular conformation in the chain due to the presence of the three prolines. Only two hydrogen

Figure 3.4.1.1. Structure of Nef:SH3 complex. (a,b). Stereo-diagrams of the polypeptide backbones of Nef_{core} and Fyn(R96I) SH3. The N-terminal helical layer of Nef_{core} (residues 71-120), which forms the SH3-interaction surface, is colored yellow. The rest of Nef_{core} (residues 121-203) is colored green. The disordered loop (residues 149-178) between βC and βD is indicated as a dotted line. The Fyn(R96I) SH3 domain is in blue. Also shown are the sidechains of the conserved tryptophan of SH3 (residue 119, in red), the specificity conferring isoleucine of SH3 (residue 96, in red) and the two prolines that define the PxxP motif of Nef (residues 72 and 75, in yellow). The views in (a) and (b) are approximately orthogonal. The figures are prepared using MOLSCRIPT (Kraulis, 1991) and Raster3D (Bacon and Anderson, 1988).

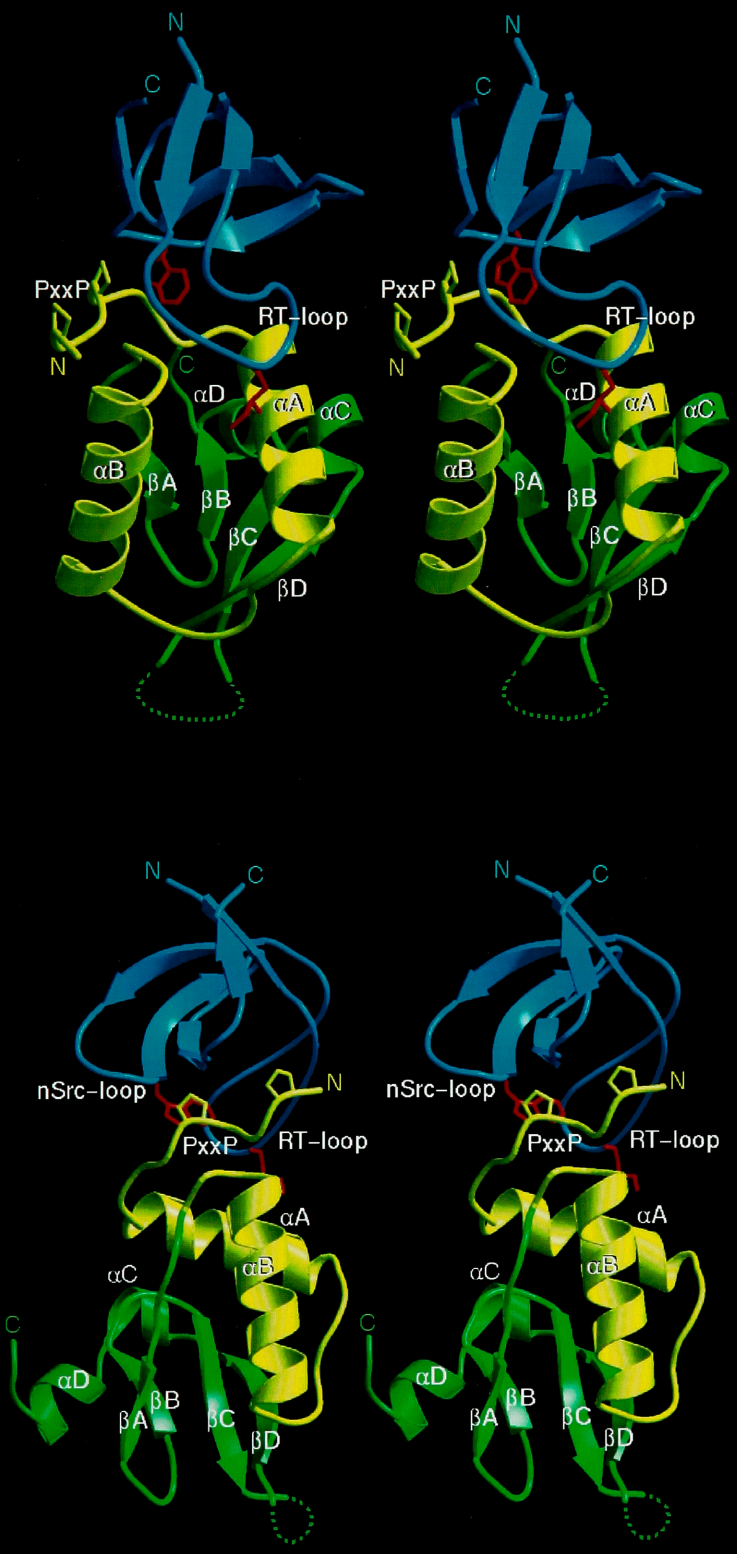

N
C
PxxP
RT-loop
N
C
αD
αA
αC
αB
βA
βB
βC
βD
N
C
PxxP
RT-loop
N
C
αD
αA
αC
αB
βA
βB
βC
βD
N
C
nSrc-loop
N
PxxP
RT-loop
αA
αC
αB
C
αD
βB
βA
βC
βD
N
C
nSrc-loop
N
PxxP
RT-loop
αA
αC
αB
C
αD
βB
βA
βC
βD

Figure 3.4.1.2. Molecular Surface of Nef. The molecular surface of Nef_{core}, with Fyn(R96I) SH3. The local electrostatic potential of Nef_{core} was calculated in the absence of the SH3 domain, using GRASP (Nicholls *et al.*, 1991). The molecular surface is colored according to the local electrostatic potential, with colors ranging from dark blue (most positive region) to deep red (most negative), through white (neutral). The SH3 domain is shown as a blue tube. The sidechains of Trp 119 and Ile 96 of SH3 are shown in yellow. Trp 113 and Phe 90 of Nef separate the binding pocket for Ile 96 of SH3 from the hydrophobic crevice that is available for potential interaction with other molecules. Arg 106 of Nef, located at the lower left lower edge of the crevice is implicated in the association of Nef with a Serine kinase activity (Sawai *et al.*, 1995).

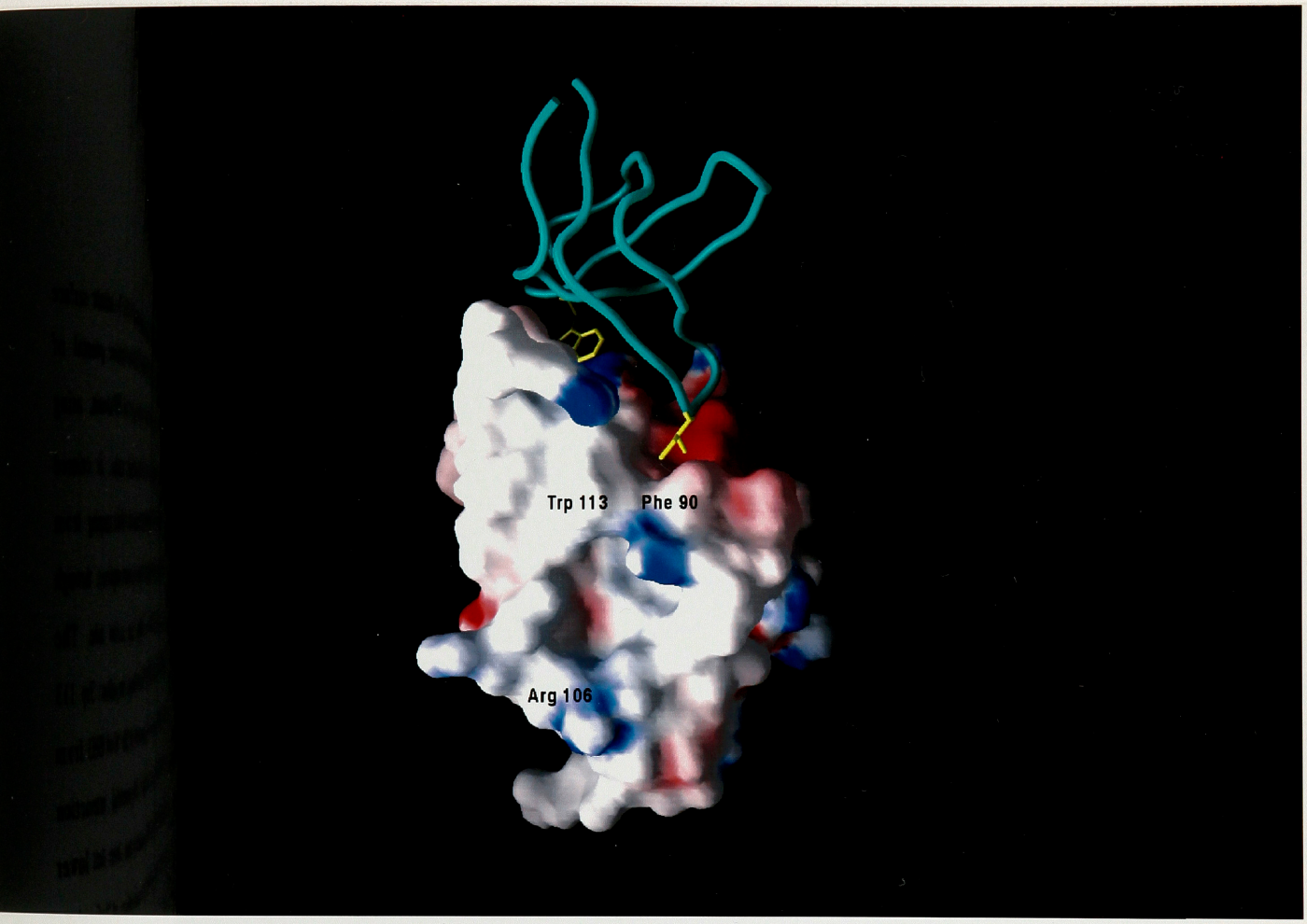
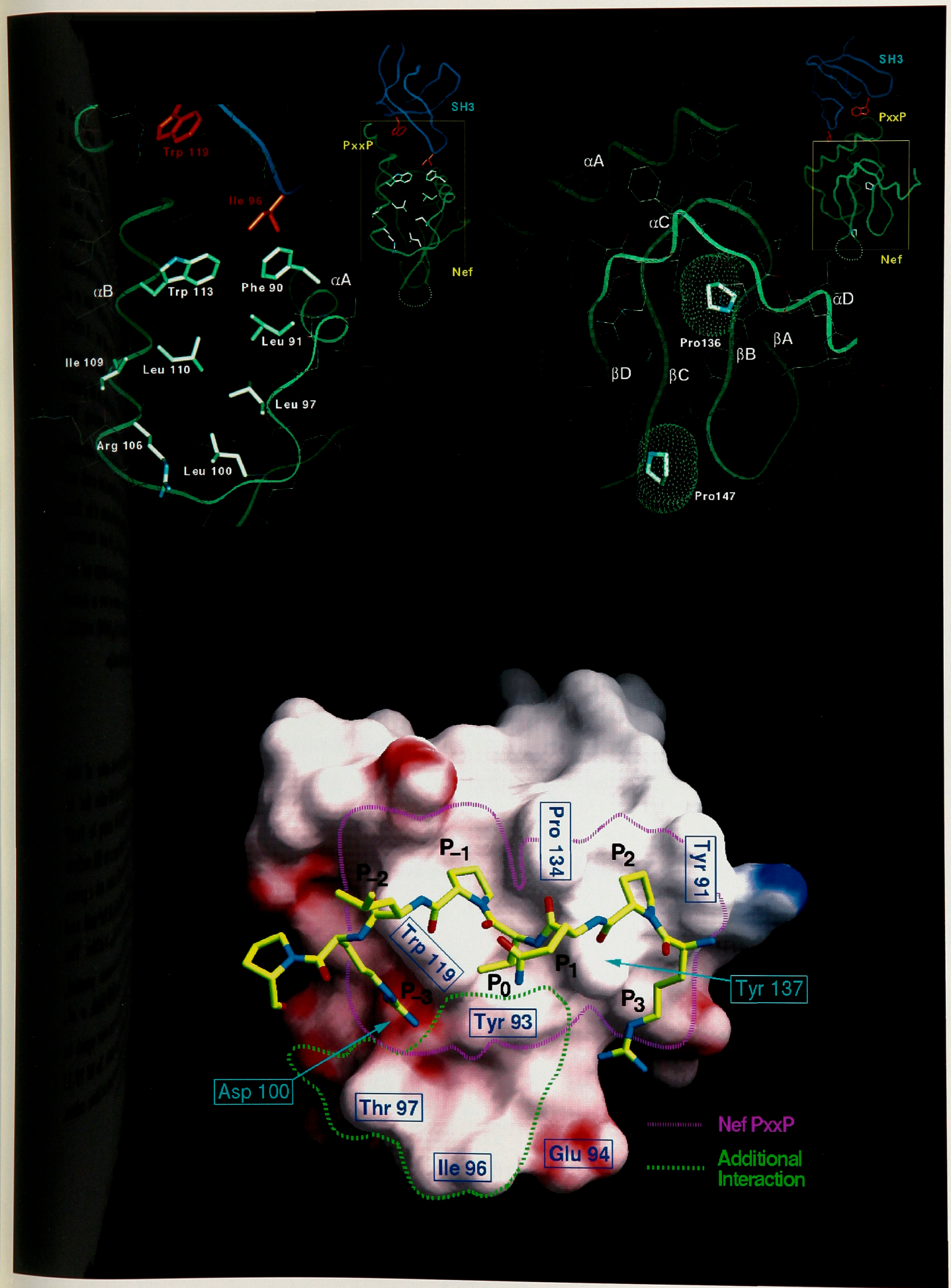

Trp 113
Phe 90
Arg 106

Figure 3.4.1.3 Architecture of Nef. **(a)** The spacing between the two helices, αA and αB, and the loop connecting them is illustrated. A hydrophobic crevice is formed between the two helices. The crevice is bordered by eight conserved residues and is separated from the SH3 interaction surface by the sidechains of two aromatic residues (Trp 113 and Phe 90). Arg 106, which has been found to be essential for the association of Nef with a Serine kinase activity (Sawai *et al.*, 1995), is located at one edge of this crevice. Insert : A ribbon diagram of the Nef:SH3 complex in the same orientation. The region that is boxed is shown in detail in the main figure. The SH3 domain and Nef_{core} are shown in dark blue and in green, respectively. The conserved tryptophan (residue 119) and the specificity conferring isoleucine (residue 96) of Fyn (R96I) SH3 domain (shown in red) are provided as landmarks. **(b)** The irregular architecture of the four anti-parallel strands (βA-βD) of Nef_{core}. Two strictly conserved prolines (residues 136 and 147) hold apart the two hair-pins and the van der Wall surfaces of these two proline are shown as dotted surfaces. Insert: ribbon drawing of the Nef_{core}:Fyn(R96I) complex. The boxed region is shown in detail in the main figure.

Figure 3.4.2.1 The footprint of Nef on the surface of the Fyn(R96I) SH3 domain. The region on the SH3 surface that is occluded by the Nef polyproline type-II helix (including the PxxP motif) is outlined in pink. An additional region on the SH3 surface that is occluded by other elements of Nef is marked with a green dotted line. Tyr 93 of SH3, is partially covered by Nef PxxP motif and is completely buried when the rest of the Nef_{core} is included. The surface is colored according to the local electrostatic potential, calculated in the absence of Nef_{core} using GRASP (Nicholls, *et al.*, 1991). Surface colors range from dark blue (most positive region) to deep red (most negative), through white (neutral). The Nef PP-II helix is shown as a stick figure, with atoms colored as follows: carbon, yellow; nitrogen, blue; oxygen, red.


SH3
PxxP
Nef
Trp 119
Ile 96
αB
Trp 113
Phe 90
αA
Leu 91
Ile 109
Leu 110
Leu 97
Arg 106
Leu 100
αA
αC
αD
βA
βB
βC
βD
Pro136
Pro147
P–1
P–2
Pro 134
P2
Tyr 91
Trp 119
P1
P0
P–3
P3
Tyr 137
Tyr 93
Asp 100
Thr 97
Glu 94
Ile 96
Nef PxxP
Additional Interaction

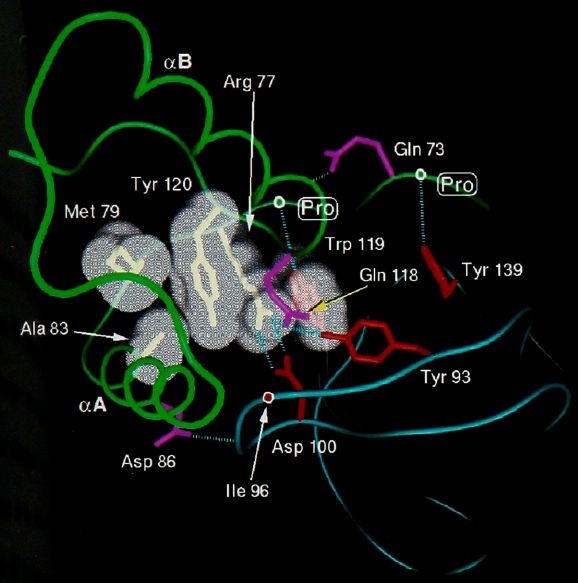
bonds are formed between strands βB and βC. The presence of two strictly conserved prolines, one at position 136 (in βB) and the other at 147 (immediately following βC) appears to hold apart the two strands, preventing the formation of a contiguous sheet structure across strands A, B, C and D (Figure 3.4.1.3b). The paucity of regular hydrogen bonding interactions between the strands, although unusual, is not without precedent. For example, the structure of the C-type lectin domain consists of a number of irregular strands with limited inter-strand hydrogen bonding (Weis *et al.*, 1991).

The C-terminal region of Nef_{core} (residues 187 to 203) only partially covers the distal surface of the β strands, and thus the bulk of the hydrophobic core of the domain is formed between the strands and the N-terminal helices. The irregular architecture of the strands coupled with the spatially separated orientation of αA and αB results in a relatively loose and open appearance to the structure, and suggests that the fold of Nef_{core} might not be particularly stable.

