5-2008

The Structural and Functional Study of GIT1 Paxillin Binding Domain

Ziwei Zhang
University of Tennessee Health Science Center

The Structural and Functional Study of GIT1 Paxillin Binding Domain

Document Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Program
Molecular Sciences

Research Advisor
Jie Zheng, Ph.D.

Committee
Stephan W. Morris, M.D. David R. Nelson, Ph.D. Susan E. Senogles, Ph.D. Stephen W. White, Ph.D.

DOI
10.21007/etd.cghs.2008.0379

**THE STRUCTURAL AND FUNCTIONAL STUDY OF GIT1 PAXILLIN
BINDING DOMAIN**

A Dissertation
Presented for
The Graduate Studies Council
The University of Tennessee
Health Science Center

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy
From The University of Tennessee

By

Ziwei Zhang
May 2008

DEDICATION

This dissertation is dedicated to my parents,
Dr. Yuan Zhang and Guohua Wang,
my sister Yiqiao, and to my wife, Xin
for their never ending
love and support

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Jie Zheng for his guidance, support and knowledge during my entire graduate education. Without the help of Dr. Zheng, this project would not have been possible. I appreciate Dr. Zheng's mentoring not only in science, but also in things beyond science.

I would also thank my dissertation committee members: Dr. Stephen White, Dr. Susan Senogles, Dr. David Nelson, Dr. Stephan Morris and Dr. Clare Sample, for their guidance, suggestion and support.

I am grateful to all Zheng lab members for their valuable advice, inspiring discussion and friendship. In particular, I thank Cristina Guibao and Youming Shao for teaching me various techniques and creating a warm, friendly atmosphere that made my graduate experience enjoyable.

I deeply appreciate my wife, Xin, for her endless love and support during my entire graduate study. I would like to thank my family whose support helped me to finish a project of this magnitude.

ABSTRACT

The G protein coupled receptor (GPCR)-kinase (GRK) interacting protein 1 (GIT1) is a multidomain protein that plays an important role in cell adhesion, motility, cytoskeletal remodeling, and membrane trafficking. GIT1 mediates the localization of p21-activated kinase (PAK) and PAK-interactive exchange factor (PIX) to focal adhesions, and its activation is regulated by the interaction between its C terminal paxillin-binding domain (PBD) and the LD motifs of paxillin.

In this dissertation, we determined the solution structure of rat GIT1 PBD by nuclear magnetic resonance (NMR) spectroscopy. The PBD folds into a four-helix bundle, which is structurally similar to the focal adhesion targeting (FAT) domain and the vinculin tail (Vt) domain. The PBD is more stable than the FAT domain and there is no evidence of helix 1 swapping.

Previous studies showed that GIT1 interacts with paxillin through the LD4 motif. However, studies in this dissertation demonstrated that in addition to the LD4 motif, the GIT1 PBD can also bind to the paxillin LD2 motif; and both LD2 and LD4 motifs competitively target the same site on the PBD surface. This dissertation also probed the function of paxillin splice variants by comparing their interaction with GIT1 PBD. It seems the paxillin isoforms did not play an important role in determining the affinity to GIT1. We also revealed that paxillin S272 phosphorylation does not influence GIT1 PBD binding *in vitro*. These results are in agreement with the notion that phosphorylation of paxillin S272 plays an essential role in regulating focal adhesion turnover.

This dissertation also computationally derived the complex structures of GIT1 PBD bound with either LD2 peptide or LD4 peptide, based on the experimental binding site information. The LD2 and LD4 peptides bound to GIT1 PBD in a manner similar to the crystal structure of FAT-LD2 complex. The complex structures visualized the reason why both LD2 and LD4 can bind to the same GIT1 binding site. It also addressed the specificity problem in determining paxillin binding to GIT1 versus FAK.

Our finding reconciles the controversial observations of earlier studies and provides a clearer picture of focal adhesion regulation. The structural studies of GIT1 PBD presented in this dissertation shed more light on the understanding of GIT functions. The novel findings also allow us to propose a working model regarding FA disassembly.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADC	analog digital converter
AIR	ambiguous interaction restraint
AUC	ananlytical ultra-centrifugation
CSP	chemical shift perturbation
EBNA	EBV Nuclear Antigen 2
EBV	Epstein - Barr virus
ECM	extracellular matrix
EGFR	epidermal growth factor receptor
ERK	extracellular signal regulated kinase
FA	focal adhesion
FAK	focal adhesion kinase
FAT	focal adhesion targeting domain
FERM	protein 4.1, ezrin, radixin and moesin domain
FID	free induction decay
FT	fourier transform
GIT	GRK interacting protein
GPCR	G protein coupled receptor
GRK	GPCR kinase
HADDOCK	high ambiguity driven biomolecular docking
HSQC	heteronuclear single quantum coherence
ITC	titration calorimetry
LIBS	ligand induced binding site
MAPK	mitogen activated protein kinase
MRI	magnetic resonance imaging
NMR	nuclear magnetic resonance
PAK	p21-activated kinase
PDB	protein data bank
PDGFR	platelet-derived growth factor receptor
PI3K	phosphatidylinositol 3 kinase
PIX	p21-activated kinase interacting exchange factor
PLC	phospholipase C
PRR	proline rich region
RDC	residual dipolar coupling
RF	radio frequency
RMSD	root mean square deviation
SAT	saturation transfer
SAXS	small angle X-ray scattering
SFK	Src family kinases
SHD	Spa2-homology domain
SPR	surface plasmon resonance
TOCSY	total correlation spectroscopy
VEGFR	vascular endothelial growth factor receptor

CHAPTER 1. INTRODUCTION

1.1 Overview of the dissertation

Focal adhesion plays a central role in cell attachment and migration; it controls embryogenesis, tissue repair, inflammatory response and tumor invasion. GIT1 is one of the important components in focal adhesions. This dissertation focuses on the structural and functional study of GIT1 and discusses the focal adhesion regulations through some novel discoveries. To briefly outline the dissertation, Chapter 1 introduces the biological background of Focal Adhesion and its major components. Chapter 2 describes the techniques used in the study, including NMR based structure determination and interaction study as well as computational complex structure modeling. Chapter 3 illustrates the GIT1 PBD structure determined by solution NMR and presents some insights into the GIT1 interactions. Chapter 4 examines GIT1 binding specificity, supported by the NMR driven complex structure models. Chapter 5 concludes the findings and discusses their implications into the understanding of focal adhesion regulations.

1.2 An introduction to the focal adhesion (FA)

The cell's ability to adhere to an external surface and to respond to outside environmental cues is essential for the survival of all the life forms, ranging from unicellular organisms to complex multi-cellular organisms. The simple unicellular organisms evolved the ability of attaching to the extracellular substrate and reacting correspondingly. They can move toward the food source and away from the repellents, as well as change cell morphology in response to either prey or mating cells. Similarly, individual cells from multi-cellular metazoans are capable of attaching to more diverse extracellular matrix (ECM) and reacting with more complicated behaviors.

For metazoans, cell adhesion and migration is critical for embryogenesis, tissue repair, inflammatory response, tumor invasion and metastasis.[1] In embryogenesis, progenitor cells migrate as sheet or loosely attached cell populations to form complicated tissues and organs. In adult organisms, tissue repair depends on the migration of epithelial cells, fibroblasts or other connective tissue cells from the nearby healthy tissues. During inflammatory response, leukocytes migrate through connections between vascular endothelial cells into the insult area, where it executes phagocytosis, cytotoxic killing or organizes more effective immune responses through cytokines. Cell adhesion signaling and migration also plays an important role in tumor invasion and metastasis. Tumor cells can migrate out of their original site and survive the apoptosis signaling due to the loss of attachment.

The cell-ECM attachment site is called focal adhesion (FA). Complicated protein complexes at the FA carry out the function of cell adhesion and signaling. To date, there are more than 50 protein components identified in FA.[2,3] They are divided into three

groups by their locations: extracellular, transmembrane and cytoplasmic.[3] The extracellular proteins are found in ECMs as FA targets, e.g. collagen, fibronectin, heparin sulfate and laminin. The transmembrane proteins are mainly the integrin family members that mediate both physical and signaling connections between the extracellular and intracellular world. Besides integrin, there are some other transmembrane proteins in the FA, including LAR-PTP receptor, syndecan4 and layilin. The majority of FA components are located inside the cell membrane and they can be further divided into three groups: structural proteins, signaling proteins and adapter proteins. Structural proteins include components forming cytoskeleton and those that execute movement, for example, actin, α-actinin, ezrin, talin, tensin and vinculin. The signaling proteins range from kinases, phosphatases to small GTPase modulators, for example, FAK, Pyk2, Src, PAK, ILK, PTP1-B, SHP-2, GIT, PIX and PKL. The adaptor proteins are multi-domain proteins mediating recruitment of other components into the FA, for example, paxillin, Hic-5, p130Cas, DOCK180 and FRNK.