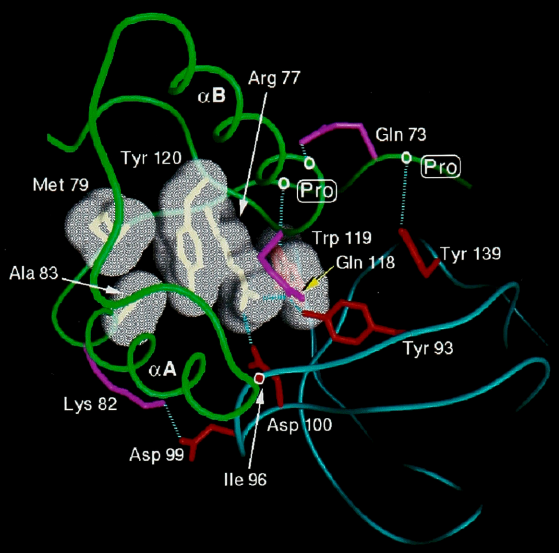
3.4.2 The Interaction between Nef PxxP Motif and the SH3 Domain

The structure of the Fyn SH3 domain has been described previously (Noble *et al.*, 1993) and the structure of the Fyn (R96I) SH3 domain seen here is essentially unchanged. The β barrel structure of the SH3 domain presents an array of conserved hydrophobic sidechains that are spaced appropriately for interaction

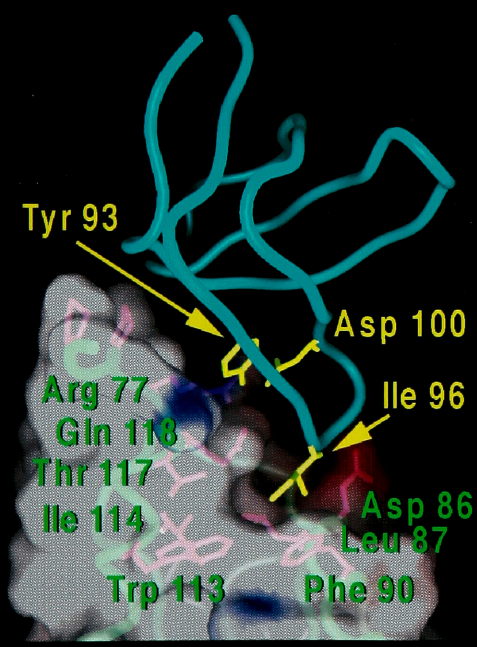
Figure 3.4.2.2 Comparison of the Nef:SH3 and peptide:Crk-SH3 interfaces. (a,b) Comparison of Nef PP-II helix interacting with Fyn(R96I) SH3 (left panel) with the Sos peptide:Crk SH3 interaction (right panel) (Wu *et al.*, 1995). In (a), the view is approximately perpendicular to the SH3 binding surface. The polypeptide backbones of the SH3 domains are shown as blue ribbons and the conserved hydrophobic residues on the interaction surfaces of the SH3 domains are colored white. The PP-II helices are shown in green, with the polypeptide backbone of the helices indicated as red ribbons. In the left panel, the rest of Nef_{core} is shown as a red Cα-backbone trace. Ile 96 of the Fyn(R96I) SH3 domain is shown in red, as is Asp 100. The view in (b) is approximately orthogonal to that in (a), and emphasizes the three-fold pseudo symmetry of the PPII helices formed by the PxxP motifs. Two of the three edges of the PP-II helices pack against conserved hydrophobic residues (in white) on the surface of the SH3 domains. In the Nef:SH3 complex (left panel), Asn 136 of Fyn(R96I) SH3 forms a hydrogen bond with the backbone of the PP-II helix of Nef. Although Crk has a proline in the corresponding position (right panel), other SH3 domains have glutamine (Lim *et al.*, 1994). The view shown is from the C-terminus of the PP-II helix in both cases.

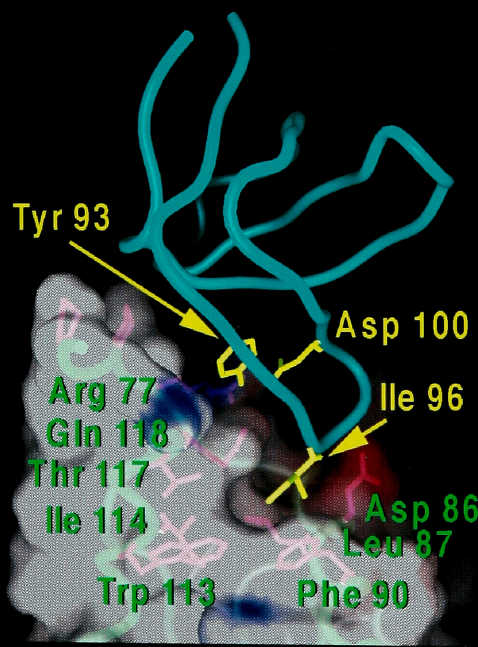


Complex-1



Complex-2





with polyproline helices (Figure 3.4.2.1). Two loops that abut this interaction surface are of particular interest in terms of SH3 specificity: the so-called "RT-Src" or "RT" loop (between the first and second strands of the SH3 domain) and the "n-Src" loop (between the second and third strands). The historical nomenclature for these SH3 loops refers to critical arginine and threonine residues in the RT loop in Src, and to insertions in the n-Src loop in neuronal forms of Src. The RT loop plays a key role in the interaction with Nef, while the n-Src loop does not.

Residues 71 to 77 of Nef form a left-handed polyproline type II helix that spans the strictly conserved PxxP motif. The PPII helix has sidechains emanating in three directions, two of which are utilized for interactions with the SH3 domain (Figures 3.4.21 & 3.4.2.2). PP-II helices can bind to SH3 domains in two orientations ("plus" and "minus"), with the particular choice of direction being dictated by electrostatic complementarity and by steric restrictions that are set by the presence of non-proline residues on one or the other interacting edge of the PP-II helix (Feng *et al.*, 1994; Lim *et al.*, 1994). The Nef PPII helix interacts with the SH3 domain in the "minus" orientation and the interface with the SH3 domain is strikingly similar to that seen in the structures of similarly oriented peptides bound to the SH3 domains of Sem5, Grb2, Src and Crk (Lim *et al.*, 1994; Feng *et al.*, 1994; Terasawa *et al.*, 1994; Wu *et al.*, 1995; Goudreau *et al.*, 1994). In the following discussion I compare the Nef PP-II:SH3 interaction with the interaction between the N-terminal SH3 domain of c-Crk and a peptide corresponding to a sequence in

the guanine-nucleotide exchange factor Sos (Wu *et al.*, 1995). Similar features are seen in structures of Sem5, Grb2 and Src SH3 domains complexed with peptides. For consistency, I adopt the notation used previously, and refer to binding sites for the PPII helix as P_{-1}, P_0, P_{+1}, etc. (Lim et al., 1994), where Val 74 of Nef occupies the P_0 position. Note that in the "minus" orientation seen here the numbering of these sites runs counter to the direction of the polypeptide chain of the PP-II helix of Nef.

One edge of the PP-II helix of Nef contains the two residues that define the PxxP motif, Pro 72 and Pro 75. These occupy the P_{-1} and P_{+2} sites on the SH3 domain and pack against conserved hydrophobic sidechains (Tyr 91, Trp 119, Pro 134 and Tyr 137 of SH3; Figures 3.4.2.2). The other edge of the PP-II helix that interacts with the SH3 includes three non-proline residues (Arg 71, Val 74 and Arg 77). The arginine residues provide hydrophobic, hydrogen bonding and electrostatic interactions with the SH3 domain. Arg 77 is strictly conserved in Nef sequences (Figure 3.4.4.1) and is critical to the integrity of the interface, since it is involved in the formation of an extensive network of interactions with other components of the Nef structure and with the SH3 domain (Figure 3.4.3.1). This sidechain stacks closely against a face of Trp 119 sidechain of SH3, and forms a salt bridge with a conserved acidic residue (Asp 100) in the RT loop of SH3. Similar ion-pairing interactions have been observed in many SH3:peptide complexes. The N-terminal Arg 71 of Nef is relatively poorly ordered, but forms a hydrogen bond via its $N\zeta$ atom with the hydroxyl group of Tyr 137 and may also contribute

to electrostatic complementarity since it is in the vicinity of two acidic residues (Asp 92 and Glu 94) of the SH3 domain. Val 74 of Nef occupies the central P_0 binding site, and packs against Tyr 93, Trp 119, and Tyr 137 sidechains of SH3. Val 74 is highly conserved in Nef sequences, and is completely buried at the interface due to tertiary interactions in the Nef protein.

Pro 72 and Pro 75 are strictly conserved in all Nef sequences, and are essential for the enhancement by Nef of viral replication in vitro (Saksela *et al.*, 1995). The critical role for these two residues in Nef function correlates with their central position at the Nef:SH3 interface. In contrast, prolines at positions 69 and 78 are not as highly conserved among primate immunodeficiency viruses (Figure 3.4.4.1) and are dispensable for the binding to SH3 domains (Lee *et al.*, 1995). Pro 69 is part of the N-terminal disordered region of Nef_{core} and is not modeled in the crystal structure. Pro 78 does not interact directly with the SH3 domain but instead plays a structural role in the recognition by causing a kink in the loop connecting the PP-II helix to helix αB. The sidechain of Tyr 120, which is part of the network of tertiary interactions at the interface (Figure 3.4.3.1a), packs against the backbone of Pro 78.

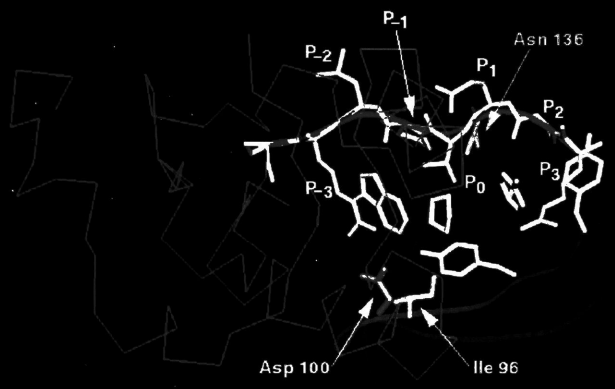
There is remarkably close correspondence between the way in which the PP-II helix of Nef docks on the SH3 domain, and the mode of interaction of peptide ligands with SH3 domains (Figure 3.4.2.2a,b). If the SH3 domains of the Nef:SH3 complex and the Crk:peptide complex are superimposed, the average deviation in the

positions of the Cα atoms of the two PP-II helices is 1.1 Å. The similarity in the general disposition of the Nef PP-II helix on the SH3 surface with that of peptide ligands is also emphasized by the fact that very similar hydrogen bonding interactions between the peptide and the SH3 domain are seen in the two cases. Hydrogen bonds between the backbone of the PP-II helix of Nef and the sidechains of Trp 119 and Tyr 137 of the Fyn (R96I) SH3 domain are similar to those seen in peptide complexes of Crk, Grb2 and Sem5. Asn 136 of Fyn SH3 forms a hydrogen bond to the backbone of the PP-II helix (Figures 3.4.2.2a,b). Many SH3 domains, including Grb2, have a glutamine at this position, and a similar hydrogen-bonding interaction is found in the Grb2-peptide complex (Lim *et al.*, 1994).

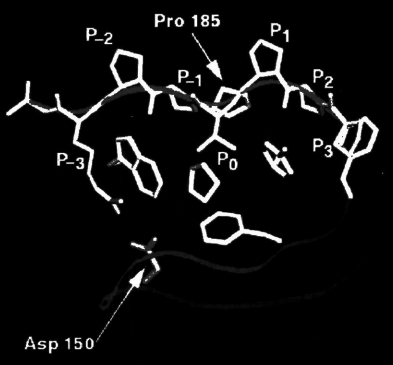
3.4.3 Tertiary Interactions Between Nef_{core} and SH3

Unlike in SH3:peptide interactions, the Nef PxxP motif is presented in the context of a folded protein that contributes additional elements to the binding interface. Consequently, the interaction surface of Nef with the SH3 domain is larger than for an isolated peptide. The total accessible surface area on Nef and on the SH3 domain that is buried upon formation of the complex is ~1,200 $Å^2$ (calculated using a probe of radius 1.4 Å), with about 600 $Å^2$ buried on each partner. In comparison, if the PP-II helix alone of Nef is considered, the total buried surface area is ~780 $Å^2$, with ~390 $Å^2$ buried on the SH3 domain. The increase in interaction area between the SH3 domain and Nef relative to the core interaction of the PP-II helix is due to contributions from the two helices αA and αB and

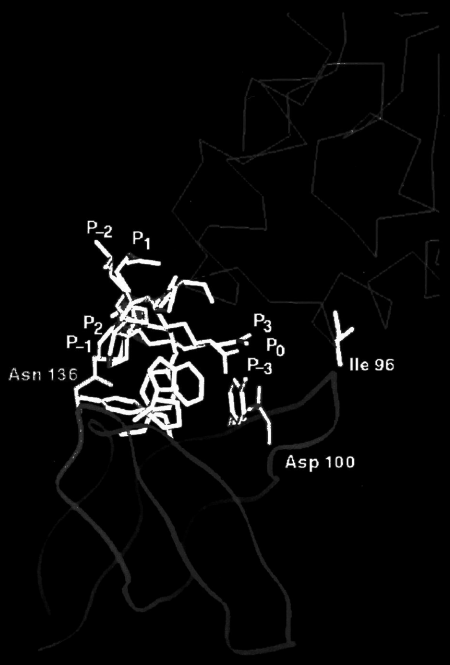
Figure 3.4.3.1. Tertiary interactions between Nef_{core} and Fyn(R96I) SH3 domain. (a) Comparison of the interactions in the two complexes in the crystal. (b) Molecular surface of Nef, showing the binding site for the isoleucine sidechain of the SH3 domain. The polypeptide backbones of Nef and the SH3 domain are shown as green and blue tubes, respectively. Sidechains of Nef are shown in pink and in yellow (displayed under their respective molecular surfaces). SH3 sidechains are shown in red. Hydrogen bonding interactions are shown as dashed lines. Hydrogen bonds to backbone positions are indicated by the placement of white circles along the backbone ribbon. For clarity, the sidechain of Ile 96 is not shown, and instead the Cα position of this residue is indicated with a red circle. The structure on the left is the complex that is focus of the major part of the discussion in the text. The structure on the right is that of the second independent complex in the crystal. Note the slight change in the relative orientation of the Nef and SH3 components of the complex. The sidechain of Asp 86 forms a hydrogen bond with Thr 97 in the RT loop of the second complex. This interaction is not shown for clarity.



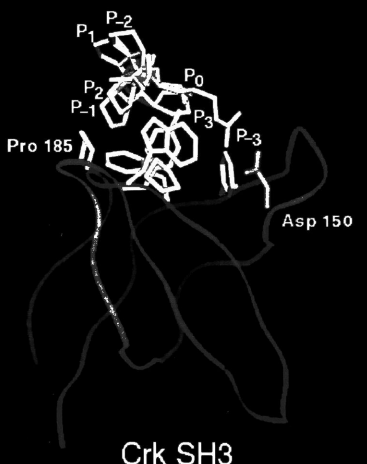
Fyn(R96I) SH3



Crk SH3



Fyn(R96I) SH3



Crk SH3

adjacent residues in Nef, and from the RT loop in the SH3 domain (Figures 3.4.2.1 & 3.4.2.2).

The PP-II helix of Nef interacts with the rest of Nef as well as with the SH3 domain. The total accessible surface area of the Nef PP-II helix, considered separately from the rest of the protein, is ~1,200 $Å^2$. In the Nef:SH3 complex approximately a third of the surface area of the PP-II helix is buried by the SH3 domain and a third by the rest of the Nef protein. The main interaction of the PP-II helix with other elements of Nef involves the C-terminal region of helix αB and the loop immediately following (Figure 3.4.3.1). The sidechain of Gln 73 in the PxxP motif is hydrogen bonded to the backbone of αB. Three other hydrogen bonding interactions between the PxxP motif and the C-terminal region of αB are likely to be important in stabilizing the position of Arg 77, a critical component of the Nef:SH3 interface. The backbone carbonyl groups of residues 118 and 119 are hydrogen bonded to the N^{ζ} atom of Arg 77 and the backbone amide of Leu 76, respectively. The conformation adopted by these two consecutive carbonyl groups is made possible by the presence of a strictly conserved glycine at position 119 of Nef (Figure 3.4.41). Finally, the amide nitrogen at position 121 of Nef hydrogen bonds to the carbonyl group of Arg 77.

In the Nef:SH3 complex the RT loop of the SH3 domain extends over the surface of Nef, such that the sidechain of Ile 96 of SH3 is inserted into a pocket formed between helices αA and αB of Nef (Figure 3.4.3.1). The interaction of the RT-loop with Nef thus involves

a hydrophobic anchor point, the isoleucine sidechain at position 96, which packs against the conserved sidechains of Leu 87, Phe 90, Trp 113 and Ile 114 of Nef. The rest of the Nef:SH3 interface is remarkably polar, with the exception of the two proline and the valine residues of the PP-II helix.

The polar nature of the central region of the interface results from the formation of a network of interconnecting hydrogen bonds between Nef and the SH3 domain. A notable feature of this network is the role played by Arg 77 of the PP-II helix of Nef. In the Nef:SH3 complex, the interaction of Arg 77 with Asp 100 is likely to be strengthened by the formation of stacking interactions between Arg 77 with a conserved tryptophan sidechain of the SH3 domain (Trp 119) on one side, and with the sidechain of Tyr 120 of Nef on the other. Tyr 120 is positioned by packing against Met 79 and Ala 83 of Nef. Asp 100 of SH3 is also hydrogen bonded to Gln 118 of Nef, which is at the C-terminus of helix αB. Gln 118 is hydrogen bonded to the hydroxyl group of Tyr 93 of the SH3 domain, which interacts with the valine sidechain at position P_0 of the PP-II helix of Nef. The buttressing of the interaction of Arg 77 with Asp 100, in the RT loop, has an additional significance in the Nef:SH3 complex, since it serves to position the Ile 96 sidechain that is also in the RT loop and is involved in the hydrophobic interaction with the crevice on the surface of Nef.

Most of the SH3 residues involved in the hydrogen bonding network at the Nef:SH3 interface are highly conserved in Src-family

SH3 domains, are therefore unlikely to play a role in determining the specificity of the interaction. However, they function to place the RT loop in a defined orientation on the Nef binding surface. In addition to being highly variable in its sequence, the RT loop of SH3 domains is a highly flexible element (see, for example, (Gosser *et al.*, 1995)) and the hydrogen bonding network at the Nef:SH3 interface is likely to localize the RT loop, allowing "readout" by the Nef surface of the specificity-determining residues of the SH3. These interactions clearly lead to quite high specificity, with a greater than 100-fold selectivity over closely related SH3 domains (Lee *et al.*, 1995). The low affinity of wild-type Fyn SH3, with an arginine at position 96, is readily understood, given the hydrophobic nature of the binding site for Ile 96 in the mutant.

Despite the similarity between the SH3 interactions of the PP-II helix of Nef and peptide ligands of SH3 domains, isolated peptides corresponding to this region of Nef do not bind with high affinity to SH3 domains. For example, the 12-mer peptide PVRPQVPLRPMT, which encompasses the PxxP motif of Nef, binds to the Hck SH3 domain with a K_d of 91 μM (Lee *et al.*, 1995). In contrast, the peptide PPPVPPRRRR (derived from Sos), used in the comparative illustrations, binds to the Crk SH3 domain with a K_d of 5.2 μM (Knudsen *et al.*, 1995). Part of the reduction of binding affinity of the Nef PxxP peptide is likely to be due to the presence of non-proline residues at the non-binding edge of the PP-II helix (Gln at P_{-1} and Leu at position P_{-2}). In the case of the isolated peptide binding to the SH3 domain, these residues do not contribute to the

binding interface and are likely to increase the entropic cost of forming the PP-II helix. The intact Nef protein binds more than 300 times more tightly to the SH3 domain than the isolated peptide, indicating that the additional interactions described here compensate for these destabilizing factors.

3.4.4 Flexibility of the Nef:SH3 Interface

There are two Nef:SH3 complexes in the asymmetric unit of the crystal, and these have different crystal packing interactions with the lattice. The structures of the SH3 and Nef domains are similar in the two complexes, but differ in the relative orientation of Nef and SH3 (Figure 3.4.3.1a). The PP-II helix of one of the two Nef molecules is rotated by about 15° with respect to the orientation of the other. However, the position of the PP-II helix with respect to the SH3 domain is essentially unchanged (the SH3 domain and the PP-II helix move as one unit), which results in a 15° rotation of Nef (excluding the PP-II helix) with respect to the SH3 domain, pivoted about a point close to the C-terminal end of the PP-II helix.