Cell attachment, migration and signaling are highly coordinated cellular functions. They are dynamically regulated in both temporal and spatial scales. For simplicity, the cell attachment, cell migration and FA signaling will be introduced in separate sections, although those functions are closely interlaced.

1.2.1   Cell attachment

Cells adhere to ECM with the transmembrane ECM receptor integrin. Cells migrate by securing attachments in the leading edge and extending forward by cytoskeleton remodeling. At the trailing edge, the FA degrades and releases attachments so that the cell can move forward. While this appears simple, in reality there is an array of elegantly regulated events taking place in order for the cells to attach and migrate.

The first step in cell adhesion involves the interaction of peri-cellular components with their complementary substrate.[4] The peri-cellular components are normally glycolproteins or glycosaminoglycan receptors, such as CD44 and galectins.[5,6] The interactions are characterized as weak, transient, non-specific interactions. It often involves multiple copies of polysaccharides binding cooperatively to achieve the necessary affinity. The step occurs on at subsecond time scale and the distance between the cell membrane and the substrate surface is several micrometers.[4] The interaction is transient; the cell can detach from the surface if further intergrin based tighter interactions can not be created. The transient attachment is important because it creates the necessary time allowing integrin to approach to its specific substrate, otherwise simply random encounters can only result in a small fraction of stable interactions.

The next step involves the development of stable, specific interactions between cell surface integrin and its substrates. Integrin is a family of transmembrane receptors that does not contain enzymatic activity and actin binding function. All the signaling and structural connection functions are mediated by other FA proteins that are recruited by integrin. Under regulation of kinases and phosphatases, FA can further mature into a

"focal contact" through recruitment of more integrins and cytoskeleton components. Without maturation, FA is easily disassociated and the attachment is released.

1.2.2 Cell migration

Cell migration is the result of dynamic FA assembly and disassembly that occurs in cooperation with actin cytoskeleton remodeling. It can be dissected into four steps.

1. At the beginning of migration, membrane extensions such as lamellipodia and filopodia are developed at the leading edge of the cell.[7,8] In the protrusions, small FA attachments form, anchoring the cell's front edge with the ECM components. Since integrin links the ECM components to the actin filaments (through adapter proteins), the linkage physically attaches the cell cytoskeleton with the ECM.

2. The polymerization of actin adds new actin protomers at the front tip of the cytoskeleton, pushing the plasma membrane protrusion at the lamellipodia and filopodia. The membrane protrusion will be anchored by newly formed FA attachments (focal complexes) in front of the older ones.

3. The focal complexes are weak transient cell-ECM adhesions, which can disassociate easily if not further stabilized. Under intracellular and extracellular tensions, more integrin can be recruited into the focal complexes and more actins polymerize. Thus, the FA attachments mature into a "focal contact". The focal contact is more stable and more robust, often found in stationary cells. Conversely, stable focal contacts are not observed in rapidly moving cells, e.g. leukocytes.[9]

4. At the trailing edge, stable focal contacts disassociate to allow the movement of the cell body. The dissociation is caused by focal adhesion disassembly, or focal adhesion turnover. Compared to FA assembly, little is known about FA disassembly, although it is as important as FA assembly. Part of this dissertation tries to understand the mechanism of FA disassembly.

1.2.3 Cell signaling through FA

Focal adhesions have traditionally been thought of as simple receptors that anchor the cell to the ECM or other cells. However, more and more evidence has shown that in addition to physical linkage, FA plays critical signaling roles in regulating cell proliferation, survival and motility. In particular, FA relays cellular signaling in both an"inside-out" and "outside-in" fashion. In addition, FA acts as a "mechanosensor" that integrates mechanical cues with biochemical responses. Also, FA signaling collaborates with various growth factor receptor pathways in promoting survival and proliferation.

1.2.3.1   Inside-out signaling

Many integrins often exist in an inactive state. The integrin cannot bind to its ECM ligand unless activated by signals from inside of the cell. The regulated integrin activation modulates the affinity between cells and ECM, allowing cells to be motile or stationary according to functional needs. For instance, the circulating platelet has high density surface integrins, mainly αIIbβ3 heterodimer. However, all the integrins are in an inactive state without activation of the thrombotic cascade. When the platelet is activated, e.g. for hemostasis, the activation signal can be sent from inside of the plasma membrane. The activation of αIIbβ3 integrin leads to rapid and strong adherence to fibrinogen, fibronectin and von Willebrand factors, which in turn crosslink more platelets and form thrombosis.

As transmembrane proteins, integrins have a large extracellular domain and a small intracellular tail for both the α and β subunits. More structure details are reviewed in section 1.3.1. Without activation, the integrin extracellular domain is in a bend conformation that corresponds to the inactivated state (Figure 1-1a). During inside-out activation, the binding of the intracellular FA protein talin to the cytoplasmic tail of integrin leads to the separation of α and β subunits. The separation of the intracellular domains results in a long-range transmembrane conformational change, which opens the integrin extracellular domain in a "switchblade" fashion.[10,11] The long-range allosteric change exposes a new binding interface, and the integrin is fully activated. The exposed epitope is termed ligand induced binding site (LIBS).

1.2.3.2   Outside-in signaling

The outside-in signaling is the classic function of FA, by which the cell detects the chemical environment. The engagement of integrin extracellular domain to the ECM component can deliver the information, such as the chemical nature of the ECM components and their density, into the plasma membrane and allow the cell to respond correspondingly. Thus, FA can function as chemical sensors by outside-in signaling.

The same allosteric conformational change occurs when the extracellular substrate binding domain is occupied by the ECM component. The interaction stabilizes the integrin extracellular domain, locking the intracellular tails in a separated active state (Figure 1-1b). Thus, this conformational change promotes the interactions between the integrin cytoplasmic tail and the complex FA signaling and structural proteins, e.g. talin. Actually, this allosteric change is bidirectional and reciprocal. The inside-out and outside-in signaling should be viewed as the two aspects of a single allosteric change. The binding of either ECM components or the intracellular FA components can stabilize integrin in an active form, facilitating the interaction at the other side.

During FA maturation, more integrins participate in the crosslinking and clustering, forming stable focal contacts. Upon "outside-in" activation, the integrin cytoplasmic tails recruit many signaling molecules into the FA, including focal adhesion

4
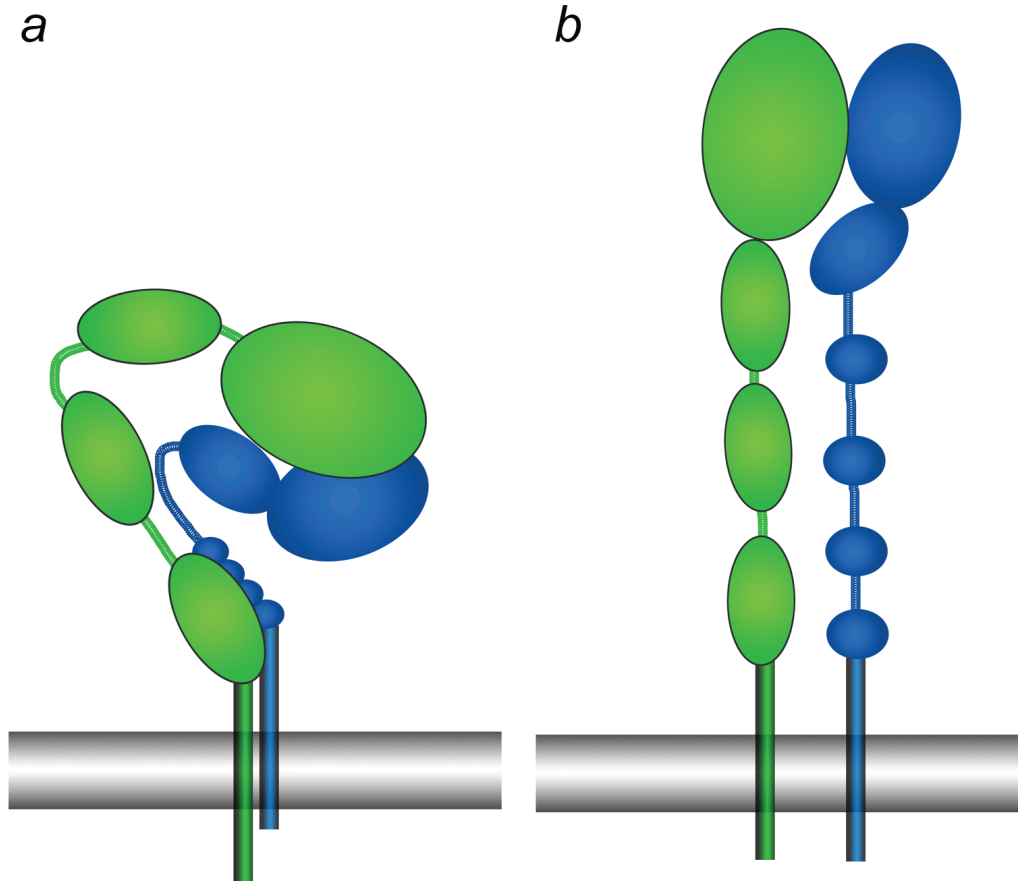
Figure 1-1   Schematic view of integrin heterodimer and activation.
Integrins are obligate heterodimers formed by a large extracellular domain and small transmembrane and intracellular domains. The α and β subunits are colored in green and blue respectively. (a) the inactive bend conformation. (b) the active extended conformation.

kinase (FAK), Src family kinases (SFK) and integrin linked kinase (ILK). To begin with, the integrin β3 tail binds talin, which recruits FAK into FA. Upon trans-autophosphorylation, FAK is activated and functions as a scaffold to further bring in SFK, activated phosphatidylinositol 3 kinase (PI3K) and extracellular signal regulated kinase (ERK)/ mitogen activated protein kinase (MAPK). The cytoskeleton remodeling and gene transcription are regulated by more complex downstream signaling. To name a few, FAK activates Rho family small GTPases, including Rac1, Cdc42, and RhoA, which differentially control the cytoskeleton formation and myosin contractility in both leading edge and trailing edge – the regulation of cell adhesion and migration. The Rac-RelA/NF-κB also influences gene transcription, promoting cell survival and proliferation. At the same time, PI3K-Akt and Ras-ERK linkages also induce transcription for survival and proliferation – the mechanism of FA signaling is required to initiate apoptosis when the cell does not adhere to the ECM.

As important as FA signaling activation, the deactivation of FA signaling helps cells detach from the adhesion and promotes cell motility by activating focal adhesion disassembly at the trailing edge. However, as the mechanism of FA activation just starts to unveil; how FA is deactivated remains even more mysterious. Rac plays an important role in FA disassembly, either directly through its effector p21-activated kinase (PAK), and/or indirectly through antagonizing Rho. One mechanism involves the recruitment of PAK by GIT into the FA. Activated PAK can catalyze the disassembly of actin stress fibers and has been found to be the central player for FA disassembly.[12-15]

1.2.3.3 FA as a mechanosensor

During FA maturation, the focal complex transforms into a more stable focal contact by crosslinking and clustering more integrins. It has been shown that local tension can reinforce the focal contacts and external forces can stimulate the growth of focal contacts.[16,17] The size of the focal contact is proportional to the contraction force applied on the FA.[18] Thus, in addition to detect the chemical composition of the ECM, FA can sense the mechanical features of the ECM too. This is of importance for cells to be able to adhere and migrate successfully. The reason is simple – only a rigid surface can provide cells enough force for movement. If the integrin happens to bind a piece of ECM component that is floating in the solution, the ECM can not provide enough tension when cell myosin contracts. Without the signal of tension, or in another word the confirmation of a rigid ECM surface, no further integrin clustering can occur. So, the focal complexes can not mature into robust focal contacts. Instead, the focal complexes will disassemble and those transient attachments are released.

The mechanism for the mechanosensor function is not very clear. One model states that for a molecular aggregate with external pulling tension, the new subunits tend to incorporate at the direction of force applied.19 The tension applied at the existing focal complexes can then promote the formation of new actin filaments, which then help crosslink more integrin, like during the formation of focal contacts.

1.2.3.4   Synergistic role of FA

In addition to function as ECM chemosensor and mechanosensor, focal adhesion also plays synergistic role to many growth factor receptors. It has been demonstrated that FA signaling is required for the potent efficient signaling of epidermal growth factor receptor (EGFR),[20] the vascular endothelial growth factor receptor (VEGFR),[21] and the platelet-derived growth factor receptor (PDGFR).[22] The association of β3 integrin with those growth factor receptors is critical for their functions. As illustrated in Figure 1-2, FA signaling shares two major downstream pathways with these growth factor signaling: the Ras/Raf-ERK/MAPK pathway and the PI3K-Akt pathway.[23] Both of them are critical for anti-apoptosis and cell proliferation. The activated ERK/MAPK promotes survival via both direct and indirect mechanisms – directly influencing the activity of the cytoplasmic apoptosis regulator bcl-2;[24] or indirectly change the level of survival proteins p53 and bax.[25] On the other hand, the activated PI3K-Akt pathway suppresses apoptosis by activating Mdm2 or inactivating Bad, Caspase9 or FKHR. It also promotes growth by mTOR or regulates the cell cycle by GSK3.

1.3   Focal adhesion components

1.3.1   Integrin

Integrin is a family of important transmembrane receptor proteins. They form obligate heterodimers of an α and β subunits. To date, there are 18 α subunits and 8 β subunits in humans.[26] The heterodimer pairing is limited to certain type of subunits; the total number of available heterodimers is limited to 24 (Figure 1-3). Depending on cell type, each cell will generally express 8 to 12 different heterodimers. Since each α or β subunit is transcribed by an independent gene, the density and type of heterodimer pairs on a cell is determined by the transcription level of both genes. The ligand specificity is determined together by both α and β subunits.

Both α and β subunits are large transmembrane proteins, each of which consists of a large extracellular ligand binding domain, a single transmembrane domain and a small intracellular signaling domain (20-50 amino acids). Currently, there are 12 integrin heterodimer structures available in the protein data bank (PDB) database (I checked the PDB online database on Jan. 19, 2008). All of them are fragments of the extracellular ligand binding domain. Among all the integrin structures, the crystal structure of integrin αVβ3 is by far the most complete one, which covers almost the entire extracellular domain (PDB ID: 1JV2, [27]). As illustrated in Figure 1-4, the entire extracellular domain is deeply bended, reflecting the inactive state. The N terminal segments of both subunits form an ovoid "head". The "tail" of αV consists of three β-sandwich domains, while the tail of β3 consists of four EGF domains. So far, no structure of the transmembrane domain and intracellular domain reported. The reason for the difficulty in obtaining a structure of the intracellular domain is because of the high flexibility of this domain.