The net effect of the conformational change is that the SH3 domain and Nef are tipped such that the isoleucine sidechain of the SH3 domain in the second complex is more buried in the binding pocket on Nef. All of the interactions described for the first domain are generally preserved in the second complex, although some of the details differ because sidechain positions have changed by 1 to 2 Å due to the inter-molecular rotation (Figure 3.4.3.1a). The most

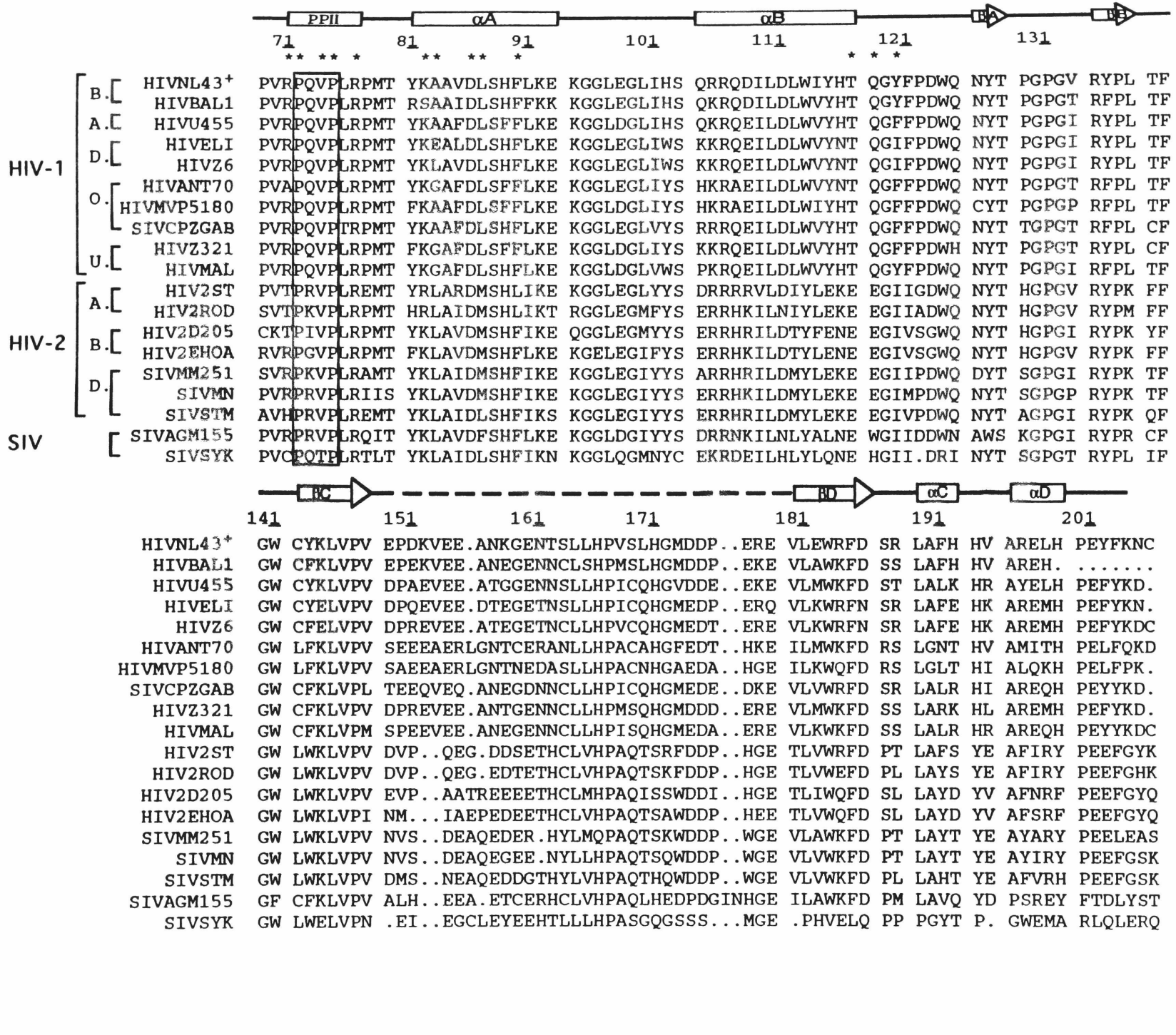
extreme difference involves the interaction of helix αA with the RT loop of the SH3 domain. In the first complex, the sidechain of Asp 86 (strictly conserved in the Nef sequences shown in Figure 3.4.41) hydrogen bonds with the backbone of the RT loop. This interaction is replaced in the second complex by a hydrogen between Asp 86 and the sidechain of Thr 97 of the SH3 and an additional ion-pairing interaction between the well-ordered sidechains of Lys 82 of Nef and Asp 99 in the RT loop of the SH3 domain (Figure 3.4.3.1a).

This conformational difference emphasizes an essential feature of the Nef:SH3 interface. Rather than a rigid inter-digitation of hydrophobic sidechains, the polar nature of the Nef:SH3 interaction results in some degree of flexibility at the interface. In particular, adjustments in the positions of hydrogen bonding groups at the interface may be assisted by the participation of water molecules in the network of interactions. Although water molecules are observed at certain positions at each of the interfaces, a detailed analysis of the role of these waters awaits the measurement of more accurate X-ray data at higher resolution.

3.4.5 Sequence Variation in Nef_{core} and its Implication in the Nef-Host Proteins Interactions

The Nef sequences shown in Figure 3.4.4.1 are representative of the range of sequence diversity observed in HIV-1, HIV-2, and SIV. For the region corresponding to the ordered structure in the crystal (residues 71 to 206, excluding the disordered loop between

Figure 3.4.4.1 Sequence alignment of Nef_{core}. The amino-acid sequences of the conserved core region of Nef protein from different strains of HIV-I, HIV-2 and SIV are shown. Residue numbers above the sequences correspond to HIV-1 NL4-3. The secondary structural elements of the Nef_{core} crystal structure are indicated by arrows for strands (labeled βA - βD) and boxes for α-helices (labeled αA - αD) and poly-proline type II helix (labeled PPII). A loop connecting βC and βD is disordered in the structure and is indicated as a broken line. Residues involved in the interaction with SH3 are labeled with an asterisk above the sequence. +The NL4-3 strain used in this paper contains a mutation that replaces Thr 71 with Arg, corresponding to most patient-derived sequences. The sequences shown were obtained from a database at Los Alamos National Laboratory, via internet (http://hiv-web.lanl.gov/) and were selected in an attempt to illustrate the diversity of Nef sequences. The particular HIV-1, HIV-2 and SIV sub-groupings shown are as given in the Los Alamos database. The groupings are based on sequence similarity, and not on the origin of the virus.


PPII
αA
αB
βA
β3
71
81
91
101
111
121
131
HIV-1
HIV-2
SIV
B.
A.
D.
O.
U.
HIVNL43+ PVRPQVPLRPMT YKAAVDLSHFLKE KGGLEGLIHS QRRQDILDLWIYHT QGYFPDWQ NYT PGPGV RYPL TF
HIVBAL1 PVRPQVPLRPMT RSAAIDLSHFFKK KGGLEGLIHS QKRQDILDLWVYHT QGYFPDWQ NYT PGPGT RFPL TF
HIVU455 PVRPQVPLRPMT YKAAFDLSFFLKE KGGLDGLIHS QKRQEILDLWVYHT QGFFPDWQ NYT PGPGI RYPL TF
HIVELI PVRPQVPLRPMT YKEALDLSHFLKE KGGLEGLIWS KKRQEILDLWVYNT QGIFPDWQ NYT PGPGI RYPL TF
HIVZ6 PVRPQVPLRPMT YKLAVDLSHFLKE KGGLEGLIWS KKRQEILDLWVYNT QGIFPDWQ NYT PGPGI RYPL TF
HIVANT70 PVAPQVPLRPMT YKGAFDLSFFLKE KGGLEGLIYS HKRAEILDLWVYNT QGFFPDWQ NYT PGPGT RFPL TF
HIVMVP5180 PVRPQVPLRPMT FKAAFDLSFFLKE KGGLDGLIYS HKRAEILDLWIYHT QGFFPDWQ CYT PGPGP RFPL TF
SIVCPZGAB PVRPQVPTRPMT YKAAFDLSHFLKE KGGLEGLVYS RRRQEILDLWVYHT QGFFPDWQ NYT TGPGT RFPL CF
HIVZ321 PVRPQVPLRPMT FKGAFDLSFFLKE KGGLDGLIYS KKRQEILDLWVYHT QGFFPDWH NYT PGPGT RYPL CF
HIVMAL PVRPQVPLRPMT YKGAFDLSHFLKE KGGLDGLVWS PKRQEILDLWVYHT QGYFPDWQ NYT PGPGI RFPL TF
HIV2ST PVTPRVPLREMT YRLARDMSHLIKE KGGLEGLYYS DRRRRVLDIYLEKE EGIIGDWQ NYT HGPGV RYPK FF
HIV2ROD SVTPKVPLRPMT HRLAIDMSHLIKT RGGLEGMFYS ERRHKILNIYLEKE EGIIADWQ NYT HGPGV RYPM FF
HIV2D205 CKTPIVPLRPMT YKLAVDMSHFIKE QGGLEGMYYS ERRHRILDTYFENE EGIVSGWQ NYT HGPGI RYPK YF
HIV2EHOA RVRPGVPLRPMT FKLAVDMSHFLKE KGELEGIFYS ERRHKILDTYLENE EGIVSGWQ NYT HGPGV RYPK FF
SIVMM251 SVRPKVPLRAMT YKLAIDMSHFIKE KGGLEGIYYS ARRHRILDMYLEKE EGIIPDWQ DYT SGPGI RYPK TF
SIVMN PVRPRVPLRIIS YKLAVDMSHFIKE KGGLEGIYYS ERRHKILDMYLEKE EGIMPDWQ NYT SGPGP RYPK TF
SIVSTM AVHPRVPLREMT YKLAIDLSHFIKS KGGLEGIYYS ERRHRILDMYLEKE EGIVPDWQ NYT AGPGI RYPK QF
SIVAGM155 PVRPRVPLRQIT YKLAVDFSHFLKE KGGLDGIYYS DRRNKILNLYALNE WGIIDDWN AWS KGPGI RYPR CF
SIVSYK PVQPQTPLRTLT YKLAIDLSHFIKN KGGLQGMNYC EKRDEILHLYLQNE HGII.DRI NYT SGPGT RYPL IF
βC
βD
αC
αD
141
151
161
171
181
191
201
HIVNL43+ GW CYKLVPV EPDKVEE.ANKGENTSLLHPVSLHGMDDP..ERE VLEWRFD SR LAFH HV ARELH PEYFKNC
HIVBAL1 GW CFKLVPV EPEKVEE.ANEGENNCLSHPMSLHGMDDP..EKE VLAWKFD SS LAFH HV AREH. ....
HIVU455 GW CYKLVPV DPAEVEE.ATGGENNSLLHPICQHGVDDE..EKE VLMWKFD ST LALK HR AYELH PEFYKD.
HIVELI GW CYELVPV DPQEVEE.DTEGETNSLLHPICQHGMEDP..ERQ VLKWRFN SR LAFE HK AREMH PEFYKN.
HIVZ6 GW CFELVPV DPREVEE.ATEGETNCLLHPVCQHGMEDT..ERE VLKWRFN SR LAFE HK AREMH PEFYKDC
HIVANT70 GW LFKLVPV SEEEAERLGNTCERANLLHPACAHGFEDT..HKE ILMWKFD RS LGNT HV AMITH PELFQKD
HIVMVP5180 GW LFKLVPV SAEEAERLGNTNEDASLLHPACNHGAEDA..HGE ILKWQFD RS LGLT HI ALQKH PELFPK.
SIVCPZGAB GW CFKLVPL TEEQVEQ.ANEGDNNCLLHPICQHGMEDE..DKE VLVWRFD SR LALR HI AREQH PEYYKD.
HIVZ321 GW CFKLVPV DPREVEE.ANTGENNCLLHPMSQHGMDDD..ERE VLMWKFD SS LARK HL AREMH PEFYKD.
HIVMAL GW CFKLVPM SPEEVEE.ANEGENNCLLHPISQHGMEDA..ERE VLKWKFD SS LALR HR AREQH PEYYKDC
HIV2ST GW LWKLVPV DVP..QEG.DDSETHCLVHPAQTSRFDDP..HGE TLVWRFD PT LAFS YE AFIRY PEEFGYK
HIV2ROD GW LWKLVPV DVP..QEG.EDTETHCLVHPAQTSKFDDP..HGE TLVWEFD PL LAYS YE AFIRY PEEFGHK
HIV2D205 GW LWKLVPV EVP..AATREEEETHCLMHPAQISSWDDI..HGE TLIWQFD SL LAYD YV AFNRF PEEFGYQ
HIV2EHOA GW LWKLVPI NM...IAEPEDEETHCLVHPAQTSAWDDP..HEE TLVWQFD SL LAYD YV AFSRF PEEFGYQ
SIVMM251 GW LWKLVPV NVS..DEAQEDER.HYLMQPAQTSKWDDP..WGE VLAWKFD PT LAYT YE AYARY PEELEAS
SIVMN GW LWKLVPV NVS..DEAQEGEE.NYLLHPAQTSQWDDP..WGE VLVWKFD PT LAYT YE AYIRY PEEFGSK
SIVSTM GW LWKLVPV DMS..NEAQEDDGTHYLVHPAQTHQWDDP..WGE VLVWKFD PL LAHT YE AFVRH PEEFGSK
SIVAGM155 GF CFKLVPV ALH..EEA.ETCERHCLVHPAQLHEDPDGINHGE ILAWKFD PM LAVQ YD PSREY FTDLYST
SIVSYK GW LWELVPN .EI..EGCLEYEEHTLLLHPASGQGSSS...MGE .PHVELQ PP PGYT P. GWEMA RLQLERQ

βC and βD), the sequences in the HIV-1 subgrouping are highly conserved. The Nef variant used here (HIVNL43) is 81% identical to the HIVMAL sequence, the most distant in the HIV-1 grouping. Sequences in the HIV-2 and SIV groupings are more divergent. For example, with respect to HIVNL43, sequence identities for HIV2EHOA and SIVAGM155 are 57% and 49%, respectively. However, substitutions at most positions across Nef_{core} are highly conservative (except in the disordered βC-βD loop), indicating that the three-dimensional structure reported here will be a reasonable model for all Nef sequences.

There are 25 positions that are conserved without exception in all the sequences shown in Figure 3.4.4.1. The strictly conserved residues that do not appear to have a major structural role fall into two categories. Five are involved in interactions with the SH3 domain (Pro 72, Pro 75, Leu 76, Arg 77 and Asp 86) and the other three (Arg 106, Ile 109 and Leu 110) are located on the edge of the crevice between helices αA and αB (Figure 3.4.1.3a). The hydrophobic nature of the sidechains lining this crevice is conserved in all Nef sequences (Figure 3.4.1.2), and Arg 106 has been implicated in the association of Nef with Ser kinase activity (Sawai *et al.*, 1995). The arrangement of hydrophobic sidechains seen here is strongly suggestive of a binding site for a ligand (Figure 3.4.1.3a). Recently, one of the Nef-associated kinases has identified as PAK68, a member of the PAK (p21-activated kinase) family. Via the ability to interact with Rho-family G-proteins, PAKs have been implicated in various signal transduction pathways involving transcription activation and

cytoskeleton modulation. The association of Nef with the PAK may also link Nef to stimulated proviral DNA synthesis, and enhanced viral infectivity (via enhanced matrix phosphorylation).

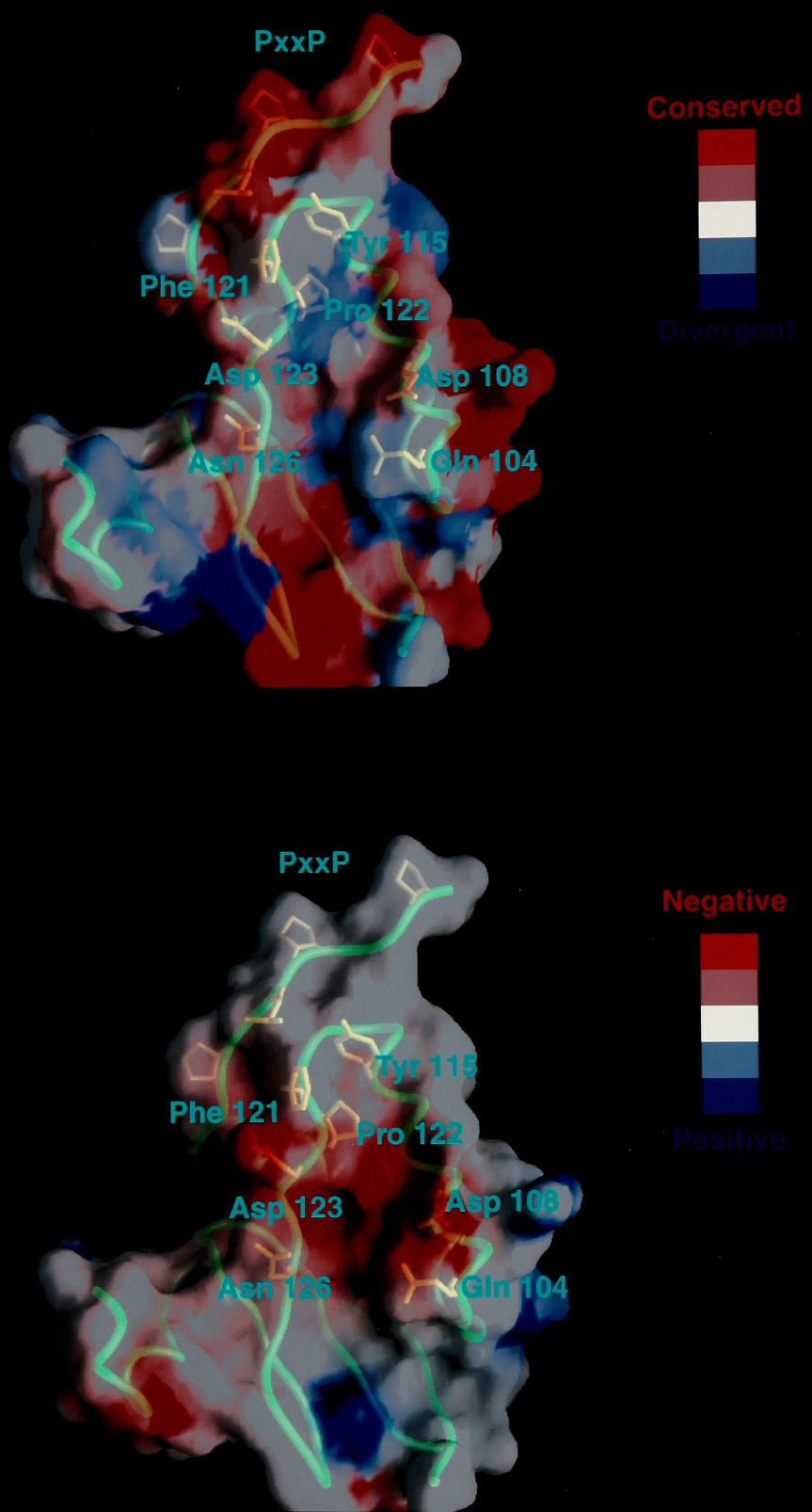
Also of interest is the interaction between Nef and the cytoplasmic tail of CD4 molecules. It was shown that Nef induced CD4 down-regulation, a process that requires a dileucine motif at the CD4 tail. Nef can interact directly with a peptide containing the dileucine motif (CD4 tail peptide) *in vitro* although the affinity is weak (~1 mM) and the biological significance of this interaction is still unclear. Nevertheless, Grzesiek and co-workers have mapped the binding site onto the Nef surface. Surprisingly, the CD4 tail peptide interacts with the region of Nef that is very close to ——or overlapping with ——the PAK binding site (Grzesiek *et al.*, 1996). In addition to the residues surrounding the hydrophobic pocket implicated in PAK binding (residues 95-97, 106, 110), the interaction between Nef and the CD4 tail peptide also involves N-terminal region of Nef, including residues 57-59. However, these structural findings are not in perfect agreement with previous functional study; a mutant Nef that is unable to recruit PAK still down-regulate CD4, albeit less efficiently than the wild type (Sawai *et al.*, 1997).