7
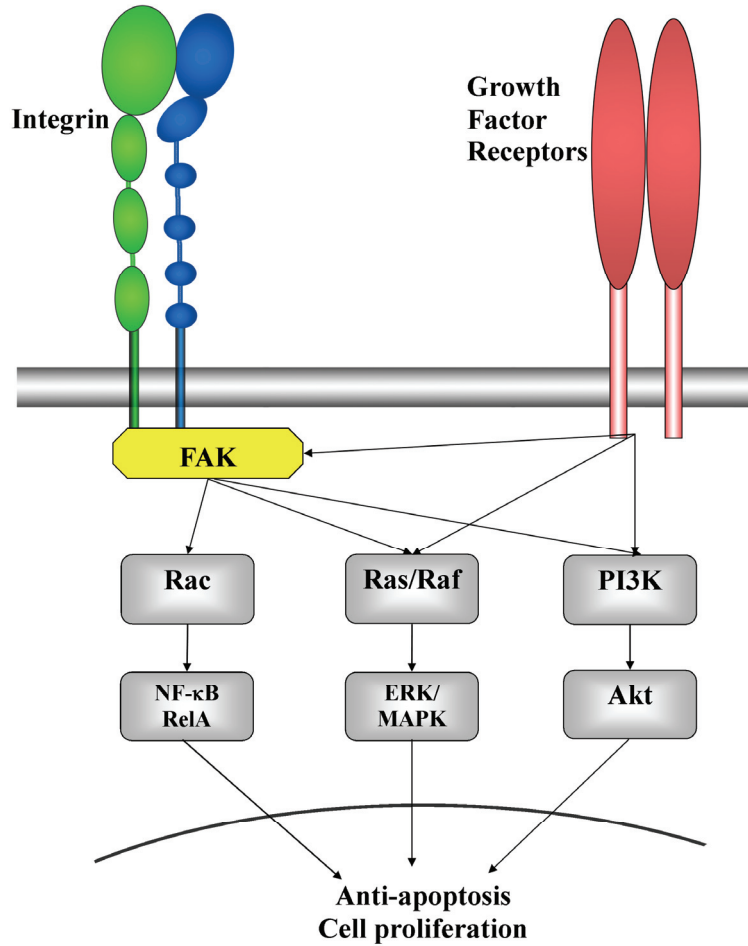
Figure 1-2   The synergistic role of FA and growth factor receptors.
*FA plays a synergistic role with growth factor receptors (receptor tyrosine kinases)
through the PI3K pathway and Ras pathway. At the same time, some receptor tyrosine
kinases can also interact with FAK and promote the Rac pathway. The activation of all
the three pathways regulates genes to promote cell survival and proliferation.*
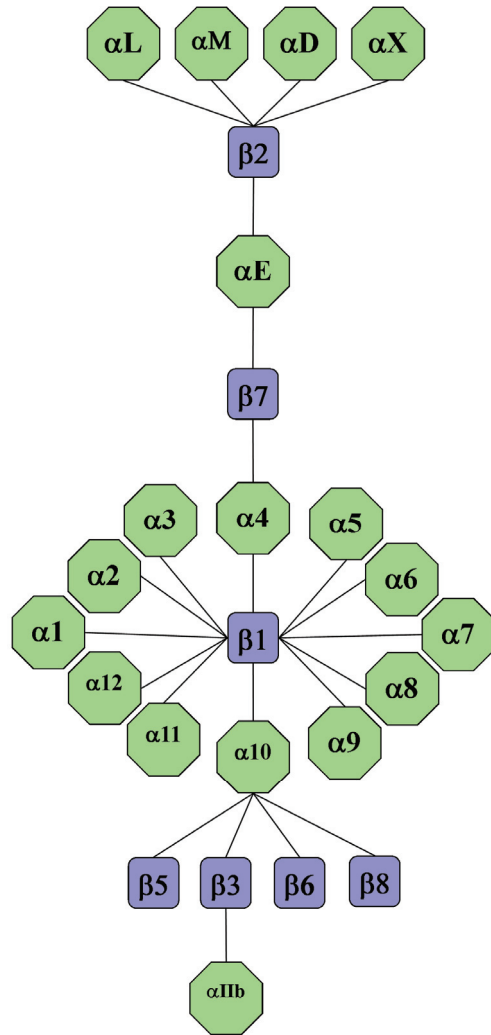
Figure 1-3   All the 24 possible integrin heterodimer pairs.
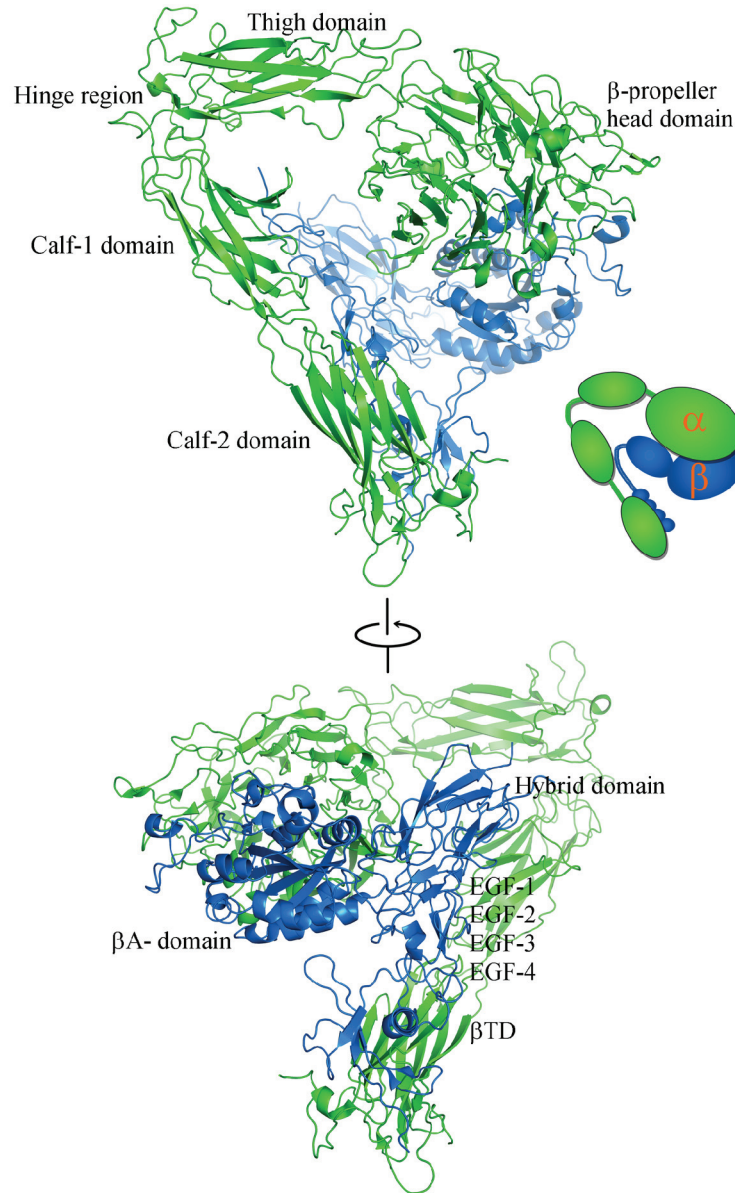
Figure 1-4   The crystal structure of the extracellular domain of integrin αVβ3.
The structure of the extracellular domain of integrin αVβ3 (1JV2, [27]) illustrated by software Pymol. The α and β subunits are colored green and blue respectively. The schematic view of domain orientation is also colored this way. The α subunit consists of the β-propeller head domain, the Thigh domain and two Calf domains. The β subunit consists of the βA-domain, Hybrid domain, 4 EGF motifs and the βTD. An oval head is formed by the β-propeller head domain from the α subunit βA-domain and the Hybrid domain from the β subunit. The overall domain structure is deeply bent, reflecting the inactive state.

The regulation of integrin activation is discussed in section 1.2.3. During activation, the bent structure opens up around the hinge region between the first and the second β-sandwich domains of αV. The conformational change exposes substrate binding sites that allow the association of integrin with its ECM target.

1.3.2   *Paxillin*

Identified over a decade ago,[28] paxillin is one of the most important adaptor proteins in the focal adhesions. Although containing no signaling activity, paxillin functions as a signaling hub by recruiting other proteins to integrate and disseminate signaling from integrins and other growth factor receptors. It also physically participates in the FA-cytoskeleton connections through the interaction with actin binding protein vinculin and actopaxin.

The N terminus of paxillin consists of five LD motifs that function as protein docking sites for many binding partners (Figure 1-5a).[29] There are four LIM domains (named after Lin-11, Isl-1, Mec-3) in the C terminus that mediate protein interactions too.[30] The LD motifs and LIM domains are modular protein interaction sites that are connected by flexible disordered sequences in between. Thus the whole protein has a very flexible beads-on-the-string architecture. The LD motifs are the consensus leucine-rich LDXLLXXL sequences found to repeat five times at the paxillin N terminal half (Figure 1-5b). With diverse binding specificity, the LD motifs mediate the interaction of paxillin to a range of target proteins (Table 1-1). It has been demonstrated that the LD motifs forms amphipathic α-helices and all the leucines are exposed to one side to create a large hydrophobic interface.[31,32] The LIM domains are tandem zinc finger motifs, which contains two antiparallel β-sheet pairs and a short α-helix at the C terminus. Cooperatively, the LIM2 and LIM3 domains mediate the localization of paxillin to FA and the serine/threonine phosphorylation of the two LIM domains regulates paxillin FA localization and cell adhesion.[33,34]

Paxillin is conserved from yeast to vertebrates; and the higher eukaryotes express three paxillin splice isoforms. While the α isoform is more ubiquitously expressed, the β and γ isoforms are more strictly expressed.[35] The human paxillin sequences are the same except for an insertion of 34 or 48 amino acids right after LD4 motif for isoform β or γ. The protein interactions are different for the splice isoforms, for example, isoform β bound to FAK but weakly to vinculin; however isoform γ bound to vinculin but weakly to FAK.[35] Since both FAK and vinculin interactions are mediated by the paxillin LD4 motif, the interaction differences might be caused by the insertions. Since the paxillin-GIT1 interaction is also mediated by LD4 motif, this interaction might also be different. Before this dissertation, no study has investigated the effect of paxillin splice isoforms on the GIT1 interaction.
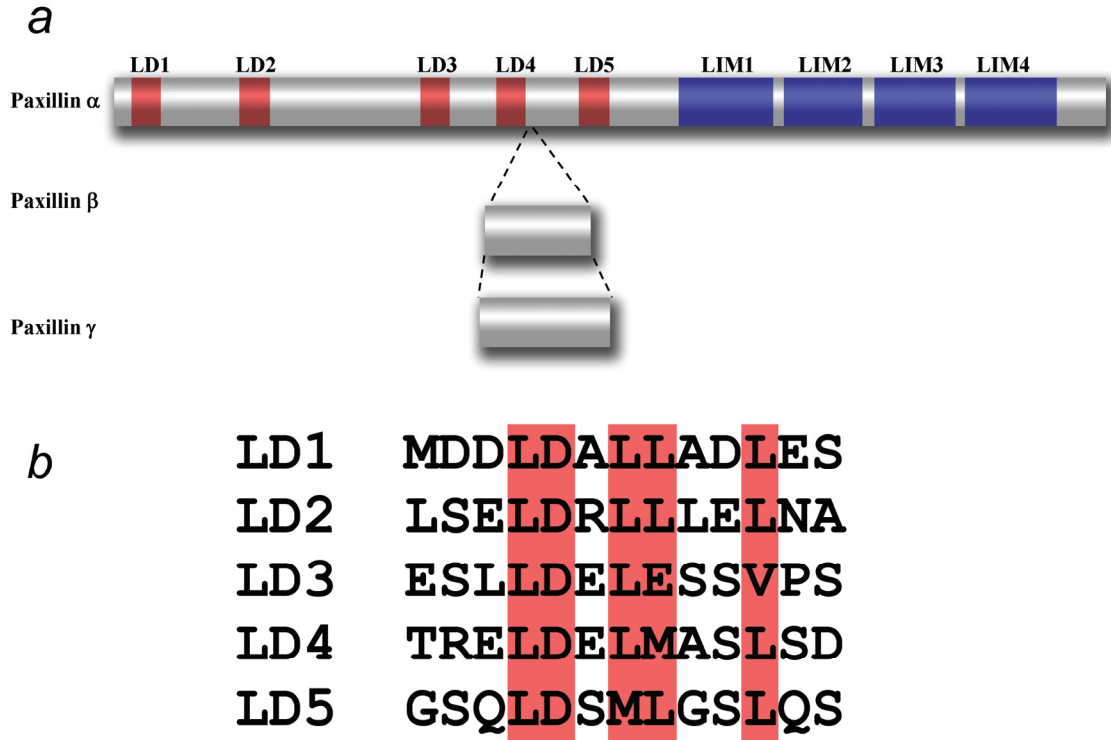
Figure 1-5   The domain structure of full length paxillin.
(a) Paxillin α consists of 5 N terminal LD motifs and 4 C terminal LIM domains. The sequence of paxillin β and γ are the same as the α splice variant except for an insertion of 34 and 48 amino acids right after the LD4 motif. (b) the sequence alignment of human paxillin LD motifs.

Table 1-1   Summary of LD motif binding partners.

LD motifs	Binding partner
LD1	Actopaxin [36]
	Intergrin-linked kinase (ILK) [37]
	Vinculin [38]
	Protein E6 from papillomavirus [39]
LD2	Vinculin [38]
	FAK [31,33]
LD3	No binding partner identified
LD4	Actopaxin [36]
	GIT1 [38]
	FAK [33]
	PAK3 [40]
	Clathrin [38]
LD5	No binding partner identified

1.3.3 FAK

The focal adhesion kinase FAK is one of the most important signal "integrators" in the FA. Since its identification in 1992, FAK has been found to play critical roles in different physiological functions, including embryo development, immunoresponse, cell adhesion and cell motility.[41] Also, an elevated FAK level has been linked with the invasion and metastasis phenotype of many malignant cells.[42]

FAK is an ubiquitously expressed protein tyrosine kinase; about 1050 amino acids long and 125 KDa in molecular weight. There are four major domains in FAK: the N terminal protein interaction FERM domain (protein 4.1, ezrin, radixin and moesin), the central tyrosine kinase domain, two proline rich regions (PRRs) and a C terminal focal adhesion targeting (FAT) domain (Figure 1-6). The FERM domain mediates the interaction with EGFR, PDGFR, ETK and ezrin,[43] which is how FAK can integrate the signal from both integrin and growth factor receptors and output in the FA control. The central kinase domain is essential for phosphorylating many critical tyrosines in activating both its targets and FAK itself. The PRRs function as SH3 binding sites for proteins like p130CAS, GRAF and ASAP1. The C terminal FAT domain promotes the localization of FAK into FA, through the interaction with integrin binding proteins paxillin and talin.[41]

Among FAK domains, the FAT is of special interests to this dissertation, because it shares many similarities with our target GIT1 PBD (section 1.3.4). The FAK FAT domain is a four-helix bundle protein that binds paxillin through both LD2 and LD4 motifs.[31] The LD2 and LD4 motifs bind to FAT at the H2/H3 face and H1/H4 face cooperatively.[44,45]

1.3.4 GIT

The G-protein coupled receptor (GPCR)-kinase (GRK) interacting protein (GIT) is an important scaffold as well as a signaling protein in the FA. It plays a critical role in

FERM domain	Kinase domain	PRR	FAT domain
EGFR	Src	p130CAS	paxillin
PDGFR	Grb7	GRAF	Talin
Ezrin	Shc	ASAP1	VEGFR-3
FAK			
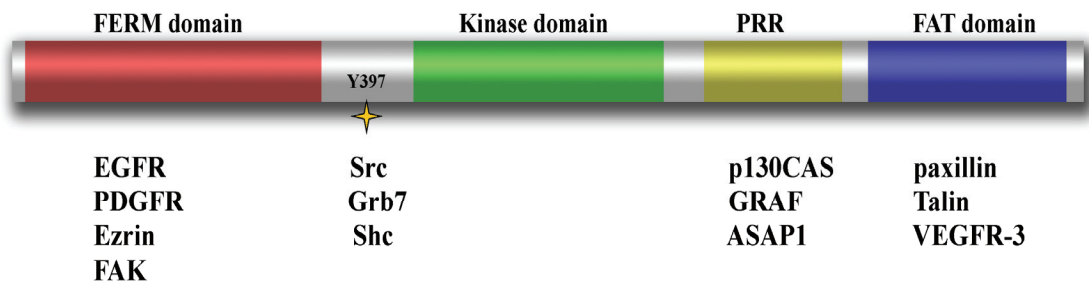
Figure 1-6   The domain structure of FAK and its interacting proteins.
The auto-phosphorylation of tyrosine 397 activates FAK and creates binding site for SH2 proteins, like Src and Grb.

controlling FA disassembly. It also controls the cytoskeleton dynamics and membrane trafficking between endosome and plasma membrane. At the same time, GIT has been linked to Huntington's disease and HIV infection.[46-48] In all mammals and birds studied, there are two GIT family members: GIT1 and GIT2. The GIT proteins are highly conserved within the orthologs, especially for the N terminal and C terminal regions.[49] GIT has been found to locate in membrane ruffles, focal adhesions and recycling endosomes. [49]

As shown in Figure 1-7, GIT is a multi-domain protein that contains distinct activities in individual domains. The N terminal Arf GTPase activating protein (Arf-GAP) domain exhibits GAP activity for Arf1, Arf2, Arf3, Arf5 and Arf6.[50-53] By interacting with vesicle surface proteins and phospholipid metabolizing enzymes, GIT regulates membrane traffic and receptor recycling. Then there are three ankyrin repeats which mediate the interaction with the endosome.[54,55] Following the ankyrin repeats is the Spa2-homology domain (SHD). The SHD is a tandem repeat that mediates the binding to many proteins, for example, the p21-activated kinase interacting exchange factor (PIX), MEK1 and phospholipase C γ (PLCγ).[56-58] The coiled-coil domain mediates GIT protein homo- or heter-dimerization by a putative leucine zipper motif.[59] The C terminal paxillin binding domain (PBD) is highly conserved.[49] It mediates the interaction with paxillin and is required for the FA localization of the GIT complex.[54,60,61] A structure model of GIT1 PBD was derived from small-angle X-ray scattering (SAXS) and homology modeling based on the FAK FAT domain.[62] To date, a high-resolution structure of the GIT1 PBD is unavailable and the mechanism of the paxillin–GIT interaction remains unclear.

Through the work in this dissertation, I tried to characterise the atomic structure of GIT1 PBD and understand the paxillin –GIT interaction, with methodology like NMR and Biacore surface plasma binding assay.
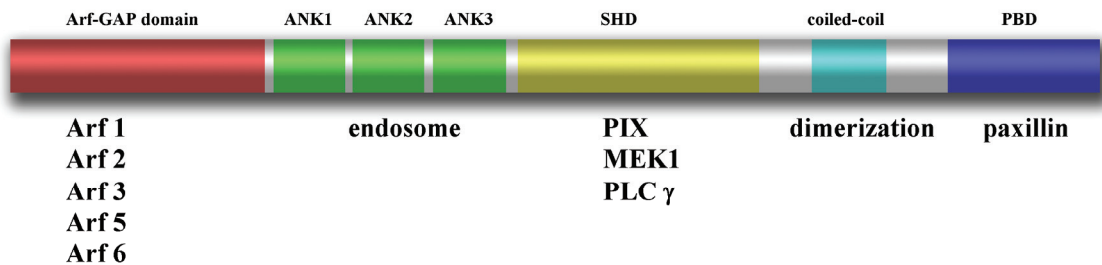
Figure 1-7   The domain structure of full length GIT and its interacting proteins.