Furthermore, Nef was recently shown to interact with another host protein, a thioesterase. This interaction is also involved in Nef-mediated CD4 down-regulation (Cohen *et al.*, unpublished results). By mutational analysis, the thioesterase binding site on Nef has been mapped to a region that is distinct from the PAK (or CD4 tail) binding

site and includes helix αB and the following loop (residues 108, 112, 115, 121-123). The Nef residues involved in the interaction with the thioesterase are not very well-conserved (Figure 3.4.4.2). In addition, Nef proteins from certain immunodeficient viruses (such as SIV-MAC239) fail to efficiently down-regulate CD4, presumably reflecting the sequence divergence in this region.

In conclusion, Nef has been shown to interact with four different host proteins, including the Src family SH3 domains, the PAK family of kinases, the CD4 cytoplasmic tail, and the thioesterase. While the first two proteins appear to be involved in enhancement of viral replication *in vivo*, the last two proteins concern the down-regulation of CD4. From the Nef structure, it can be seen that the SH3 domain and thioesterase bind to distinct regions of Nef although PAK and CD4 tail might share a common binding site. Consistent with structural analysis, recent biochemical data indicated that the binding of the SH3 domain to Nef is independent of the thioesterase (Cohen *et al.*, unpublished). For example, a Nef molecule can bind to both an SH3 domain and thioesterase simultaneously. It is conceivable that Nef might function as a adapter that bridges several different signaling molecules, an idea that awaits further study (see discussion section 4.4).

Figure 3.4.4.2 The putative thioesterase binding site. The molecular surfaces of Nef_{core} are colored according to the conservation score (a) and the local electrostatic potential (b). In (a), the most conserved regions are colored in red and the least are in blue. In (b), the coloring ranges from dark blue (most positive) to deep red (most negative). The residues implicated in binding to thioesterase (see main text) are labeled in blue. The view is close to that in Figure 3.4.1.1b.


PxxP
Tyr 115
Phe 121
Pro 122
Asp 123
Asp 108
Asn 126
Gln 104
Conserved
Divergent
PxxP
Tyr 115
Phe 121
Pro 122
Asp 123
Asp 108
Asn 126
Gln 104
Negative
Positive

3.4.6 Summary

The crystal structure of the conserved core of HIV-1 Nef has been determined in complex with the SH3 domain of a mutant Fyn tyrosine kinase (Arg 96 -> Ile), to which Nef binds tightly. The conserved "PxxP" sequence motif of Nef, known to be important for optimal viral replication, is part of a polyproline type II helix that engages the SH3 domain in a manner resembling closely the interaction of isolated peptides with SH3 domains. The Nef:SH3 structure also reveals how high affinity and specificity in the SH3 interaction is achieved by the presentation of the PxxP motif within the context of the folded structure of Nef.

4 Discussions

4.1 The Binding Specificity of the SH2 Domains

4.1.1 The SH2 Specificity and its Prediction

In the search for the structural principles underpinning the SH2 binding specificity, the structures of several different SH2 domains in complexes with their cognate ligand peptides have been determined (reviewed in Kuriyan and Cowburn 1997). An important implication of these structural studies (including the structures of the SH-PTP2 SH2 domain presented in this thesis) is that the overall topology of the SH2 domains and the peptide binding modes appear to be *conserved*. To illustrate in the simplest way, the SH2 domain is divided by its central ß-sheet into two halves, with the phosphotyrosine and the residues immediately following the phosphotyrosine plunging into each of the halves (Figure 1.2.1). The "N-terminal half" which forms the phosphotyrosine binding pocket is highly conserved among different SH2 domains (Waksman *et al.*, 1992). On the other hand, the more divergent "C-terminal half" can form different binding surfaces for 3 to 5 residues C-terminal to the phosphotyrosine residue. The interactions between the SH2 domain and the peptide residues C-terminal to phosphotyrosine appears to confer the sequence specificity (Songyang, *et al.*, 1993; Larose, *et al.*, 1995).

Most SH2 domains show relatively strong selectivity for particular side chains at the +1, +3, and (in some cases) +5 positions of the peptide (except for GRB-2, discussed separately below). The reason for these observations can be readily understood from the structural data. High affinity phosphopeptides often bind to the SH2 domains in an extended conformation. As a result, significant contacts are only made by odd numbered residue (+1, +3, and in some cases, +5). In addition, since the peptide binding modes utilized by the SH2 domains are often conserved, peptide sequence preference could be correlated with side-chains of residues at several critical positions of the SH2 domains. The clearest example of such correlation is provided for the residue at the βD5 position of the SH2 domains, which contacts the peptide sidechains at position +1 (and +3). Certain SH2 domains (including Src family tyrosine kinases as well as GAP and the adaptor proteins GRB2 and Nck) have aromatic residues at βD5, and preferentially bind to peptides containing polar sidechains at +1. In contrast, other SH2 domains (including the tyrosine phosphatases, p85, and phospholipase C-γ1) contain relatively smaller hydrophobic sidechains at βD5 and select for hydrophobic residues at +1.

The structures determined so far provide an explanation for the correlation between residue βD5 (of the SH2 domains) and the peptide sequence preference at +1 position. The aromatic sidechain of residue βD5 (a tyrosine in the case of Src) closes up the binding site for the +1 residues while a less bulky sidechain (such as an isoleucine in the cases of SH-PTP2 and p85) will open up the +1 binding site

and allow it (+1 sidechain) to interact with several hydrophobic residues of the SH2 domains.

On the other hand, it would be difficult to predict the peptide sequence preference at the +3 and +5 positions solely on the basis of the primary sequence. The binding sites for peptide residue +3 and +5 are primarily determined by the BG and EF loops where large insertion or deletion often occur. For example, the SH-PTP2 SH2 domain contains a large insertion in the BG loop and distinct sequence in the EF loop with respect to the Src SH2. As described in the Result section, these alterations in SH-PTP2 result in a structural rearrangement of these two loops and a more open peptide binding surface which can accommodate a hydrophobic sidechain of the peptide residue at +5 position. A similar groove-like binding surface is also observed in the PLC-γ1 SH2 domain (Pascal, *et al.*, 1994), although it has an amino acid sequence rather different from that of SH-PTP2 in the BG and EF loops. Such drastic changes in the backbone conformation would be difficult to predict correctly by current homology modeling techniques.

Selectivity for polar sidechains at the +2 or +4 position has been described in certain SH2 domains. Since the peptide residues at +2 and +4 often make limited contacts with the SH2 domains, sequence preference at these position tends to be weak and rather difficult to predict. In several known cases (such as Src and SH-PTP2), the preference at +2 and +4 positions are determined by local electrostatic interactions provided by polar sidechains located in the

proximity of these binding sites (see Results section). The GRB2 SH2 domain is unique in that it selects most strongly for an asparagine at +2 position. This unusual target specificity is due to the presence of a Trp residue at the EF1 position, which fills up the +3 binding site, as seen in the recently determined GRB2-peptide structure (Furet *et al.*, 1996). The peptide in the sequence of Y*VNV adopts a β turn conformation and the sidechain of Asn +2 forms a specific hydrogen bonding network with the GRB2 SH2 domains. Although the importance of the Trp EF1 in determining the GRB2 SH2 specificity was predicted (based on the sequence comparison and the Src/Lck SH2 structures), the selectivity of Asn +2 was *not* understood until the structure of the GRB2-peptide complex became available (Marengere *et al.*, 1994). Thus, further structural analysis of the SH2 domains, particularly the most divergent ones (such as STAT), might be required to obtain a more complete understanding of the SH2 specificity.

4.1.2 The Discrepancy Between Structural and Binding Data

Even with a high affinity peptide-SH2 structure in hand, certain aspects of the SH2 specificity ——particularly the relative contribution of each specificity-determining residue ——are still not completely understood. For example, based on the structure of the Lck SH2 in complex with the pYEEI peptide, Lck SH2 is predicted to have a strong preference for a large hydrophobic residue at the +3 position and a weak selectivity for an acidic residue at the +1 position (see Result sections). However, a previous mutagenesis study

indicated that the peptide residue at +1 position is as sensitive, if not more so, to mutation as that at +3 position (Payne *et al.*, 1993). Most surprisingly, the replacement of the hydrophobic residue, Ile +3, with an acidic residue (Glu) results in only a modest loss of binding affinity (Payne *et al.*, 1993). It is not clear how the apparently hydrophobic pocket (for +3 residue) of the Lck SH2 can accommodate a charge residue such as Glu. One might suspect that the mutant peptide adopts a different conformation in binding to the Lck SH2, an possibility that remains to be tested.

The SH2 domain of PLC-γ1 represents a perplexing example in which the result of the mutagenesis study can not be readily explained by the structural findings. Forman-Kay and co-workers have determined the solution structure of the C-terminal SH2 domain of PLC-γ1 in complex with its cognate ligand peptide (PDGFR-1021; sequence, NNDY*IIPLPDPK) (Pascal *et al.*, 1994). In the structure, the binding site for the peptide consists of an extended groove that binds the 6 residues that follow the phosphotyrosine. A surprising aspect of the recognition of the PDGFR-1021 peptide by the PLC-γ1 SH2 domain is that although 8 residues of the peptide are seen to interact with the SH2 domain, truncation of the peptide to just 3 residues (DY*I) resulted in only a 15 fold reduction in binding affinity (Shoelson, unpublished). The binding data thus clearly shows that the majority of the binding energy comes from just the phosphotyrosine and the two residues that immediately flank it although, as seen in the structure, an extensive set of residues of the peptide interacts closely with the SH2 domain.

The discrepancy between the structural results and the binding data has been further investigated by examining the changes in the dynamic properties of the SH2 domain upon peptide binding (Pascal, *et al.*, 1994). This analysis demonstrates that the residues interacting with the phosphotyrosine (which contribute to binding energy) undergo a significant restriction in dynamic flexibility upon binding. On the other hand, the residues interacting with the C-terminal end of the peptide (which contribute little binding energy) do not. These results have general implications in studying molecular interactions. It has been suggested that the analysis of dynamic behavior in response to ligand binding could be used to distinguish residues that contribute most to binding energy from those that do not. Such methodology, if applicable in general, would serve as an efficient way for understanding SH2 specificity. In contrast, the current methodology which relies on the combination of structural analysis and systematic mutagenesis is relatively slow.

4.1.3 Does the *in vitro* Specificity Observed for Peptide Account for the *in vivo* Specificity?

A direct correlation between the cellular functions of the SH2 domains and their specificity observed *in vitro* was reported by Pawson and co-workers (Marengere *et al.*, 1994). Their work has shown that a single amino acid mutation of the Src SH2 domain, from Thr to Trp at the EF1 position, can alter its binding specificity to mimic that of the Sem-5 SH2 domain *in vitro* (as described in Section

4.1.2). Furthermore, this mutant Src SH2 domain can functionally substitute for the Sem-5 SH2 domain in vulval induction of *C. elegans* ——presumably by mimicking the binding specificity of the Sem-5 SH2 domain *in vivo*. Thus, Pawson's study indicated that the *in vitro* peptide binding specificity of the SH2 domain could, at least partially, account for the specificity observed *in vivo*.

Conceivably, the tertiary structure in which the phosphotyrosine-containing region is presented may influence the binding affinity and specificity. However, the observation that a mutant PDGF receptor containing a deletion near the GAP-SH2 binding site binds to the GAP-SH2 domain to nearly the same degree as the wild type PDGF receptor suggests that the tertiary structure may not be a primary factor in determining binding affinity (Kashishian *et al.*, 1992). Nevertheless, *in vivo* mutational analysis indicated that, for certain receptors, the difference between the binding affinities of specific and non-specific phosphorylation sites are much greater than that observed for isolated peptides. In several cases, it is possible to eliminate the binding of an individual SH2 domain by removing a single tyrosine phosphorylation site. For example, single mutation of tyrosine 766 in EGF receptor results in undetectable binding of phospholipase C-γ1 (PLC-γ1) to the receptors. The all-or-none effect demonstrated by mutagenesis study *in vivo* suggests that the SH2 domains might be capable of mediating —— to a certain extent ——tertiary interactions with tyrosine-phosphorylated proteins and such interactions may "fine-tune" the binding affinity and specificity.

4.2 Peptide Recognition by the X-11 PTB Domain

4.2.1 The Comparison with the SH2 Domains

The discovery of the PTB domains originated from their ability to bind to phosphorylated tyrosine residues in signaling proteins (Kavanaugh *et al.*, 1994), but with sequence specificity that is fundamentally distinct from that of the SH2 domains (Songyang, *et al.*, 1992). Subsequent structural analyses of phosphopeptide complexes of the Shc (Zhou, *et al.*, 1995) and IRS-1 (Eck, *et al.*, 1996; Zhou, *et al.*, 1996) PTB domains showed that the architecture of these domains is related to that of the PH domains, but not to that of SH2 domains. Significantly, the recognition of phosphotyrosine residues by the PTB domains relies on basic residues in the domain that are not conserved (Figure 3.2.2.3). The binding site for the phosphotyrosine residue lies on the edge of the domain, and the positively charged sidechains that interact with the phosphate group are exposed to solvent.

The structures of the X11/APP peptide complexes show that the recognition of the unphosphorylated form of the APP peptide by the X11 PTB domain occurs by a mechanism that is essentially the same as that utilized by the Shc and IRS-1 PTB domains to bind to phosphorylated targets. The critical elements of the recognition are the β-turn formed by the NPxY motif, and the incorporation of the peptide into a β sheet of the PTB domain. A striking feature of the

X11/APP complex is that the unphosphorylated tyrosine residue of the NPxY motif is bound in virtually the same conformation as is seen for the corresponding phosphotyrosine residue in Shc and IRS-1.

Quantitative binding studies demonstrate that addition of a phosphate group to the tyrosine residue of the APP peptide does not significantly change the binding affinity for the X11 PTB domain (Table 3.2.1.1). Several charged residues that are in the vicinity of the tyrosine hydroxyl group are available to coordinate the phosphate group in the phosphorylated form of APP (Figure 3.2.3.2), but clearly they do so with no net stabilization of the complex. This is in contrast to the situation for Shc and IRS-1, each of which has a similar constellation of positively charged residues at the phosphotyrosine binding site. Shc and IRS-1, however, have been shown to clearly discriminate between phosphorylated and unphosphorylated targets (Mandiyan *et al.*, 1996; Zhou *et al.*, 1996). It thus appears that different PTB domains have evolved different requirements for phosphorylation of the tyrosine of the NPxY motif in their targets.

The binding specificity of well characterized SH2 domains, on the other hand, appears to be restricted quite sharply to phosphorylated peptides. In contrast to the PTB domains, which bind tightly to their peptide targets by extensive hydrogen bonding and packing interactions, the SH2 domains have relatively limited interactions with their target peptides. Consequently, the

phosphotyrosine residue provides the major anchor point for the peptide to the SH2 domain. A comparison of the phosphotyrosine binding sites of the PTB and SH2 domains reveals why phosphorylation of the target is dispensable in the case of certain PTB domains, but is absolutely required in the case of the SH2 domains. In the following, the mechanism of phosphotyrosine recognition by the SH2 domains is illustrated using the structure of the v-Src SH2 domain, for which crystal structures in the complexed and uncomplexed form are available (Waksman *et al.*, 1993).

The structure of the uncomplexed v-Src SH2 domain was determined at 2.5 Å resolution, with 4 crystallographically independent molecules in the crystallographic asymmetric unit (Waksman *et al.*, 1993). The conformation of the phosphotyrosine binding site differs slightly in each case. In one molecule, the binding site resembles closely that seen in peptide complexes, due to the binding of a phosphate ion. In another, no phosphate ions have been localized, and the binding site is somewhat more open than in peptide complexes. The other two molecules are intermediate in structure. In each of the 4 molecules, the strictly conserved residue Arg 175 (βB5 in the SH2 notation) rises up from the interior of the domain into a position where it can engage two oxygens of the phosphate group (Figure 3.2.3.2). This sidechain is quite buried, even in the most open of the four SH2 molecules, with total surface accessibility ranging from 2 to 6 $Å^2$ in the four molecules (Figure 3.2.3.2). This may be compared with a surface accessibility of ~100 $Å^2$ for Arg 155 (αA2), an SH2 residue that engages the

phosphotyrosine from the surface. Analysis of the structures of other uncomplexed SH2 domains, such as that of SH-PTP2 (Lee *et al.*, 1994; presented in this thesis), reveals that the buried nature of Arg βB5 is a conserved feature of the SH2 architecture (the solvent accessible surface for Arg βB5 is ~11 $Å^2$ in SH-PTP2).

The burial of Arg βB5 in the SH2 structure is likely to have a significant energetic penalty, and is probably a critical architectural feature that then demands that the ligand be phosphorylated in order to neutralize the residue. In X11, five basic residues ((Lys 347, Arg 432, Arg 433, Arg 434 and Arg 354) are in a position to potentially interact with the phosphate group of phosphorylated ligands. These residues are all on the surface, and the lack of stabilization of phosphorylated versus non-phosphorylated APP peptide is likely to be a consequence of an unfavorable desolvation energy for these sidechains. However, the situation for the PTB domains is clearly subtle, since the Shc and IRS-1 PTB domains do bind much more strongly to phosphorylated peptides than they do to nonphosphorylated ones (Kavanaugh *et al.*, 1994; Trub *et al.*, 1995; Wolf *et al.*, 1995; Batzer *et al.*, 1995). As in X11, in the Shc and IRS-1 PTB the charged sidechains that interact with the phosphotyrosine are solvent accessible and on the surface. The net stabilization of phosphorylated peptides in this case must be the result of particularly favorable geometry of interaction, an issue that will be of interest to investigate further.

4.2.2 Biological Implications of the X11-APP Peptide Complex Structure

Sequence-specific recognition of peptides by the X11 PTB domain involves a number of residues distributed over a relatively long region (10~13 residues) of the peptide ligands. The differences in binding specificity between X11 and the other well-characterized PTB domains indicates that the PTB domains are capable of mediating a broad spectrum of sequence-specific interactions. In particular, the finding that C-terminal specificity can also be a feature of PTB-mediated interactions further extends the specificity of the PTB domain. Thus, the overlapping specificity between Shc and IRS-1 PTB domains likely reflects their shared biological functions in targeting common growth factor receptors instead of a limited repertoire of the PTB-mediated interactions.

The use of secondary structural elements for peptide/protein recognition ——particularly by incorporating β-strands in the ligand into existing structure ——has been identified in several other cases, including the recently determined crystal structure of a PDZ/peptide complex (Doyle *et al.*, 1996). Such a setup creates a continuous β-sheet structure that would allow additional strands to be added on (Harrison, 1996). It is thus interesting to note that the secondary structure of the cytoplasmic domain of β-APP was predicted to contain a short β-strand and a long helix before the NPxY motif at its C-terminus.