CHAPTER 2. METHODS

2.1 Nuclear magnetic resonance (NMR)

2.1.1 *Overview of NMR spectroscopy*

Nuclear magnetic resonance (NMR) is a physical phenomenon describing the behavior of certain atoms in the presence of external magnetic fields. In 1946, Felix Bloch and Edward Purcell independently discovered the nuclear magnetic resonance phenomenon and shared the Nobel Prize in Physics in 1952. NMR spectroscopy was originally developed by Richard Ernst and Weston Anderson as pulsed Fourier transform NMR spectroscopy. With this contribution, Ernst was awarded the Nobel Prize in Chemistry in 1991. After six decades of work, scientists have developed NMR spectroscopy into a widely used technique in studying the chemical and physical properties of small and macro molecules. In 2002, Kurt Wüthrich was awarded the Noble Prize in Chemistry for his contribution in the development of NMR to study biological macromolecules such as proteins. At the same time, NMR also found an important application in biomedical imaging, a technique called magnetic resonance imaging (MRI). In 2003, Paul Lauterbur and Peter Mansfield were jointly awarded the Nobel Prizes in Physiology and Medicine based on their discoveries that lead to the development of MRI.

In the early 1980s, Prof. Wüthrich first provided the idea of sequential assignment, which provided the foundation for the study of biological macromolecules by NMR spectroscopy. It opened a new page for its applications in Structural Biology. Since then, numerous new methods and pulse sequences have been developed. Now, NMR spectroscopy is one of the major techniques used in Structural Biology. To date, there are 7115 structures in the protein data bank (PDB) solved by NMR spectroscopy, accounting for 14.4% of all structures (Table 2-1).

Table 2-1   PDB statistics.
Data source: RCSB website (http://www.rcsb.org/pdb/statistics/holdings.do, accessed on March 4, 2008).[63] Permitted to reprint.

		Molecule type				
		Proteins	Nucleic acids	Protein/NA complexes	Other	Total
Experiment methods	X-Ray	39082	1018	1791	24	41915
	NMR	6168	803	137	7	7115
	EM	114	11	42	0	167
	Other	88	4	4	2	98
	Total	45452	1836	1974	33	49295

Currently X-ray crystallography and NMR spectroscopy are the two main techniques used in Structural Biology. These are the only methods that are capable of determining biological macromolecule structure at atomic level resolution. A comparison between them can help us have a better understanding of the techniques available to structural biologists (Table 2-2).

In general, X-ray crystallography can solve a larger structure with better resolution. However, this structure is static, packed in the crystal, which is like a snapshot of a moving object. Sometimes, the crystal packing can artificially alter the structures. On the other hand, NMR spectroscopy is capable of observing molecular dynamics and associated conformational equilibrium in solution. NMR spectroscopy is able to obtain a lot of molecular parameters, such as individual angles, distances, chemical shifts, coupling constants and tumbling rates. At the same time, this makes it difficult to interpret NMR data, since these parameters are normally observed as a mixed (average) signal of thousands of individual signals from a macromolecule. Because of the complexity of the NMR derived data, it normally takes more time to solve a structure by NMR compared to X-ray crystallography. In general, the size of a structure one can solve by using NMR spectroscopy is smaller than X-ray crystallography. There are two reasons why NMR spectroscopy is limited by the molecule size: 1. the signal overlapping makes unambiguous assignment impossible; 2. the large molecule has a slower tumbling rate, which makes the signal decay too fast to detect (small T_2).

2.1.2 *Physics behind NMR spectroscopy*

2.1.2.1 Spin

Spin describes a property of subatomic particles, like electrons, protons and neutrons. Each unpaired particle possesses a spin of ½. Since they come with plus or minus sign, they can normally pair up and cancel out the NMR observable signal. However, for certain atom types, the spin of particles can not pair up, resulting in net spin. It is these unpaired nuclear spins that are of importance for NMR detection.

Each spin possesses a micro magnetic field, which can be considered as a tiny magnetic moment vector. In the absence of an external magnetic field, these micro magnets are randomly distributed and the total magnetic effect is canceled out. There is only one energy level for the entire population. In the presence of an external magnetic field, the energy distribution of the nuclei will be split into two or more levels, depending on how many unpaired spins. (Figure 2-1) The population in the lower energy level (N^+) is slightly larger than the population in the higher energy level (N^-). The ratio between the two populations is determined by the energy difference and temperature. This is described by the Boltzmann statistics, where E is the energy difference; K is Boltzmann's constant (1.3805×10^{-23} J/Kelvin) and T is the temperature in degrees (Kelvin).

Table 2-2   The comparison of NMR spectroscopy and X-ray crystallography.

	NMR spectroscopy	X-ray crystallography
Sample preparation	Purified stable protein in high concentration in a solution ideally with low ionic strength	Purified protein forming single crystal. Need to screen the best condition, which is the most difficult step
Sample labeling	Depends on application. Sample can be ^{15}N or ^{13}C labeled or both. Deuterium labeling is required for some applications	Selenium-methionine labeling is necessary for MAD based phasing
Observing	All proton atom signals	Diffraction pattern of all heavy atoms
Structure solving	Calculated structure based on the distance (NOE) restraints	Observe the 3-D position of heavy atoms from the electron density map
Molecule size limit	Normally smaller than X-ray method	Can solve larger structures
Resolution	Lower	Higher
Molecule dynamics	Easy to observe	More difficult to observe. Only by B-factor
Binding study	Can obtain binding kinetics, binding affinity and binding site information from simple titration experiments	Can only obtain the binding site information from solving a complex structure
Number of data sets	Can obtain multiple data sets from one sample with different pulse sequences. Can be used for cross validation	One data set from one crystal

No magnetic field   With magnetic field

E

N^-

N^+
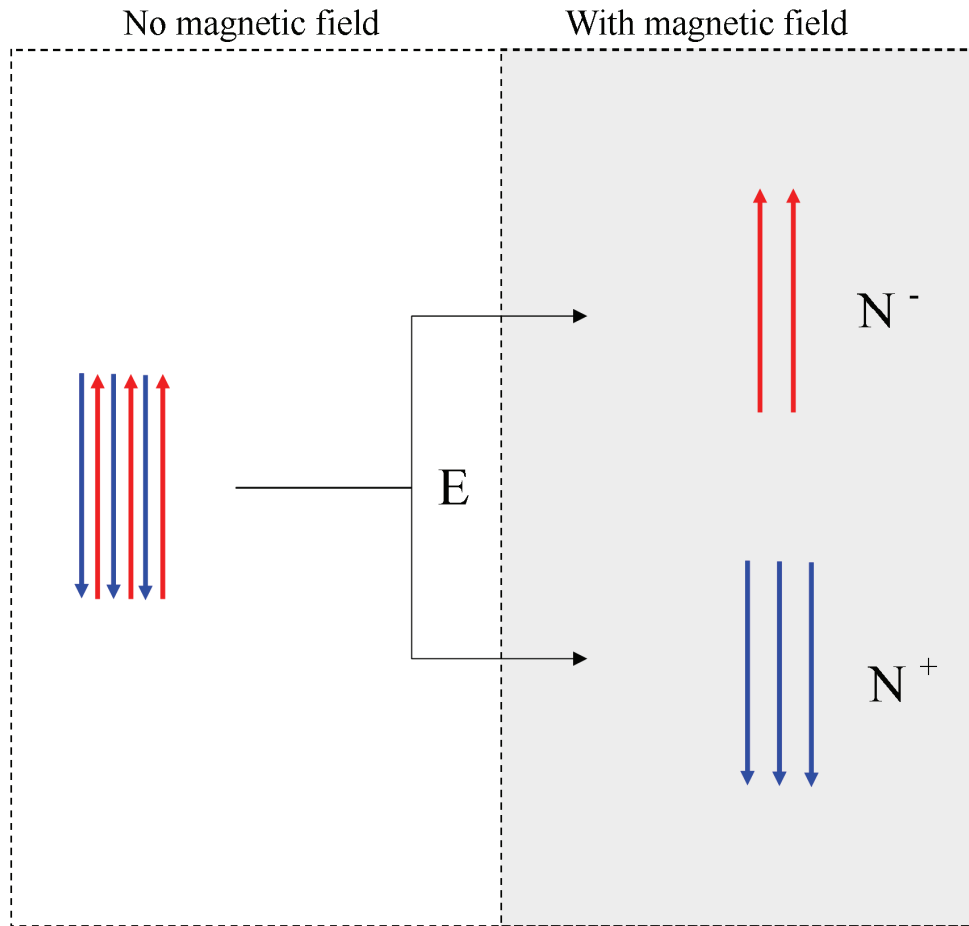
Figure 2-1   Spin energy splitting in an external magnetic field.

$$\frac{N^-}{N^+} = e^{\frac{-E}{KT}}$$ Equation 2-1

Thus, in the presence of an external magnetic field, there is a difference in the spin population of different energy states. The magnetic moment vectors can not cancel each other out, which results in a net magnetic moment vector. The energy gap is determined by the gyromagnetic ratio, γ and the strength of the external magnetic field, B_0. In the following equation, h is the Planck's constant (h=6.626x10^{-34} J s).

$$E = h\,\gamma\,B_0$$ Equation 2-2

Each NMR detectable atom has a characteristic spin frequency, Larmor frequency. It is different according to the magnetic field strength. The Larmor frequency is defined by the magnetic field strength B_0 and the gyromagnetic ratio γ:

$$\nu = \gamma\,B_0$$ Equation 2-3

The actual frequency for one atom is modulated by the chemical environment of that atom. It is the actual frequency that NMR spectroscopy utilizes to study the chemical environment of each individual atom.