The understanding of the biological function of the X11/β-APP interaction has been impeded by our limited knowledge of these two proteins. While β-APP appears to play a role in learning processes through an unknown mechanism (Muller *et al.*, 1994), there has been no known function assigned to X11. However, the measured high specificity and high affinity as well as the conserved peptide binding mode observed in the X11/APP peptide structure suggests that the interaction might be relevant. The PTB domain of X11 appears to be solely responsible for its interaction with β-APP, since no enhanced or cooperative binding to β-APP was observed for the full-length X11 protein (Margolis *et al.* unpublished results). The present study reveals the critical determinants for X11/β-APP interactions. Such information should facilitate further dissection of physiological functions of β-APP as well as the pathology of Alzheimer's disease.

4.3 The SH3-mediated Interactions

4.3.1 Interactions with the PP-II Helix: SH3 and WW Domains

A conserved feature of the PxxP/SH3 interaction is that the peptide ligands usually adopt a PP-II helix conformation in binding to the SH3 domains. The use of the PP-II helix appears to be an economic solution for the relatively small interaction surfaces of the SH3 domains. However, the pseudo three-fold symmetry, an intrinsic property of the PP-II helix, allows ligands to bind to the SH3 domains in two opposite orientations. In most of the SH3 domains, the duality in binding orientation is resolved by employing a conserved charge pairing between basic residues of the peptides and acidic residues of the SH3 domains —— thereby restricting the ligands to bind in only one of the two possible directions.

The use of a PP-II helix is not, however, a unique feature of SH3-mediated interactions. Another peptide-interacting module, the WW domain, also interacts with proline-rich peptides (in the sequence of PPxY) that adopts a PP-II helix conformation. The WW domains, found in a variety of cytoskeletal and signal-transducing proteins, are small protein modules (~38 amino acids) that share no sequence homology nor structural fold with the SH3 domains. Like the SH3-mediated interactions, the PPxY-WW interaction involves mostly hydrophobic interactions that are mediated by a set of conserved aromatic residues on the binding surface of the WW

domains. However, there is no specific ion pairing found. Instead, the orientation of the bound peptide appears to be defined by the position of the tyrosine of the PPxY motif, which binds to a large hydrophobic pocket on the surface of the WW domains. Furthermore, several peptides bind to MHC class I and II molecules in a left-handed helical conformation that is very similar to the PP-II helix. In these cases, the orientation of the bound peptides appear to be determined by the interactions between the termini of the peptide and the two pockets on the opposite ends of the binding groove of the MHC molecules.

4.3.2 The Affinity and Specificity of the SH3-mediated Interactions

The SH3 domain/PxxP peptide interactions are, in general, of low affinity (K_D = 5~100 μM) and low selectivity. The low affinity is likely due to the relatively small and flat interaction interface (~700Å^2 including both SH3 and the peptide). Although low affinity doesn't necessarily confer low selectivity (see appendix A), the highly conserved binding surface of the SH3 domains (used for binding to the PP-II helix) leaves little room for specificity. Except for a conserved surface salt-bridge (described above), the interactions between SH3 and the bound peptide are mostly hydrophobic in nature. The residues that form this surface salt-bridge (Arg_{-3} of the peptide and a conserved acidic residue of the SH3) are exposed to solvent —— thereby suggesting that this salt-bridge is most likely destabilizing. On the other hand, the Abl

SH3 domain, which contains a threonine in the place of the conserved acidic residue (referenced above), appears to bind with higher affinity to certain proline-rich peptides (such as 3BP1 and 3BP2) that contain a large hydrophobic residue at the -3 position (instead of a basic residue). These features of the SH3-PxxP interaction appears to be consistent with the general view that binding energy is provided by hydrophobic effects while complementary hydrogen bonds and electrostatic interactions confer specificity (Fersht *et al.*, 1985).

Furthermore, to bind to the SH3 domains, peptides are required to adopt a specific secondary structure (i.e. the PP-II helix), a process that would incur a conformational entropy penalty. Free SH3 ligand peptides are known to adopt the PP-II conformation to a certain degree in solution. Nevertheless, increasing the number of the prolines in the peptides often results in enhanced SH3 binding ability (Yu *et al.*, 1994; Hanafusa *et al.*, unpublished results), thereby suggesting that the conformational entropy penalty might be considerable for most of the SH3 ligand peptides. Likewise, the Nef PxxP peptide which contains only the two PxxP-defining prolines binds to the Hck or Fyn SH3 domains with very low affinity (100~200 μM).

The biological function of the SH3 domains is supposed to be to mediate highly specific protein-protein interactions. As the interaction between SH3 domains and PxxP peptides is intrinsically weak and promiscuous, additional mechanisms are expected to exist for enhancing both affinity and specificity. The simplest way to

achieve this is to place the SH3 domain in the proximity of its ligand, thus increasing the local concentrations of both interacting partners. A example for using such a mechanism is provided by the recently determined structure of a fragment of Itk tyrosine kinase (Andreotti *et al.*, 1997). In this structure, the Itk SH3 domain and the adjacent proline-rich region forms an intramolecular interaction that resembles a typical SH3-PxxP interaction. It was suggested that this SH3-mediated intramolecular interaction might play a regulatory role in the Itk kinase as well as in other Tec family kinases (Andreotti *et al.*, 1997).

The Nef-SH3 interaction, described in this thesis, utilizes two distinct mechanisms to achieve tight and specific binding. The Nef PxxP motif, present in the context of full-length Nef protein adopts a PP-II helix even in the absence of the SH3 domain (Grzesiek *et al.*, 1996). Thus, the entropic penalty for forming the PP-II helix has been paid during protein folding. In addition, the interaction involves regions of Nef other than the PxxP motif, thereby further enhancing the specificity —— and possibly the affinity. This region of Nef targets the RT-loop, one of the two most divergent regions of the SH3 domains. The interactions that are additional to the typical SH3-PxxP interactions allow Nef to distinguish between two highly homologous SH3 domains, those of Fyn and Hck.

The additional interaction provided by the divergent RT-loop of SH3 domain and regions other than the PxxP motif may be a general mechanism utilized for enhancing SH3 binding specificity. A similar

mechanism is also suggested by the structures of two Src family tyrosine kinases, Hck and Src, recently determined in their autoinhibited forms (Sicheri *et al.*, 1997; Xu *et al.*, 1997). The autoinhibition of Src/Hck TK involves an intramolecular interaction mediated by the SH3 domain, in addition to an intramolecular interaction between the C-terminal tyrosyl-phosphorylated tail and the SH2 domain. The linker region between SH2 and kinase domains adopts a PP-II conformation in binding to the SH3 domain. However, the interaction seems to be less optimal since several hydrophobic pockets on the SH3 surface are not fully utilized. Additional contacts with the kinase domain are made by the RT-loop of the SH3 domains. The intramolecular interaction mediated by the SH3 domain is crucial for positioning the N-lobe of the kinase in an inactive conformation. Mutations in the RT-loop of the SH3 domain, or occupying the binding surface of the SH3 domain, activates the Src/Hck kinase activity —— presumably by disruption the SH3-mediated intramolecular interaction. Taken together, the intramolecular interaction mediated by the Src/Hck SH3 domain includes a SH3-PxxP interaction as well as the additional interactions via the RT-loop. The intramolecular interactions mediated by both SH2 and SH3 domains are required for the kinase to maintain in an inactivated state. Furthermore, substrate recognition by Src tyrosine kinases also involves cooperative binding by both SH2 and SH3 (Alexandropoulos and Baltimore, 1996).

In summary, previous studies have established that the SH3/PP-II interactions are rather promiscuous in general and an

additional level of complexity is often required to give rise to more specificity. Identified so far are several mechanisms including (i) increased local concentration, (ii) tertiary interaction, and (iii) synergistic interactions mediated by multiple domains. These mechanisms do not, however, appear to be specific for the SH3 domain and may be as well utilized by other peptide recognition domains (such as PDZ, WW, and SH2 domains). Furthermore, the binding surface of the SH3 domain appears to be optimal for, but not restricted to, the interaction with PP-II helix from the structure point of view. Indeed, it was recently reported that the P53 core domain binds to the SH3 domain of 53BP2 (P53 binding protein 2) in a manner that is distinct to the typical SH3-PxxP interaction (Gorina and Pavletich, 1996). Although its biological significance is still unclear at the moment, this interaction is of high affinity (in nM range). Thus, other types of PPII-independent interaction mediated by the SH3 domain might as well exist.

4.4 The Biological Functions of Nef: The Implication of the Nef Structure

Pray, do not mock me:
I am a very foolish fond old man,
Fourscore and upward, not an hour more nor less;
And, to deal plainly,
I fear I am not in my perfect mind.
Methinks I should know you, and know this man;
Yet I am doubtful for I am mainly ignorant...

King Lear

William Shakespeare, *King Lear*

Perhaps, it is not surprising that our understanding of how Nef functions is still limited after years of research and hundreds of published papers. As one might guess, the problem is due to the great complexity of Nef functions as well as the difficulty of relating cellular and molecular findings to the results of animal and clinical studies. The determination of Nef structure provides a structural framework for interpreting the properties of various Nef mutants and makes possible a thorough structure-function analysis.

The importance of Nef in HIV pathogenesis and the development of AIDS has been well-established by clinical and animal studies. In addition, using the tissue culture systems, a panel of cellular functions of Nef has been demonstrated, which include (i) down-regulation of CD4 and MHC-I molecules, (ii) activation/inhibition of lymphocytes, and (iii) enhanced viral replication and infectivity. As suggested by previous mutational studies, these functions might not be inter-dependent or directly

related. It is not clear, however, which particular cellular function(s) of Nef is most critical for disease progression. In addition, little is known about how Nef exerts these cellular functions.

Nef has no demonstrable catalytic activity and the determined Nef structure resembles no known protein. It is likely that Nef might function through its interactions with host proteins. Along the same line, Nef has been shown to interact with three host proteins (including the thioesterase, β-COP and the cytoplasmic tail of CD4) that are implicated in accelerated internalization of CD4 and MHC-I molecules (Mangasarian *et al.*, 1997). It has been proposed that Nef accelerates the internalization of these cell-surface receptors by acting as a physical connector to the endocytic apparatus although the mechanism utilized is still unclear. As described in the Result section, the interacting sites for the thioesterase and the CD4 tail have been mapped onto two distinct regions of Nef, although the binding site for β-COP is still not known). Further mutagenesis study is likely to sort out which interaction is most critical for the down-regulation of CD4 (or MHC-I) molecules and whether these cellular functions can account for the biological effects of Nef in animals and humans.

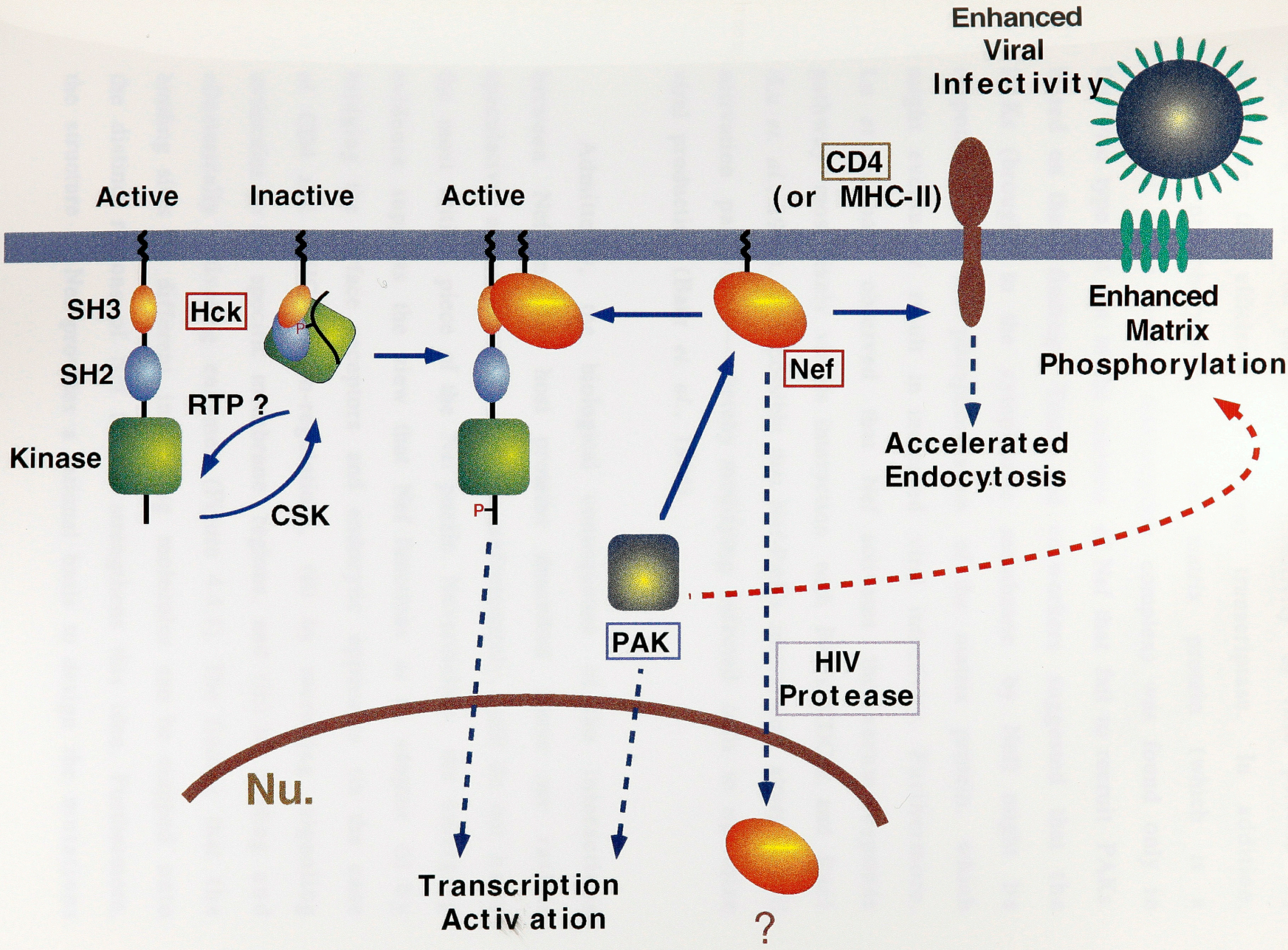
Also of interest is the interactions of Nef with two types of protein kinases, namely, Src family tyrosine kinases (Src TKs) and p21-activated kinases (PAKs). The Src family tyrosine kinases associate with Nef through their SH3 domains. In addition, these SH3 domains also mediate intramolecular interactions (with an internal

PxxP motif in Src TKs) required for the auto-inhibition of the kinase activity. Our recent biochemistry study has shown that the binding of Nef to Hck (via SH3 domain) leads to activated kinase activity, apparently by disrupting the SH3-mediated intramolecular interaction. Given the broad range of signaling pathways mediated by Src-TKs, it is conceivable that its activation by Nef may result in an altered activation status of T-cell (or macrophage), thereby, facilitating viral replication (Skowronski *et al.*, 1993).

The fact that the SH3-interacting elements defined by the structural analysis are highly conserved in Nef sequences suggests the biological relevance of the Nef-SH3 interaction. However, it is worth noting that the *in vivo* interaction between Nef and Src-TKs has not yet been demonstrated and the authentic Nef ligand might be an as-yet-unidentified SH3 domain. In addition to its critical roles in Nef's association with the SH3 domains and in the enhanced viral replication, the functional PxxP motif is also required for the membrane localization of Nef (Saksella, *et al.* unpublished result). As several cytoskeleton proteins are known to contain SH3 domains, one might speculate that Nef anchors on cytoskeleton in order to resume membrane location that is critical for its functions (Baur *et al.*, 1994).

The recently identified interaction between Nef and PAKs has drawn considerable attention because it might provide a clue for how Nef exerts several of its functions ——such as the enhanced proviral DNA synthesis, elevated matrix phosphorylation, and increased viral infectivity (Nunn and Marsh, 1996; Sawai *et al.*, 1996). It has been

Figure 4.4.1 Proposed model for the pleiotropic functions of Nef.


Enhanced
Viral
Infectivity
CD4
(or MHC-II)
Active
Inactive
Active
SH3
SH2
Kinase
Hck
P
RTP ?
CSK
P
Nef
Enhanced
Matrix
Phosphorylation
Accelerated
Endocytosis
PAK
HIV
Protease
Nu.
Transcription
Activation
?

suggested that Nef enhances the infectivity of viral particle by increasing the efficiency of reverse transcriptase. In addition, increased phosphorylation of the matrix protein (which is a component of the reverse transcription complex) was found only in the wild type but not in the mutants of Nef that fail to recruit PAKs. Based on these findings, Trono and co-workers suggested that the PAKs (brought to the cytoplasmic membrane by Nef) might be responsible for the phosphorylation of the matrix protein, which might eventually result in increased viral infectivity. Furthermore, Lu *et al.* have observed that Nef activates the serum-response pathway most likely via its interaction with PAKs, CDC42, and Rac1 (Lu *et. al.*, 1996). It is possible that Nef-PAKs interaction alters T-cell activation pathway ——thereby modifying infected cells to optimize viral production (Baur *et. al.*, 1994).

Admittedly, the biological consequence of the interactions between Nef and the host proteins described above are rather speculative at the moment. And we, unfortunately, still do not have the most critical piece of the Nef puzzle. Nevertheless, the emerging evidence supports the view that Nef functions as an adaptor (i) by bridging the surface receptors and endocytic apparatus (in the case of CD4 and MHC-I down-regulation), (ii) by recruiting signaling molecules to the specific membrane regions, and (iii) by binding and allosterically activating enzymes (Figure 4.4.1). The finding that the binding sites for different interacting molecules can be mapped onto the distinct regions of Nef further strengthens this idea. Furthermore, the structure of Nef provides a rational basis to design the mutations

that are likely to disrupt specific interactions without perturbing the structural integrity. Such a study will not only provide insight into the pleiotropic functions of Nef but also ——as it frequently happens —— lead to a better understanding of normal cell biology.

Appendix A. Specificity, Selectivity, and Affinity

Researchers working in different disciplines such as enzymology, pharmacology, and physiology, might perceive the term "specificity" in several different ways. In the following discussion, I will use a definition recently described by Eaton *et. al.* who referred to "specificity" as "a functional discrimination between molecules competing for a common ligand". For simplicity, such discrimination will be mainly referred to as the difference in free energy of a given interaction (*i.e.* from a thermodynamic point of view). For simplicity, kinetic discrimination will not be discussed Also, another ambiguous term, "selectivity", is defined as the ability of a given molecule to discriminate against certain modifications of its cognate ligand (see below).

First of all, let's consider the case of a simple bimolecular interaction in which a protein P interacts with its ligand L. By definition, the equilibrium dissociation constant, K_D (which defines affinity), relates the concentration of the complexed and uncomplexed species:.

$$P + L \Leftrightarrow P{:}L; \quad K_D = [P]\cdot[L]/[P{:}L]$$

Similarly, in the case of a second ligand N binds to P with a constant, K_n, and competes with L. It follows then:

$$P + N \Leftrightarrow P{:}N; \quad K_n = [P]\cdot[N]/[P{:}N]$$

If the competing ligand, N is a mutant form of L, I then define selectivity, s_s as the ability of protein P to discriminate N from L (*i.e.* select against the mutation). In short, $s = K_N / K_D$ It is also interesting

to see that selectivity could be additive, if the effects exerted by the mutations are not inter-dependent. For example, if protein P can discriminate i number of independent mutations with selectivity s_1, s_2 .. s_i, the total selectivity s (against all types of mutations) will be $s = \prod_{m=1}^{i} s_m$.