2.1.2.2   NMR signal and Fourier Transform (FT)

In the presence of external magnetic field B_0, the net magnetic moment vector is aligned with B_0. This vector can be flipped down to the plane perpendicular to B_0 by another magnetic pulse B_1. The flipped vector will then relax back to its original position through a spiral processing trace. During the relaxing, the spinning vector generates electron magnetic waves that can be detected by coils located in the perpendicular plane. Due to the vector relaxation, the detected signal decays exponentially, generating the raw data called free induction decay (FID).

For biological NMR, a proton is the most common atom that is directly detected. Each proton processes at a special frequency, because it is modulated by its unique chemical environment, including atoms via direct bonding and electrostatic interactions through space. The FID recorded by an NMR spectrometer is a mixture of the relaxation signal of all the protons in the sample. Thus thousands of individual frequencies are mixed and represented as the time domain data FID. Then a mathematical operation called Fourier transform (FT) is used to extract the frequency data out of an FID (Figure 2-2**).

FT is a mathematical method for converting time domain data to frequency domain data.[64] An FT is defined by the function:
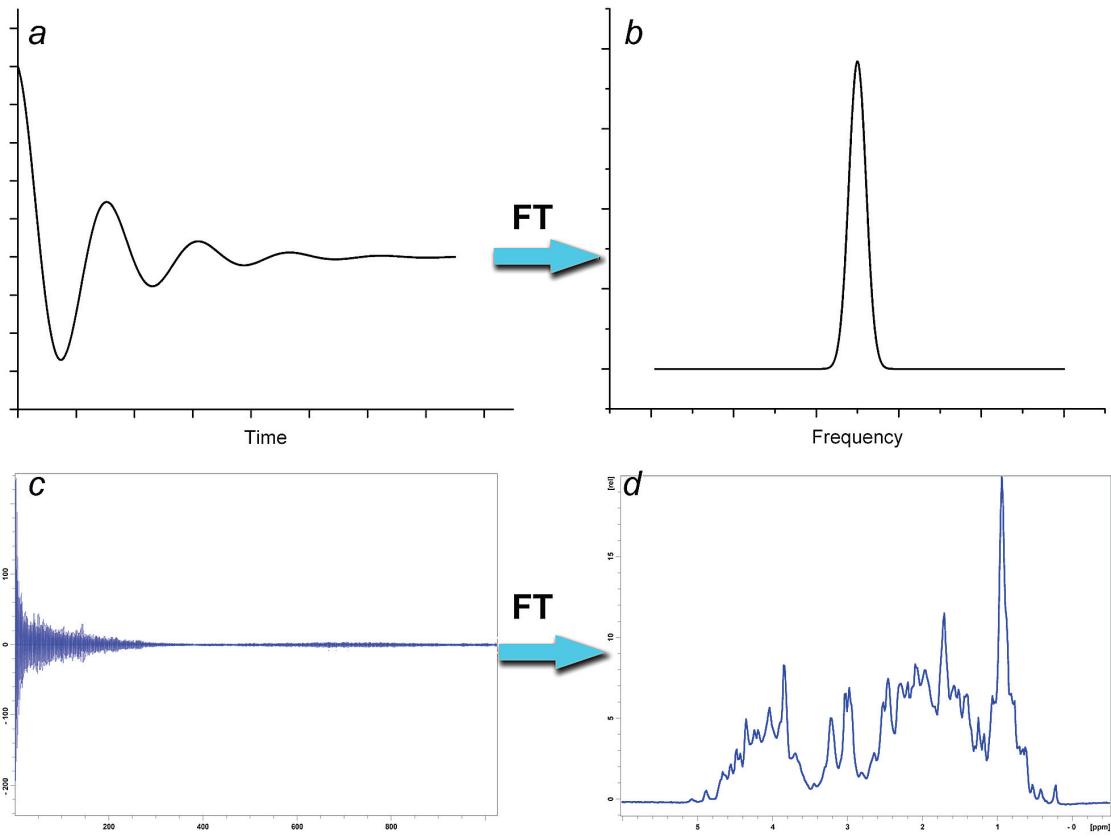
Figure 2-2   Fourier transform converts time domain data to frequency domain data.

$$f(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt = \int_{-\infty}^{+\infty} f(t)[\cos(\omega t) - i\sin(\omega t)]dt \qquad \text{Equation 2-4}$$

After FT, the FID is transformed into readable frequency data, in which each peak corresponds to a frequency. As the atom frequency is a function of external magnetic field, it is difficult to compare frequency directly if two results are coming from different magnetic fields. In practice, an internal reference is used to normalize the frequencies. The normalized frequency is called chemical shift (δ) in the unit of ppm.

$$\delta = \frac{\upsilon - \upsilon_{REF}}{\upsilon_{REF}} \times 10^6 \qquad \text{Equation 2-5}$$

2.1.2.3 NMR spectrometer and pulse sequence

NMR spectroscopy is a very complex system. This diagram is simplified to illustrate the most essential components in an NMR spectrometer (Figure 2-3). A strong magnetic field B_0 is generated by a superconducting coil in an outer shell of the spectrometer. The sample is inserted into the center of B_0. It is wrapped by receiver coils and radio frequency (RF) coils. A central consol controls RF pulses that manipulate the spins through the frequency synthesizer and RF transmitter. When the spins in the sample relax back to the equilibrant state, they can release energy in the form of radio frequency that is detected by the receiver coils. The analog signal is then amplified and converted to a digital one by the amplifiers and analog digital converter (ADC). The control console passes the digitized FID to the terminal computer, allowing it to be converted to the frequency domain data. The data processing and presentation are also accomplished in the terminal computer.

Besides the above components, there are many other critical subsystems. The homogeneity of the magnetic field is critical for modern high resolution NMR spectroscopy. A set of small shimming coils are placed between the sample and the main magnetic field. By altering the current in those coils, the field inhomogeneity can be compensated. For the same reason, the magnetic field is also required to be stable over time. The lock system continuously monitors the signal (frequency) of certain nuclei, normally deuterium in the sample. A drift of the observed deuterium frequency is considered as a drift of the magnetic field. Then the magnetic drift is corrected by another coil in the main magnet, in a feed-back manner. Since most chemical and physical properties are influenced by the temperature, a stable sample temperature is also important. The temperature is precisely regulated by a heated gas system.

A computer controls the spectrometer through programmed RF pulses called pulse sequences. A typical pulse sequence consists of preparation, mixing and