An analysis for the definition of "specificity" was provided by von Hippel and Berg from a thermodynamic viewpoint at the equilibrium state. Although the original treatment was given for protein-DNA interactions, the idea can be generalized for almost any type of molecular interaction. Given the interactions described above, let us now consider that protein P may interact with a variety of other ligands, B_i.

$$P + B_i \Leftrightarrow P{:}B_i; \qquad K_{bi} = [P]\cdot[B_i]/[B_i{:}N]$$

A measure (α_s) for the discrimination of Protein P in terms of its specific ligand L relative to the nonspecific ligands B_i (i.e. specificity) is defined as following:

$$\alpha_s = [P{:}L] / \sum_i [P{:}B_i] \quad \text{or} \quad \alpha_s = ([P]\cdot[L]/K_D) / (\sum_i [Bi]\cdot[P]/K_{Bi})$$

$$\alpha_s = ([L]/K_D) / (\sum_i [B_i]/K_{Bi}) \quad \text{or} \quad \alpha_s = [L] / (\sum_i [B_i]/s_{Bi})$$

Thus the specificity doesn't depend on the concentration of protein P. It depends, however, on the affinities of ligand L and all other competing species —— as well as their concentrations. Also, the absolute affinity of L is less important than the ratio between the affinity of L and that of the competing species B_i. Therefore, one can have a high specificity with a relatively low affinity provided that

the affinity of L is much greater than the affinity of nonspecific ligands B_i.

With these terms defined, now we can turn to the question: "how specific is specific enough?" Since we often don't know the effective concentrations of the proteins of interest, a few assumptions have to be made. First, let's assume the protein of interest (say, a SH2-containing molecule) and its cognate ligand have a cellular concentration of 25 nM which is equal to the average protein concentration for a cell with 10^5 different proteins of average mass 40 kD at a total concentration of 100 mg/ml. Without the competing ligands, the fraction of the desired complex, $[P:L]/P_{total}$, is calculated as 53% (for K_D=10 nM) and 2% (for K_D=1 μM). However, if one considers the presence of 50 competing ligands whose concentrations are all 25 nM, the fraction of the desired complex decreases depending on the selectivity (i.e. the ratio between the affinity of the cognate ligand and those of the competing ligands). If the selectivity is 1000, the fraction of the desired complex, $[P:L]/P_{total}$, is calculated as 51% (for K_D=10 nM), 17% (for K_D=100 nM) and 2.3% (for K_D=1 μM). However, if the selectivity is as low as 100, the fraction of the desired complex is calculated as 40% (for K_D=10 nM), 15% (for K_D=100 nM) and 2% (for K_D=1 μM). Thus, in order to have the fraction of the desired complex greater than 50%, the K_D and the selectivity have to be equal to or better than 10 nM and 1000, respectively. It is worth noting that these calculations account for only one protein that binds to multiple ligands. One would imagine that in the case of multiple proteins competing for multiple

ligands, the fraction of the desired complex would be even lower than those values described above. In addition, the *in vitro* measurements indicated that the affinity and selectivity for the known peptide recognition modules (such as SH2, SH3, and PTB domains) are about one or two order of magnitude lower than these values. It might be reasonable to assume that additional mechanisms are used in cells.

As discussed in the main text, it is likely that tertiary interactions play a role in the enhancement of the affinity and selectivity for the peptide-recognition modules. The effects exerted by tertiary interactions are certainly case-dependent. Here I will only discuss the factors that have more general implications. For example, one would imagine that the peptide-recognition modules might be compartmentalized or restricted to certain regions of the cells. Thus, the effective concentrations for both modules and ligands would be much higher than those described above. As a result, for the binding modules with moderate affinity and selectivity (ex. 100 nM and 100 respectively), the fraction of the desired complex would appear to be reasonable high (40%, compared 17% without compartmentalization) if the effective concentrations for all the interacting partners (including non-specific ones) increase by 10 fold.

Another possibility to increase the local concentration of the binding modules is to have significant amount of non-specific or low-affinity ligands located near the cognate ligands. This is analogous to sequence-specific DNA-protein interactions which can

be facilitated by including non-specific DNA. The effect is a reduction in the search dimension from three to two, as in the case of DNA binding proteins. Thus, it is interesting to note that certain peptide binding modules, such as Shc PTB domains also interact with phospholipids, albeit with much lower affinity. Shc PTB domain uses two distinct binding regions for peptide and phospholipids (Zhou *et al.*, 1995). It is conceivable that the binding of PTB domains to phospholipids would recruit the PTB domains to the cytoplasmic membrane where the tyrosyl-phosphorylated receptors located. On the other hands, the SH2 domains appear to use the same binding sites for both phosphopeptides and phospholipids. It has been suggested that PtdIns [3,4,5]P3 compete the binding of the SH2 domains to tyrosine-phosphorylated proteins (Rameh *et al.*, 1995). It remains a critical issue as how the phospholipids actually regulate the interactions mediated by the phosphopeptide binding domains.

Appendix B. Molecular Replacement

As the number of determined structures increases rapidly, molecular replacement which utilizes known structures to solve the "phase problem" has become exceedingly important. The structure of the SH-PTP2 SH2 domain, presented in this thesis, was determined using this technique. This appendix does not intend to review the theory of molecular replacement in detail. Instead, the main purpose of this appendix is to provide an overview of this technique and to illustrate how it can be applied using Xplor (Brünger, 1992).

The method of molecular replacement relies on the availability of the three-dimensional structure of a homologous protein, which will serve as a "phasing model". Once the known structure is placed in the corrected position of the crystallographic frame of the new protein, it provides an initial phase set to generate the initial electron density map according to the following equation:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}^{obs}|\cdot Exp(-2\pi i(hx+ky+lz-\alpha_{hkl}^{model}))$$

The structure can then be further modified and refined to reveal the desired structure. Thus, the main problem is to transfer the know protein molecular structure to the crystal arrangements of the protein for which the structure is not known.

To place the known structure in the target unit cell requires the knowledge of the proper orientation and precise position of the unknown structure. In other words, the know structure relates to the

target structure (in its crystallographic frame) by a rotation and a translation matrices that are needed to be solved.

$$X_{unknown} = [C_{rotation}] \cdot X_{known} + [C_{translation}]$$

A direct six-dimensional search for both rotation and translation matrices is, for most cases, still not computationally feasible at the moment. Nevertheless, without knowing the position of the unknown structure, its orientation (with respect to the model) in the target unit cell can be determined by using the Patterson function (Rossmann and Blow, 1962). Thus, the search can be broken down into two three-dimensional searches ——first a rotation and then a translation search. The two main functions involved, rotation and translation functions are described in more detail below.

The principle of the rotation search depends on the properties of the Patterson function which is given as the following:

$$P(u,v,w) = \frac{1}{V}\sum_{hkl}|F(h,k,l)|^2 \cdot \cos[2\pi(hu+kv+lw)]$$

or in short $$P(\mathbf{u}) = \frac{1}{V}\sum_{s}|F(\mathbf{S})|^2 \cdot \cos[2\pi\,\mathbf{u}\cdot\mathbf{S}]$$

where u,v, w are relative coordinates in the Patterson unit cell (in real space). It can be shown that the Patterson function, $P(\mathbf{u})$ can be written as : $$P(\mathbf{u}) = \int_{r_1} \rho(\mathbf{r}_1) \times \rho(\mathbf{r}_1 + \mathbf{u})\, dv$$

Thus, the Patterson map is a vector map: vectors between atoms in the real structure show up as vectors from the origin to maxima in the Patterson map. These vectors can be divided into two categories: self-Patterson vectors and cross-Patterson (Hoppe, 1957). Self-Patterson vectors are derived from pairs of atoms that belong to the same molecule while cross-Patterson are those from different

molecules. The lengths of the self-Patterson vectors depend on the diameters of the structure and are usually shorter than cross-Patterson vectors. If there were no cross-Patterson vectors, the inner region of the Patterson map would be similar for the homologous molecule in different crystal frame, apart from a rotation difference. Thus, the self-Patterson vectors can supply us with the rotational relationship between the known and the unknown molecular structures.

The rotation function used for rotation search is defined by Rossman and Blow (1962) : $R(\alpha,\beta,\gamma) = \int_U P(\mathbf{u}) \times P_r(\mathbf{u}_r) d\mathbf{u}$ where α, β, and γ are the rotation angles defined in Eulerian angles (θ_1, θ_2, and θ_3 in Xplor notation) and U is the volume in the Patterson map where the self-Patterson peaks are located. The original Rossman and Blow method is performed in the reciprocal space. A conceptually simpler method is proposed by Huber and Steigemann (1985) in which the Patterson map is rotated in real space (after origin removed) and then superimposed on the original map. Xplor employed this real-space Patterson search method.

In addition to the Eulerian angular angles, two other conventions are also used for the rotation angles: the pseudo-orthogonal Eulerian angles (Lattman, 1972) and the spherical polar angles (ψ, ϕ, and κ in Xplor notation). The Lattman angles (θ_+, θ_-, and θ_2) are related to the Eulerian angles by the following: $\theta_+ = \theta_1 + \theta_3$; $\theta_- = \theta_1 - \theta_3$; $\theta_2 = \theta_2$

The Eulerian angles (or Lattman angles) are usually used for cross-rotation search (*i.e.* for relating the known structure to the unknown structure in the target unit cell) because the symmetry of the rotation function shows up clearly. On the other hand, the spherical polar angles is best for self-rotation function which is used to locate noncrystallographically related molecules in the same unit cell. This is because the angular angle, κ, directly related the noncrystallographically-related molecules. For example, a proper noncrystallographic two fold symmetry will show up nicely at $\kappa = 180^{o}$.

The rotation search should be limited to the asymmetric unit of the rotation function which is related to the Patterson symmetry of the target space group. The Patterson symmetry for all space groups has been derived and tabulated by Rao (Rao *et. al.*, 1980) and has been incorporated in the new version of Xplor (v3.851; Brunger, 1996). The grid for the first coarse rotational search should be smaller than the convergence (~10°) of the Patterson-correlation refinement that refines the highest peaks generated from the first rotational search (see below).

It is obviously important to include regions of vector space where the self-vector density is high, and to omit those where cross-vector density is high. For an approximately spherical molecule, Blow suggested to include only Patterson vector lengths less than 75-80% of the diameter of the molecule, which would include about 90% of the integrated Patterson density for self-vectors (Blow, 1985). It is

also critical to omit very low resolution terms since they are determined primarily by the shape of the solvent. It has been suggested to omit all terms with Bragg spacing greater than 7 or 8 Å (Blow, 1985). The high resolution cut on the data depends on the differences between the known (or model) and unknown structure. It is usually set to 2 or 3 times of the expected rms deviation between the search model and the unknown structure.

After a coarse rotational search, the produced peaks are then subject to the Patterson-correlation refinement procedure or PC-refinement (Brunger, 1990) which is implemented in Xplor. The PC-refinement procedure carries out refinements of a large number of orientations, increases the radius of convergence and clusters peaks of equivalent orientations. This is performed by minimizing a target pseudoenergy function defined as $E_{total}(\Omega,T) = (1 - PC(\Omega,T)$

where $PC(\Omega,T) = \frac{<|E_{obs}|^2 |E_m(\Omega,T)|^2 - <|E_{obs}|^2><|E_m(\Omega,T)|^2>>}{\sqrt{<|E_{obs}|^4 - <|E_{obs}|^2>^2><|E_m(\Omega,T)|^4 - <|E_m(\Omega,T)|^2>^2>}}$

E_{obs} denotes the normalized observed structure factor, and $E_m(\Omega,T)$ denotes the normalized structure factors of the search model placed in a triclinic unit cell identical in geometry to that of the target crystal. The Ω and T are the rotation and translation (if any) matrices. The target function for the PC-refinement is thus proportional to the negative correlation coefficient between the squared amplitudes of the observed structure factors and those calculated structure from the model. This refinement procedure is particularly important if the protein molecule has flexible parts,

which have different relative positions in the unknown structure, compared with the search model.

The orientation that gave the highest value for the correlation coefficient (from PC-refinement) is then chosen for the translation search. In the translation search, the known structure rotated in the appropriate orientation (according to the result of PC-refinement) is moved through the asymmetric unit. Then, the structure factors are calculated and compared with the observed structure factor by calculating an R-factor or the correlation coefficient as a function of the molecular position. The asymmetric units of the translation function for the 230 space groups (known as Cheshire groups) have been derived by Hirshfeld (1968) and are now incorporated in Xplor v. 3.851 (euclidean.lib).

Once the orientation and position of the unknown structure (with respect to the search model) are determined, the structure is subject to rigid body refinement (Sussman *et al.*, 1997) to move the model into the optimal position. The packing of the model in the target crystal unit cell is then checked to identify any bad contact. Good crystal packing would suggest the correctness of the molecular replacement solution. At this stage, the electron density map (calculated by using the observed structure factors and the phase from the model) should start to reveal unique features of the unknown structure. The phasing model can be modified accordingly and further refined.

References:

Even Jehovah,
After Moses had got the Commandments
Committed to stone
Probably thought:
I always forget the things
I really intended to say.

Christopher Morley

Abrams, C. S., and Zhao, W. (1995). SH3 domains specifically regulate kinase activity of expressed Src family proteins. J. Biol. Chem. *270*, 333-339.

Alexandropoulos, K., and Baltimore, D. (1996). Coordinate activation of c-Src by SH3- and SH2-binding sites on a novel p130Cas-related protein, Sin. Genes Dev. *10*, 1341-1355.

Alexandropoulos, K., Cheng, G., and Baltimore, D. (1995). Proline-rich sequences that bind to Src homology 3 domains with individual specificities. Proc. Natl. Acad. Sci. USA *92*, 3110-3114.

Andreotti, A. H., Bunnell, S. C., Feng, S., Berg, L. J., and Schreiber, S. L. (1997). Regulatory intramolecular association in a tyrosine kinase of the Tec family. Nature *385*, 93-7

Bacon, D., and Anderson, W. F. (1988). A fast algorithm for rendering space-filling molecule pictures. J. Mol. Graphics *6*, 219-220.

Batzer, A. G., Blaikie, P., Nelson, K., Schlessinger, J., and Margolis, B. (1995). The phosphotyrosine interaction domain of Shc binds an

LXNPXY motif on the epidermal growth factor receptor. Molecular & Cellular Biology *15*, 4403-9.

Baur, A. S., Sawai, E. T., Dazin, P., Fantl, W. J., Cheng-Mayer, C and Peterlin, B. M. (1994). HIV-1 Nef leads to inhibition or activation of T cells depending on its intracellular localization. Immunity *1*, 373-384.

Blaikie, P., Immanuel, D., Wu, J., Li, N., Yajnik, V., and Margolis, B. (1994). A region in Shc distinct from the SH2 domain can bind tyrosine-phosphorylated growth factor receptors. Journal of Biological Chemistry *269*, 32031-4.

Blow D. M. (1985). Introduction to rotation and translation functions. In P.A. Machin (Ed.), Molecular Replacement (pp. 2-7). Daresbury Laboratory.

Bolen, J. B. (1993). Nonreceptor tyrosine protein kinases. Oncogene *8*, 2025-2031.

Booker, G. W., Breeze, A. L., Downing, A. K., Panayotou, G., Gout, I., Waterfield, M. D., and Campbell, I. D. (1992). Structure of an SH2 domain of the p85a subunit of phosphatidylinositol-3-OH kinase. Nature *358*, 684-687.

Borg, J.-P., Ooi, J., Levy, E., and Margolis, B. (1996). The Phosphotyrosine Interaction Domains of X11 and FE65 Bind to

Distinct Sites on the YENPTY Motif of Amyloid Precursor Protein. Molecular and Cellular Biology *16*, 6229-6241.

Brady-Kalnay, S. M., and Tonks, N. K. (1994). Protein tyrosine phosphatases: from structure to function. Trends Cell Biol. *4*, 73-76.

Brünger, A. T. (1988). X-plor manual, Howard Hughes Medical Institute, and Dept. of Molecular Biophysics and Biochemistry, Yale University, New Haven , Connnecticut.

Brünger, A. T. (1990). Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. Acta Crystallogr. Sect. A. *46*, 46-57.

Brünger, A. T. (1992). *X-PLOR (Version 3.1) Manual*: The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Avenue, New Haven, CT 06511).

Brünger, A. T. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. Acta Cryst. *D49*, 24-36.

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

Case, R. D., Piccione, E., Wolf, G., Lechleider, R. J., Chaudhuri, M., Neel, B. G., and Shoelson, S. E. (1994). SH-PTP2 SH2 domain binding specificity is defined by direct interactions with PDGF b-receptor, EGF receptor, and IRS-1 derived phosphopeptides. J. Biol. Chem. *269*, 10467-10474.

Charbonneau, H., and Tonks, N. K. (1992). 1002 protein phosphatases? Annu. Rev. Cell Biol. *8*, 463-493.

Charest, A., Wagner, J., Jacob, S., McGlade, C. J., and Tremblay, M. L. (1996). Phosphotyrosine-independent binding of SHC to the NPLH sequence of murine protein-tyrosine phosphatase-PEST. Evidence for extended phosphotyrosine binding/phosphotyrosine interaction domain recognition specificity. Journal of Biological Chemistry *271*, 8424-9.

Cheng, G., Ye, Z. S., and Baltimore, D. (1994). Binding of Bruton's tyrosine kinase to Fyn, Lyn, or Hck through a Src homology 3 domain-mediated interaction. Proc. Natl. Acad. Sci. USA *91*, 8152-8155.

Chothia, C., and Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. ***EMBO J. 5***, 823-826.

Cicchetti, P., Mayer, B. J., Theil, G., and Baltimore, D. (1992). Identification of a protein that binds to the SH3 region of abl and is similar to Bcr and GAP-rho. Science *257*, 803-806.

Cohen, B., Yoakim, M., Piwnica-Worms, H., Roberts, T., and Schaffhausen, B. S. (1990). Tyrosine phosphorylation is a signal for the trafficking of pp85, a polypeptide associated with phosphatidylinositol kinase activity. Proc. Natl. Acad. Sci. USA *87*, 4458-4462.

Cohen, G. B., Ren, R., and Baltimore, D. (1995a). Modular binding domains in signal transduction proteins. Cell *80*, 237-48.

Cohen, S. L., Ferre-D'Amare, A. R., Burley, S. K., and Chait, B. T. (1995b). Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. Protein Science *4*, 1088-1099.

Daniel, M. D., Kirchoff, F., Czajak, S. C., Sehgal, P. K., and Derosiers, R. C (1992). Protective effects of a live attenuated SIV vaccine with a deletion in the nef gene. Science *258*, 1938-1941.

Deacon, N. J., Tsykin, A., Solomon, A., Smith, K., Ludford-Menting, M., Hooker, D. J., McPhee, D. A., Greenway, A. L., Ellett, A., Chatfield, C., Lawson, V. A., Crowe, S., Maerz, A., Sonza, S., Learmont, J., Sullivan, J. S., Cunningham, A., Dwyer, D., Dowton, D., and Mills, J. (1995). Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipents. Science *270*, 988-991.

Dechert, U., Adam, M., Harder, K. W., Clark-Lewis, I., and Jirik, F. (1994). Characterization of protein tyrosine phosphatase SH-PTP2.

Study of phosphopeptide substrates and possible regulatory role of SH2 domains. Journal of Biological Chemistry *25*, 5602-11.

Dikic, I., Batzer, A. G., Blaikie, P., Obermeier, A., Ullrich, A., Schlessinger, J., and Margolis, B. (1995). Shc binding to nerve growth factor receptor is mediated by the phosphotyrosine interaction domain. Journal of Biological Chemistry *270*, 15125-9.

Domchek, S. M., Auger, K. R., Chatterjee, S., Burke, T. R., and Shoelson, S. E. (1992). Inhibition of SH2 Domain/Phosphoprotein Association by a Nonhydrolyzable Phosphonopeptide. Biochemistry *31*, 9865-9870.

Doyle, D. A., Lee, A., Lewis, J., Kim, E., Sheng, M., and MacKinnon, R. (1996). Crystal structures of a complexed and peptide-free membrane protein-binding domain-molecular basis of peptide recognition by PDZ domains. Cell *85*, 1067-76.

Duclos, F., Boschert, U., Sirugo, G., Mandel, J. L., Hen, R., and Koenig, M. (1993). Gene in the region of the Friedreich ataxia locus encodes a putative transmembrane protein expressed in the nervous system. Proceedings of the National Academy of Sciences of the United States of America *90*, 109-13.

Eck, M. J., Atwell, S. K., Shoelson, S. E., and Harrison, S. C. (1994). Structure of the regulatory domains of the Src-family tyrosine kinase Lck. Nature *368*, 764-769.

Eck, M. J., Dhepaganon, S., Trub, T., Nolte, R. T., and Shoelson, S. E. (1996). Structure of the IRS-1 PTB domain bound to the juxtamembrane region of the insulin receptor. Cell *85*, 695-705.

Eck, M. J., Pluskey, S., Trüb, T., Harrison, S. C., and Shoelson, S. E. (1996). Spatial constraints on the recognition of phosphoproteins by the tandem SH2 domains of the phosphatase SH-PTP2. Nature *379*, 277-280.

Eck, M., Atwell, S. K., Shoelson, S. E., and Harrison, S. C. (1994). Crystal structure of the regulatory domains of the Src-family tyrosine kinase lck. Nature *368*, 764-769.

Eck, M., Shoelson, S. E., and Harrison, S. C. (1993). Recognition of a high affinity phosphotyrosyl peptide by the Src homology 2 domain of $p56^{lck}$. Nature *362*, 87-91.

Erpel, T., Superti-Furga, G., and Courtneidge, S. A. (1995). Mutational analysis of Src SH3 domain: the same residues of the ligand binding surface are important for intra- and intermolecular interactions. EMBO J. *14*, 963-975.

Fantl, W. J., Escobedo, J. A., Martin, G. A., Turck, C. W., Rosario, M., McCormick, F., and Williams, L. T. (1992). Distinct phosphotyrosines on a growth factor receptor bind to specific molecules that mediate different signalling pathways. Cell *69*, 413-423.

Felder, S., Zhou, M., Hu, P., Urena, J., Ullrich, A., Chaudhuri, M., White, M., Shoelson, S. E., and Schlessinger, J. (1993). SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. Mol. Cell. Biol. *13*, 1449-1455.

Feng, G.-S., Hui, C.-C., and Pawson, T. (1993). SH2-containing phosphotyrosine phosphatases as a target of protein-tyrosine kinases. Science *259*, 1607-1614.

Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. A. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. Science *266*, 1241-1247.

Feng, S., Chen, J. K., Yu, H., Simon, J. A., and Schreiber, S. L. (1994). Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. Science *266*, 1241-7.

Feng, S., Kasahara, C., Rickles, R. J., and Schreiber, S. L. (1995). Specific interactions outside the proline-rich core of two classes of Src homology 3 ligands. Proc. Natl. Acad. Sci. USA *92*, 12408-12415.

Fiore, F., Zambrano, N., Minopoli, G., Donini, V., Duilio, A., and Russo, T. (1995). The regions of the Fe65 protein homologous to the phosphotyrosine interaction/phosphotyrosine binding domain of Shc

bind the intracellular domain of the Alzheimer's amyloid precursor protein. Journal of Biological Chemistry *270*, 30853-6.

Freund, J., Kellner, R., Houthaeve, T., and Kalbitzer, H. R. (1994a). Stability and proteolytic domains of Nef protein from human immunodeficiency virus (HIV) type 1. Eur. J. Biochem. *221*, 811-819.

Freund, J., Kellner, R., Konvalinka, J., Wolber, V., Kräusslich, H.-G., and Kalbitzer, H. R. (1994b). A possible regulation of negative factor (Nef) activity of human immunodeficiency virus type 1 by the viral protease. FEBS Lett. *223*, 589-593.

Gorina, S., Pavletich, N. P. (1996). Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domain of 53BP2. Science *274*, 1001-1005.

Gosser, Y. Q., Zheng, J., Overduin, M., Mayer, B. J., and Cowburn, D. (1995). The solution structure of Abl SH3, and its relationship to SH2 in the SH(32) construct. Structure *3*, 1075-1086.

Goudreau, N., Cornille, F., Duchesne, M., Parker, F., Tocqué, B., Garbay, C., and Roques, B. P. (1994). NMR structure of the N-terminal SH3 domain of GRB2 and its complex with a proline rich peptide from Sos. Nature Struct. Biol. *1*, 898-907.

Grzesiek, S., Bax, A., Clore, G. M., Gronenborn, A. M., Hu, J.-S., Kaufman, J., Palmer, I., Stahl, S. J., and Wingfield, P. T. (1996). The solution structure of HIV-1 Nef reveals an unexpected fold and permits

delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. Nature Struct. Biol. *3*, 340-345.

Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J. D., and Bax, A. (1995). Four-dimensional 15N-separated NOESY of slowly tumbling perdeuterated 15N-enriched proteins. Applicatons to HIV-1 Nef. J. Am. Chem. Soc. *117*, 9594-9595.

Gustafson, T. A., He, W., Craparo, A., Schaub, C. D., and O'Neill, T. J. (1995). Phosphotyrosine-dependent interaction of SHC and insulin receptor substrate 1 with the NPEY motif of the insulin receptor via novel non-SH2 domain. Mol. Cell. Biol. *15*, 2500-2508.

Harrison, S. C. (1996). Peptide-surface association: the case of PDZ and PTB domains. Cell *86*, 341-3.

Hatada, M. H., Lu, X., Laird, E. R., Green, J., Morgenstern, J. P., Lou, M., Marr, C. S., Phillips, T. B., Ram, M. K., Theriault, K., Zoller, M. J., and Karas, J. L. (1995). Molecular basis for the interaction of the protein tyrosine kinase ZAP-70 with the T-cell receptor. Nature *377*, 32-38.

Hendrickson, W. A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron data. Science *254*, 51-58.

Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength

anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO Journal *9*, 1665-72.

Hirshfeld, F. L, (1968). Symmetry in the generation of trial structures. Acta Cryst., *A24*, 301-311.

Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. Science *268*, 1144-9.

Hope, H. (1990). Crystallography of Biological Macromolecules at Ultra-Low Temperature. *Annu. Rev. Biophys. Biophys. Chem.* ***19***, 107-126.

Hoppe, W. (1957). Die Faltmolekulmethode - eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannter Molekulstruktur. Acta Cryst., *10*, 750-751.

Huang, Y., Zhang, L., and Ho, D. D. (1995). Characterization of *nef* sequences in long-term survivors of human immunodeficiency virus type 1 infection. J. Virol. *69*, 93-100.

Huber, R. (1985). Experience with the application of Patterson search techniques. In P.A. Machin (Ed.), Molecular Replacement (pp. 58-61). Daresbury Laboratory.

Isakoff, S. J., Yu, Y. P., Su, Y. C., Blaikie, P., Yajnik, V., Rose, E., Weidner, K. M., Sachs, M., Margolis, B., and Skolnik, E. Y. (1996). Interaction between the phosphotyrosine binding domain of Shc and the insulin

receptor is required for Shc phosphorylation by insulin in vivo. Journal of Biological Chemistry *271*, 3959-62.

Isakov, N., Wange, R. L., Burgess, W. H., Watts, J. D., Aebersold, R., and Samelson, L. E. (1995). ZAP-70 binding specificity to T cell receptor tyrosine-based activation motifs: the tandem SH2 domains of ZAP-70 bind distinct tyrosine-based activation motifs with varying affinity. Journal of Experimental Medicine *181*, 375-80.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A47*, 110-119.

Kashishian, A., Kazlauskas, A., and Cooper, J. A. (1992). Phosphorylation sites in the PDGF receptor with different specificities for binding GAP and PI3 kinase in vivo. EMBO J. *11*, 1373-1381.

Kato, J. Y., Takeya, T., Grandori, C., Iba, H., Levy, J. B., and Hanafusa, H. (1986). Amino acid substitutions sufficient to convert the nontransforming p60c-src protein to a transforming protein. Mol. Cell. Biol. *6*, 4155-4160.

Kavanaugh, W. M., and Williams, L. T. (1994). An alternative to SH2 domains for binding tyrosine-phosphorylated growth factor receptors. Science *266*, 1862-1865.

Kavanaugh, W. M., Turck, C. W., and Williams, L. T. (1995). PTB domain binding to signaling proteins through a sequence motif containing phosphotyrosine. Science *268*, 1177-9.

Kawakami, T., Pennington, C. Y., and Robbins, K. C. (1986). Isolation and oncogenic potential of a novel human src-like gene. Mol. Cell. Biol. *6*, 4195-201.

Kazlauskas, A., Feng, G.-S., Pawson, T., and Valius, M. (1993). The 64-kDa protein that associates with the platelet-derived growth factor receptor β subunit via Tyr-1009 is the SH2-containing phosphotyrosine phosphatase Syp. Proc. Natl. Acad. Sci. (USA) *90*, 6939-6942.

Kazlauskas, A., Kashishian, A., Cooper, J. A., and Valius, M. (1992). GTPase activating protein and phosphatidylinositol 3-kinase bind to a distinct region of the platelet-derived growth factor receptor b subunit. Mol. Cell. Biol. *12*, 2534-2544.

Kestler, H. W., III, Ringler, D. J., Mori, K., Panicali, D. L., Sehgal, P. K., Daniel, M. D., and Desrosiers, R. C. (1991). Importance of the *nef* gene for maintenance of high virus loads and for development of AIDS. Cell *65*, 651-662.

Kirchhoff, F., Greenough, T. C., Brettler, D. B., Sullivan, J. L., and Desrosiers, R. C. (1995). Absence of intact nef sequences in a long-term survivor with nonprogressive HIV-1 infection. New Engl. J. Med. *332*, 259-260.

Knudsen, B. S., Feller, S. M., and Hanafusa, H. (1994). Four proline-rich sequences of the guanine-nucleotide exchange factor C3G bind with unique specificity to the first Src homology domain of Crk. J. Biol. Chem. *269*, 32781-32787.

Knudsen, B., Zheng, J., Feller, S. M., Mayer, J. P., Burrell, S. K., Cowburn, D., and Hanafusa, H. (1995). Affinity and specificity requirements for the first Src homology 3 domain of the Crk protein. EMBO J. *14*, 2191-2198.

Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991). SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. Science *252*, 668-674.

Kraulis, P. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946-950.

Kuriyan, J., and Cowburn, D. (1993). Structures of SH2 and SH3 domains. Curr. Op. Struct. Biol. *3*, 828-837.

Ladbury, J. E., Lemmon, M. A., Zhou, M., Green, J., Botfield, M. C., and Schlessinger, J. (1995). Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. Proc. Natl. Acad. Sci. USA *92*, 3199-3203.

Laminet, A. A., Apell, G., Conroy, L., and Kavanaugh, W. M. (1996). Affinity, specificity, and kinetics of the interaction of the SHC phosphotyrosine binding domain with asparagine-X-X-phosphotyrosine motifs of growth factor receptors. Journal of Biological Chemistry *271*, 264-9.

Lattman, E. E. (1985). Use of the rotation and translation functions. Methods Enzymol., *115*, 55-77.

Lechleider, R. J., Sugimoto, S., Bennett, A. M., Kashishian, A. S., Cooper, J. A., Shoelson, S. E., Walsh, C. T., and Neel, B. G. (1993). Activation of the SH2-containing phosphotyrosine phosphatase SH-PTP2 by its binding site, phosphotyrosine 1009 on the PDGF receptor. J. Biol. Chem *268*, 21478-21481

Lee, C.-H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E., and Kuriyan, J. (1994). Crystal structures of peptide complexes of the amino-terminal SH2 domain of the Syp tyrosine phosphatase. Structure *2*, 423-438.

Lee, C.-H., Leung, B., Lemmon, M. A., Zheng, J., Cowburn, D., Kuriyan, J., and Saksela, K. (1995). A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein. EMBO J. *14*, 5006-5015.

Lee, C.-H., Saksela, K., Mirza, U. A., Chait, B. T., and Kuriyan, J. (1996). Crystal structure of the conserved core of HIV-1 Nef complexed with a Src family SH3 domain. Cell *85*, 931-942.

Lemmon, M. A., and Ladbury, J. E. (1994). Thermodynamic studies of tyrosyl-phosphopeptide binding to the SH2 domain of $p56^{lck}$. Biochemistry *33*, 5070-5076.

Lemmon, M. A., Ferguson, K. M., and Schlessinger, J. (1996). PH domains: diverse sequences with a common fold recruit signaling molecules to the cell surface. Cell *85*, 621-4.

Li, W., Nishimura, R., Kashishian, A., Batzer, A., Kim, W. J. H., Copper, J. A., and, and Schlessinger, J. (1994). A new function for a phosphotyrosine phosphatase: linking Grb2-Sos to a receptor tyrosine kinase. Mol. Cell. Biol. *14*, 509-517.

Lim, W. A., and Richards, F. M. (1994). Critical residues in an SH3 domain from Sem-5 suggest a mechanism for proline-rich peptide recognition. Nature Struct. Biol. *1*, 221-225.

Lim, W. A., Richards, F. M., and Fox, R. O. (1994). The structural determinants of peptide binding orientation and of sequence specificity in SH3 domains. Nature *372*, 375-379.

Lu, X., et al., and Peterlin, B.M. (1996). Cdc42 and Rac1 are implicated in the activation of the Nef-associated kinase and replication of HIV. Curr. Biol. *6*, 1677-1684.

Macias, M. J., Musacchio, A., Ponstingl, H., Nilges, M., Saraste, M., and Oschkinat, H. (1994). Structure of the pleckstrin homology domain from beta-spectrin. Nature *369*, 675-7.

Maignan, S., Guilloteau, J.-P., Fromage, N., Arnoux, B., Becquart, J., and Ducruix, A. (1995). Crystal structure of the mammalian GRB2 adaptor. Science *268*, 291-293.

Mandiyan, V., O'Brien, R., Zhou, M., Margolis, B., Lemmon, M. A., Sturtevant, J. M., and Schlessinger, J. (1996). Thermodynamic studies of SHC phosphotyrosine interaction domain recognition of the NPXpY motif. Journal of Biological Chemistry *271*, 4770-5.

Mangasarian, A., Foti, M., Aiken, C., Chin, D., Carpentier, J.-L and Trono, D. (1997). The HIV-1 Nef protein acts as a connector with sorting pathways in the Golgi and at the plasma membrane. Immunity, in press.

Marengere, L. E. M., and Pawson, T. (1992). Identification of residues in GTPase activating protein Src homology 2 domains that control binding to tyrosine phosphorylated growth factor receptors and p62. J. Biol. Chem. *267*, 22779-22786.

Marengere, L. E. M., Songyang, Z., Gish, G. D., Schaller, M. D., Parsons, J. T., Stern, M. J., Cantley, L. C., and Pawson, T. (1994). SH2 domain specificity and activity modified by a single residue. Nature *369*, 502-505.

Margolis, B., Silvennoinen, O., Comoglio, F., Roonprapunt, C., Skolnik, E., Ullrich, A., and, and Schlessinger, J. (1992). High-efficiency expression/cloning of epidermal growth factor-receptor-binding proteins with Src homology 2 domain. Proc. Natl. Acad. Sci. USA *89*, 8894-8898.

Matsuda, M., Mayer, B. J., Fukui, Y., and Hanafusa, H. (1990). Binding of Transforming Protein, P47gag-crk, to a Broad Range of Phosphotyrosine-Containing Proteins. Science *248*, 1537-1539.

Mayer, B. J., and Baltimore, D. (1993). Signalling through SH2 and SH3 domains. Trends in Cell Biol. *3*, 8-13.

Mayer, B. J., Jackson, P. K., and Baltimore, D. (1991). The noncatalytic src homology region 2 segment of abl tyrosine kinase binds to tyrosine-phosphorylated cellular proteins with high affinity. Proc Natl Acad Sci Usa *88*, 627-631.

Mayer, B. J., Jackson, P. K., Van Etten, R. A., and Baltimore, D. (1992). Point mutations in the abl SH2 domain coordinately impair phosphotyrosine binding in vitro and transforming ability in vivo. Mol. Cell. Biol. *12*, 609-618.

Mayer, B. J., Jackson, P. K., Van Etten, R. A., and Baltimore, D. (1992). Point mutations in the abl SH2 domain coordinately impair phosphotyrosine binding in vitro and transforming ability in vivo. Mol. Cell. Biol. *12*, 609-618.

Milia, E., Di Somma, M. M., Baldoni, F., Chiari, R., Lanfrancone, L., Pelicci, P. G., Telford, J. L., and Baldari, C. T. (1996). The aminoterminal phosphotyrosine binding domain of Shc associates with ZAP-70 and mediates TCR dependent gene activation. Oncogene *13*, 767-75.

Moran, M. F., Koch, C. A., Anderson, D., Ellis, C., England, L., Martin, G. S., and Pawson, T. (1990). Src homology region 2 domains direct protein-protein interactions in signal transduction. Proc Natl Acad Sci Usa *87*, 8622-8626.

Muller, A. J., Pendergast, A.-M., Havlik, M. H., Puil, L., Pawson, T., and Witte, O. N. (1992). A limited set of SH2 domains binds BCR through a high-affinity phosphotyrosine-independent interaction. Mol. Cell. Biol. *12*, 5087-5093.

Muller, U., Cristina, N., Li, Z. W., Wolfer, D. P., Lipp, H. P., Rulicke, T., Brandner, S., Aguzzi, A., and Weissmann, C. (1994). Behavioral and anatomical deficits in mice homozygous for a modified beta-amyloid precursor protein gene. Cell *79*, 755-65.

Murphy, S. M., Bergman, M., and Morgan, D. O. (1993). Suppression of c-Src activity by C-terminal Src kinase involves the c-Src SH2 and SH3 domains: analysis with Saccharomyces cerevisiae. Mol. Cell. Biol. *13*, 5290-5300.

Musacchio, A., Noble, M., Pauptit, R., Wierenga, R., and Saraste, M. (1992). Crystal structure of a Src-homology 3 (SH3) domain. Nature *359*, 851-855.

Musacchio, A., Saraste, M., and Wilmanns, M. (1994). High-resolution crystal structures of tyrosine kinase SH3 domain complexed with proline-rich peptides. Struct. Biol. *1*, 546-551.

Musacchio, A., Wilmanns, M., and Saraste, M. (1994). Structure and function of the SH3 domain. Progr. Biophys. Mol. Biol. *61*, 283-297.

Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins: Struct. Funct. and Genetics *11*, 281-296.

Noble, M. E. M., Musacchio, A., Saraste, M., Courtneidge, S. A., and Wierenga, R. K. (1993). Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. EMBO J. *12*, 2617-2624.

Noble, M. E., Musacchio, A., Saraste, M., Courtneidge, S. A., and Wierenga, R. K. (1993). Crystal structure of the SH3 domain in human Fyn; comparison of the three-dimensional structures of SH3 domains in tyrosine kinases and spectrin. EMBO J. *12*, 2617-2624.

Nolte, R. T., Eck, M. J., Schlessinger, J., Shoelson, S. E., and Harrison, S. C. (1996). Crystal structure of the PI 3-kinase p85 amino-terminal

SH2 domain and its phosphopeptide complexes. Nature Struct. Biol. *3*, 364-373.

Nunn, M.F. and Marsh, J.W. (1996). Human immunodeficiency virus-1 Nef associates with a member of the p21-activated kinase (PAK) family. J. Virol. *70*, 6157-6161.

O'Bryan, J. P., Songyang, Z., Cantley, L., Der, C. J., and Pawson, T. (1996). A mammalian adaptor protein with conserved Src homology 2 and phosphotyrosine-binding domains is related to Shc and is specifically expressed in the brain. Proceedings of the National Academy of Sciences of the United States of America *93*, 2729-34.

O'Shannessy, D. J., Brigham, B. M., Soneson, K. K., Hensley, P., and Brooks, I. (1993). Determination of rate and equilibrium binding constants for macromolecular interactions using surface plasmon resonance: use of nonlinear least squares analysis methods. Anal. Biochem. *212*, 457-468.

Okada, M., Howell, B. W., Broome, M. A., and Cooper, J. A. (1993). Deletion of the SH3 domain of Src interferes with regulation by the phosphorylated carboxyl-terminal tyrosine. J. Biol. Chem. *268*, 18070-5.

Overduin, M., Rios, C. B., Mayer, B. J., Baltimore, D., and Cowburn, D. (1992). Three-dimensional solution structure of the src homology 2 domain of c-abl. Cell *70*, 697-704.

Panayotou, G., Gish, G., End, P., Truong, O., Gout, I., Dhand, R., Fry, M. J., Hiles, I., Pawson, T., and Waterfield, M. D. (1993). Interactions between SH2 domains and tyrosine phosphorylated platelet-derived growth factor beta-receptor sequences: analysis of kinetic parameters by a novel biosensor-based approach. Mol. Cell. Biol. *13*, 3567-3576.

Pascal, S. M., Singer, A. U., Gish, G., Yamazaki, T., Shoelson, S. E., Pawson, T., Kay, L. E., and Forman-Kay, J. D. (1994). Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-g1 complexed with a high affinity binding peptide. Cell *77*, 461-472.

Pascal, S. M., Yamazaki, T., Singer, A. U., Kay, L. E., and Forman-Kay, J. D. (1995). Structural and dynamic characterization of the phosphotyrosine binding region of a Src homology 2 domain-phosphopeptide complex by NMR relaxation, proton exchange and chemical shift approaches. Biochemistry *34*, 11353-11362.

Pawson, T. (1992). Tyrosine kinases and their interactions with signalling molecules. Curr. Opin. Genet. Dev. *2*, 4-12.

Pawson, T. (1995). Protein modules and signalling networks. Nature *373*, 573-580.

Pawson, T., and Schlessinger, J. (1993). SH2 and SH3 domains. Curr. Biol. *3*, 434-442.

Payne, G., Shoelson, S. E., Gish, G. D., Pawson, T., and Walsh, C. T. (1993). Binding kinetics of Lck and Src SH2 domain/phosphopeptide interactions by competition assay and surface plasmon resonance. Proc. Natl. Acad. Sci. (USA) *90*, 4902-4906.

Pelicci, G., Dente, L., De Giuseppe, A., Verducci-Galletti, B., Giuli, S., Mele, S., Vetriani, C., Giorgio, M., Pandolfi, P. P., Cesareni, G., and Pelicci, P. G. (1996). A family of Shc related proteins with conserved PTB, CH1 and SH2 regions. Oncogene *13*, 633-41.

Plutzky, J., Neel, B. G., and Rosenberg, R. D. (1992). Isolation of a src homology 2-containing tyrosine phosphatase. Proc. Natl. Acad. Sci. *89*, 1123-1127.

Potts, W. M., Reynolds, A. B., Lansing, T. J., and Parsons, J. T. (1988). Activation of pp60c-src transforming potential by mutations altering the structure of an amino terminal domain containing residues 90-95. Oncogene Res. *3*, 343-355.

Pratt, J. C., Weiss, M., Sieff, C. A., Shoelson, S. E., Burakoff, S. J., and Ravichandran, K. S. (1996). Evidence for a physical association between the Shc-PTB domain and the beta c chain of the granulocyte-macrophage colony-stimulating factor receptor. Journal of Biological Chemistry *271*, 12137-40.

Rahuel, J., Gay, B., Erdmann, D., Strauss, A., Garcia-Echeverria, C., Furet, P., Caravatti, G., Fretz, H., Schoepfer, J., and Grütter, M. (1996).

Structural basis for specificity of GRB2-SH2 revealed by a novel ligand binding mode. Nat. Struct. Biol. *3*, 586-589.

Rao, N. S., Jih, J.-H. and Hartsuck, J. A. (1980). Rotation funciton space groups. Acta. Cryst., *A36*, 878-884.

Rossmann, M.G. and Blow, D.M. (1962). The detection of sub-units within the cyrstallographic asymmetric unit. Acta Cryst., *15*, 24-31.

Ravichandran, K. S., Igras, V., Shoelson, S. E., Fesik, S. W., and Burakoff, S. J. (1996). Evidence for a role for the phosphotyrosine-binding domain of Shc in interleukin 2 signaling. Proceedings of the National Academy of Sciences of the United States of America *93*, 5275-80.

Ren, R., Mayer, B. J., Cicchetti, P., and Baltimore, D. (1993). Identification of a ten-amino acid proline-rich SH3 binding site. Science *259*, 1157-1161.

Rickles, R., Botfield, M. C., Weng, Z., Taylor, J., Green, O. M., Brugge, J., and Zoller, M. J. (1994). Identification of Src, Fyn, Lyn, PI3K, and Abl SH3domain ligands using phage display libraries. EMBO J. *13*, 5598-5604.

Sadowski, I., Stone, J. C., and Pawson, T. (1986). A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of fujinami sarcoma virus p130$^{gag\text{-}fps}$. Mol. Cell. Biol. *6*, 4396-4408.

Saksela, K., Cheng, G., and Baltimore, D. (1995). Proline-rich (PxxP) motifs in HIV-1 Nef bind to SH3 domains of a subset of Src kinases and are required for the enhanced growth of Nef+ viruses but not for down-regulation of CD4. EMBO J. *14*, 484-491.

Sawai, E. T., Khan, I. H., Montbriand, P. M., Peterlin, B. M., Cheng-Mayer, C. and Luciw, P.A. (1996). Activation of PAK by HIV and SIV Nef: importance for AIDS in rhesus macaques. Curr. Biol. *6*, 1519-1527.

Sawai, E. T., Baur, A. S., Peterlin, B. M., Levy, J. A., and Cheng-Mayer, C. (1995). A conserved domain and membrane targeting of Nef from HIV and SIV are required for association with a cellular serine kinase activity. J. Biol. Chem. *270*, 15307-15314.

Sawai, E. T., Baur, A., Struble, H., Peterlin, B. M., Levy, J. A., and Cheng-Mayer, C. (1994). Human immunodeficiency virus type I Nef associates with a cellular serine kinase in T lymphocytes. Proc. Natl. Acad. Sci. *91*, 1539-1543.

Sawka-Verhelle, D., Tartare-Deckert, S., White, M. F., and Van Obberghen, E. (1996). Insulin receptor substrate-2 binds to the insulin receptor through its phosphotyrosine-binding domain and through a newly identified domain comprising amino acids 591-786. Journal of Biological Chemistry *271*, 5980-3.

Schlessinger, J. (1994). SH2/SH3 signaling proteins. Curr. Opinion Gen. Devel. *4*, 25-30.

Sheldrick, G. (1991). Heavy Atom Location using SHELXS-90. Daresbury CCP4 Meeting (Patterson interpretation and the use of macromolecular delta-F data).

Shen, S.-H., Bastein, L., Posner, B. I., and Chrétien, P. (1991). A protein-tyrosine phosphatase with sequence similarity to the SH2 domain of the protein tyrosine kinases. Nature *352*, 736-739.

Shugars, D. C., Smith, M. S., Glueck, D. H., Nantermet, P. V., Seillier-Moiseiwitsch, F., and Swanstrom, R. (1993). Analysis of human immunodeficiency virus type 1 nef gene sequences present in vivo. J. Virol. *67*, 4639-4650.

Skolnik, E. Y., Margolis, B., Mohammadi, M., Lowenstein, E., Fischer, R., Drepps, A., Ullrich, A., and and Schlessinger, J. (1991). Cloning of PI-3 kinase-associated p85 utilizing a novel method for expression/cloning of target proteins for receptor tyrosine kinases. Cell *65*, 83-90.

Skowronski, J., Parks, D. and Mariani, R. (1993). Altered T cell activation and development in transgenic mice expressing the HIV-1 *nef* gene. EMBO J. *12*, 707-713.

Songyang, Z., Gish, G., Mbamulu, G., Pawson, T., and Cantley, L. C. (1995). A single point mutation switches the specificity of group III

src homology (SH) 2 domains to that of group I SH2 domains. J. Biol. Chem. *270*, 26029-26032.

Songyang, Z., Shoelson, S. E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W. G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R. J., Neel, B. G., Birge, R. B., Fajardo, J. E., Chou, M. M., Hanafusa, H., Schaffhausen, B., and Cantley, L. C. (1993). SH2 domains recognize specific phosphopeptide sequences. Cell *72*, 767-778.

Songyang, Z., Shoelson, S. E., McGlade, J., Olivier, P., Pawson, T., Bustelo, X. R., Barbacid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R. A., and Cantley, L. C. (1994). Specific motifs recognized by the SH2 domains of Csk, 3BP2, fps/fes, GRB-2, HCP, SHC, Syk and Vav. Mol. Cell. Biol. *14*, 2777-2785.

Sudol, M., Chen, H. I., Bougeret, C., and Bork, P. (1995). Characterization of a novel protein-binding modular - the WW domain. FEBS Lett. *369*, 67-71.

Sugimoto, S., Lechleider, R. J., Shoelson, S. E., Neel, B. G., and Walsh, C. T. (1994). Expression, purification and characterization of SH2-containing protein tyrosine phosphatase, SH-PTP2. J. Biol. Chem. *268*, 22771-22776.

Sujimoto, S., Wandless, T. J., Shoelson, S. E., Neel, B. G., and Walsh, C. T. (1994). Activation of the SH2-containing protein tyrosine phosphatase SH-PTP2 by phosphotyrosine containing peptides derived from insulin receptor substrate-1. submitted.

Sun, X. J., Crimmins, D. L., Myers, M. G., Jr., Miralpeix, M., and White, M. F. (1993). Pleiotropic insulin signals are engaged by multisite phosphorylation of IRS-1. Mol. Cell Biol. *13*, 7418-7428.

Superti Furga, G., Fumagalli, S., Koegl, M., Courtneidge, S. A., and Draetta, G. (1993). Csk inhibition of c-Src activity requires both the SH2 and SH3 domains of Src. EMBO J. *12*, 2625-2634.

Superti-Furga, G., and Courtneidge, S. A. (1995). Structure-function relationships in Src family and related protein kinases. Bioessays *17*, 321-330.

Sussman, J. L., Holbrook, S. R., Church, G. M., and Kim, S. H. (1977). A structure-factor least-squares refinement procedure for macromolecular structures using constrained and restrained parameters. Acta Cryst., *A33*, 800-804.

Talmadge, D. A., Freund, R., Young, A. T., Dahl, J., Dawe, C. J., and Benjamin, T. L. (1989). Phosphorylation of middle T by $pp60^{c\text{-}src}$: a switch for binding of phosphatidylinositol 3-kinase and optimal tumorigenesis. Cell *59*, 55-65.

Terasawa, H., Kohda, D., Hatanaka, H., Tsuchiya, S., Ogira, K., Nagata, K., Ishii, S., Mandiyan, V., Ullrich, A., Schlessinger, J., and Inagaki, F. (1994). Structure of the N-terminal SH3 domain of GRB2 complexed with a peptide from the guanine nucleotide releasing factor Sos. Nature Struct. Biol. *1*, 891-897.

Trono, D. (1995). HIV accessory proteins: leading roles for the supporting cast. Cell *82*, 189-192.

Trub, T., Choi, W. E., Wolf, G., Ottinger, E., Chen, Y., Weiss, M., and Shoelson, S. E. (1995). Specificity of the PTB domain of Shc for beta turn-forming pentapeptide motifs amino-terminal to phosphotyrosine. Journal of Biological Chemistry *270*, 18205-8.

van der Geer, P., Wiley, S., Gish, G. D., Lai, V. K., Stephens, R., White, M. F., Kaplan, D., and Pawson, T. (1996). Identification of residues that control specific binding of the Shc phosphotyrosine-binding domain to phosphotyrosine sites. Proceedings of the National Academy of Sciences of the United States of America *93*, 963-8.

Waksman, G., Kominos, D., Robertson, S. R., Pant, N., Baltimore, D., Birge, R. B., Cowburn, D., Hanafusa, H., Mayer, B. J., Overduin, M., Resh, M. D., Rios, C. B., Silverman, L., and Kuriyan, J. (1992). Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. Nature *358*, 646-653.

Waksman, G., Krishna, T. S. R., Sweet, R. M., Williams, C. H., Jr., and Kuriyan, J. (1993). Crystal structure of *Escherichia coli* thioredoxin reductase refined at 2 Å resolution. Requirement for a large conformational change during catalysis. J. Molec. Biol. *submitted*.

Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D., and Kuriyan, J. (1993). Binding of a high affinity phophotyrosyl peptide to the src SH2 domain: crystal structures of the complexed and peptide-free forms. Cell *72*, 779-790.

Wange, R. L., Malek, S. N., Desiderio, S., and Samelson, L. E. (1993). Tandem SH2 domains of ZAP-70 bind to T cell antigen receptor zeta and CD3 epsilon from activated Jurkat T cells. Journal of Biological Chemistry *268*, 19797-801.

Weis, W. I., Brünger, A. T., Skehel, J. J., and Wiley, D. C. (1990). Refinement of the influenza virus hemagglutinin by simulated annealing. J. Molec. Biol. *212*, 737-761.

Weis, W. I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W. A. (1991). Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. Science *254*, 1608-1615.

Weiss, A. (1993). T Cell Antigen Receptor Signal Transduction: A Tale of Tails and Cytoplasmic Protein-Tyrosine Kinases. Cell *73*, 209-212.

Weng, Z., Thomas, S., Rickles, R., Taylor, J., Brauer, A., Seidel-Dugan, C., Michael, W., Dreyfuss, G., and Brugge, J. (1994). Identification of Src, Fyn, and Lyn SH3-binding proteins: implications for a function of SH3 domains. Mol. Cell. Biol. *14*, 4509-4921.

White, M. F. (1996). The IRS-signalling system in insulin and cytokine action. Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences *351*, 181-9.

Williams, K. P., and Shoelson, S. E. (1993). A photaffinity scan maps regions of the p85 SH2 domain involved in phosphoprotein binding. J. Biol. Chem. *in press*.

Wilson, A. J. C. (1949). The probability distribution of X-ray intensities. Acta Cryst. *2*, 318-321.

Wiseman, T., Williston, S., Brandts, J. F., and L.-N., L. (1989). Rapid measurement of binding constants and heats of binding using new titration calorimeter. Anal. Biochem. *179*, 131-137.

Wittekind, M., Mapelli, C., Farmer, B. n., Suen, K. L., Goldfarb, V., Tsao, J., Lavoie, T., Barbacid, M., Meyers, C. A., and Mueller, L. (1994). Orientation of peptide fragments from Sos proteins bound to the N-terminal SH3 domain of Grb2 determined by NMR spectroscopy. Biochemistry *33*, 13531-9.

Wolber, V., Rensland, H., Brandmeier, B., Sagemann, M., Hoffmann, R., Kalbitzer, H. R., and Wittinghofer, A. (1992). Expression, purification and biochemical characterisation of the human immunodeficiency virus 1 nef gene product. Eur. J. Biochem. *205*, 1115-21.

Wolf, G., Trub, T., Ottinger, E., Groninga, L., Lynch, A., White, M. F., Miyazaki, M., Lee, J., and Shoelson, S. E. (1995). PTB domains of IRS-1

and Shc have distinct but overlapping binding specificities. Journal of Biological Chemistry *270*, 27407-10.

Wolfgang Vogel, Reiner Lammers, Jiaoti Huang, and Ullrich, A. (1993). Activation of a phosphotyrosine phosphatase by tyrosine phosphorylation. Science *259*, 1611-1614.

Wu, X., Knudsen, B., Feller, S. M., Zheng, J., Sali, A., Cowburn, D., Hanafusa, H., and Kuriyan, J. (1995). Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. Structure *3*, 215-226.

Yi, T., Cleveland, J. L., and Ihle, J. N. (1992). Protein tyrosine phosphatase containing SH2 domains: characterization, preferential expression in hematopoietic cells, and localization to human chromosome 12p12-p13. Mol. Cell. Biol. *12*, 836-846.

Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W., and Schreiber, S. L. (1994). Structural basis for binding of proline-rich peptides to SH3 domains. Cell *76*, 933-945.

Yu, H., Rosen, M. K., Shin, T. B., Seidel-Duggan, C., Brugge, J. S., and Schreiber, S. L. (1992). Solution structure of the SH3 domain of Src and Identification of Its Ligand-Binding Site. Science *258*, 1665-1668.

Zhang, K. Y. J., and Main, P. (1990). The use of Sayre's equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. Acta Crystallogr. *A46*, 377-381.

Zhou, M. M., Harlan, J. E., Wade, W. S., Crosby, S., Ravichandran, K. S., Burakoff, S. J., and Fesik, S. W. (1995). Binding affinities of tyrosine-phosphorylated peptides to the COOH-terminal SH2 and NH2-terminal phosphotyrosine binding domains of Shc. Journal of Biological Chemistry *270*, 31119-23.

Zhou, M. M., Meadows, R. P., Logan, T. M., Yoon, H. S., Wade, W. S., Ravichandran, K. S., Burakoff, S. J., and Fesik, S. W. (1995). Solution structure of the Shc SH2 domain complexed with a tyrosine-phosphorylated peptide from the T-cell receptor. Proc. Natl. Acad. Sci. USA *92*, 7784-7788.

Zhou, M. M., Ravichandran, K. S., Olejniczak, E. F., Petros, A. M., Meadows, R. P., Sattler, M., Harlan, J. E., Wade, W. S., Burakoff, S. J., and Fesik, S. W. (1995). Structure and ligand recognition of the phosphotyrosine binding domain of Shc. Nature *378*, 584-592.

Zhou, M.-M., Huang, B., Olejniczak, E. T., Weadows, R. P., Shuker, S. B., Miyazaki, M., Trub, T., Shoelson, S. E., and Felsik, S. W. (1996). Structural basis for IL-4 receptor phosphopeptide recognition by the IRS-1 PTB domain. Nature structural biology *3*, 388-393.

Zhou, S., Margolis, B., Chaudhuri, M., Shoelson, S. E., and Cantley, L. C. (1995). The phosphotyrosine interaction domain of SHC recognizes

tyrosine-phosphorylated NPXY motif. Journal of Biological Chemistry *270*, 14863-6.

THE LIBRARY

19010000038360

End