22

B₀

Sample

Receiver
Coil

RF
coil

Magnet Pole

Magnet Pole

RF transmitter, Fre-
quency Synthesizer

Control Console

Computer
User interface

Receiver, Ampli-
fier, ADC
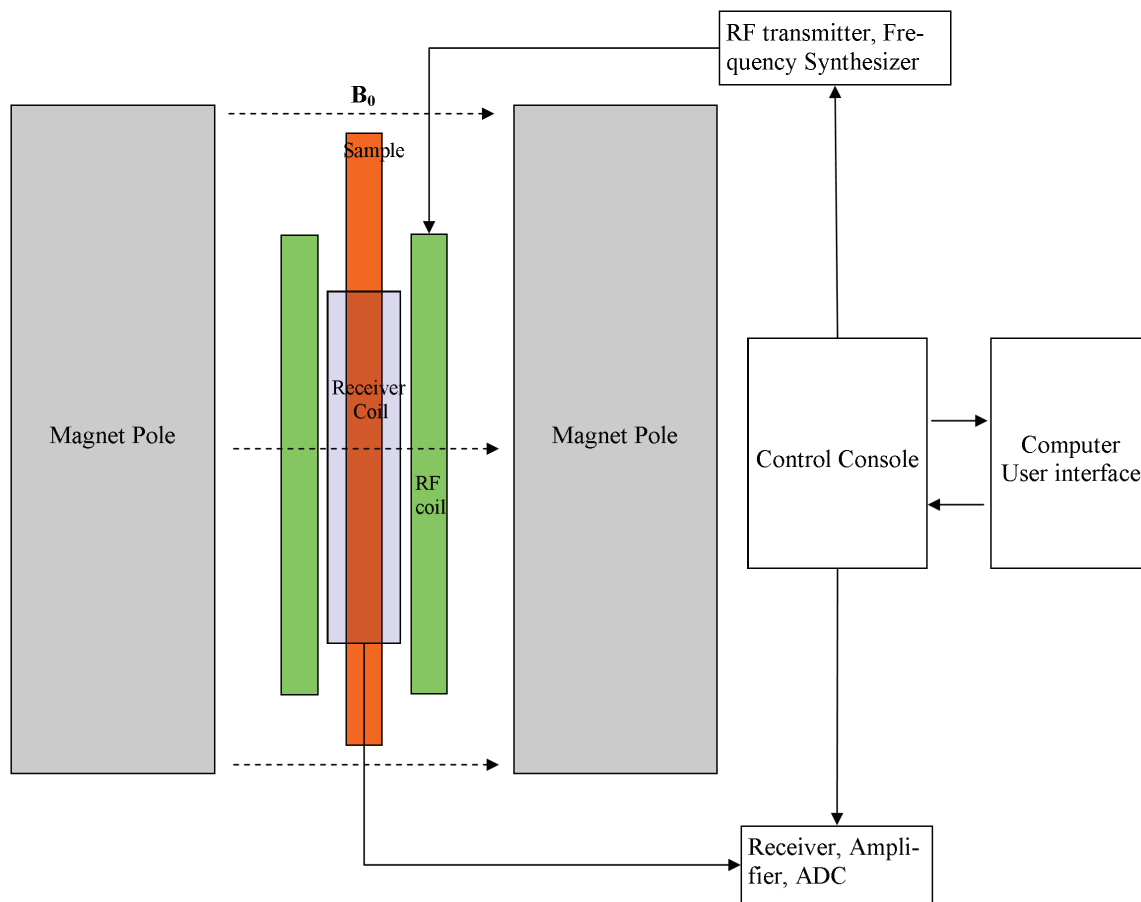
Figure 2-3   Components of an NMR spectrometer.

acquisition time (Figure 2-4). In the preparation, the spins are excited, preparing them to evolve during the following mixing stage. Depending on the specific pulse sequence, the mixing time may be present or absent. During the mixing time, different spin terms are allowed to evolve under the control of pulses. At the end of the mixing time, normally a 'read' pulse is used to prepare the spins for detection. During the acquisition, the signal from the spins is detected by the coils in the plane perpendicular to the main magnetic field. The detected raw data is the FID that can be converted to the frequency domain data.

A multi dimensional pulse sequence is based on the simple one dimensional sequence (Figure 2-4b). The difference is that for the multi dimensional sequence there is a mixing time that is variable for each scan. The mixing time increases equally during the course of a multi dimensional NMR experiment. It allows the coherence transfers to the acquisition stage to be recorded. For each variable mixing time, the same pulse sequence is executed to record multiple FIDs. This will generate an array of 1D NMR spectra that is modulated by the length of the mixing time. The coherent information is extracted by another Fourier transform along the array of the frequency domain data.

2.1.3 *Protein structure determination by solution NMR*

2.1.3.1   Overview

In principle, NMR based structure determination uses experimentally derived constraints, including distance, angle and torsion angle constraints, to build a structure model that agrees with all the constraints while energy is minimized. In order to determine the constraints, all the resonances have to be specifically assigned. Depending on the size of the target, different strategies can be used for spin assignment (Figure 2-5).

A sequential assignment strategy is best for small proteins. For sequential assignment, correlation spectra are analyzed to identify individual spin systems (amino acids). Then the complete sequential assignment is achieved by sequential NOE analysis. Further analysis is needed to resolve more long range NOEs toward a global folding. This strategy works well until the target reaches a size limit, where the overlapping signals make unambiguous sequential assignment practically impossible. For this reason, this strategy is currently used for peptides or very small proteins.

Currently, a main chain directed strategy is widely used for larger protein assignment. First, the main chain assignments and sequential connections are achieved by comparing pairs of backbone specific spectra. Then correlation spectra are analyzed to yield side chain assignment of individual residues. Similar to the sequential assignment strategy, long range NOEs need to be identified for structure calculation. This method simplifies sequential assignment by dealing with backbone only for the first step. Larger proteins are required to be assigned in this manner. It has become the main strategy used for protein structure determination.

a

Transmitter
Channel

d1   p1   d2   pw

Decoupling
Channel   Preparation   Mixing   Acquisition

b

Mixing time

Frequency domain data
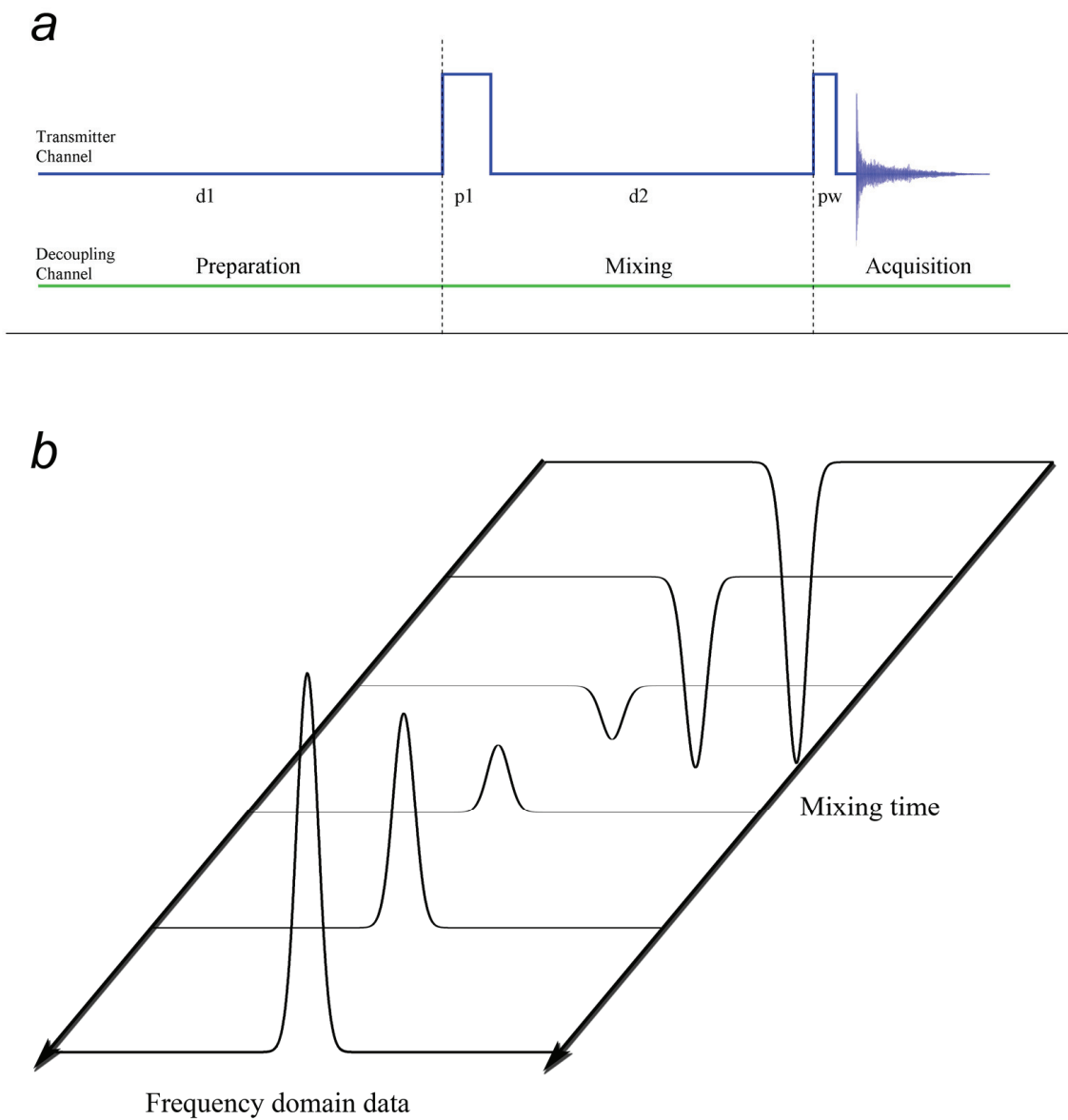
Figure 2-4   **Example of a simple pulse sequence.**
(a) Example of simple one dimensional pulse sequence. (b) Schematic view of the two dimensional NMR spectrum.
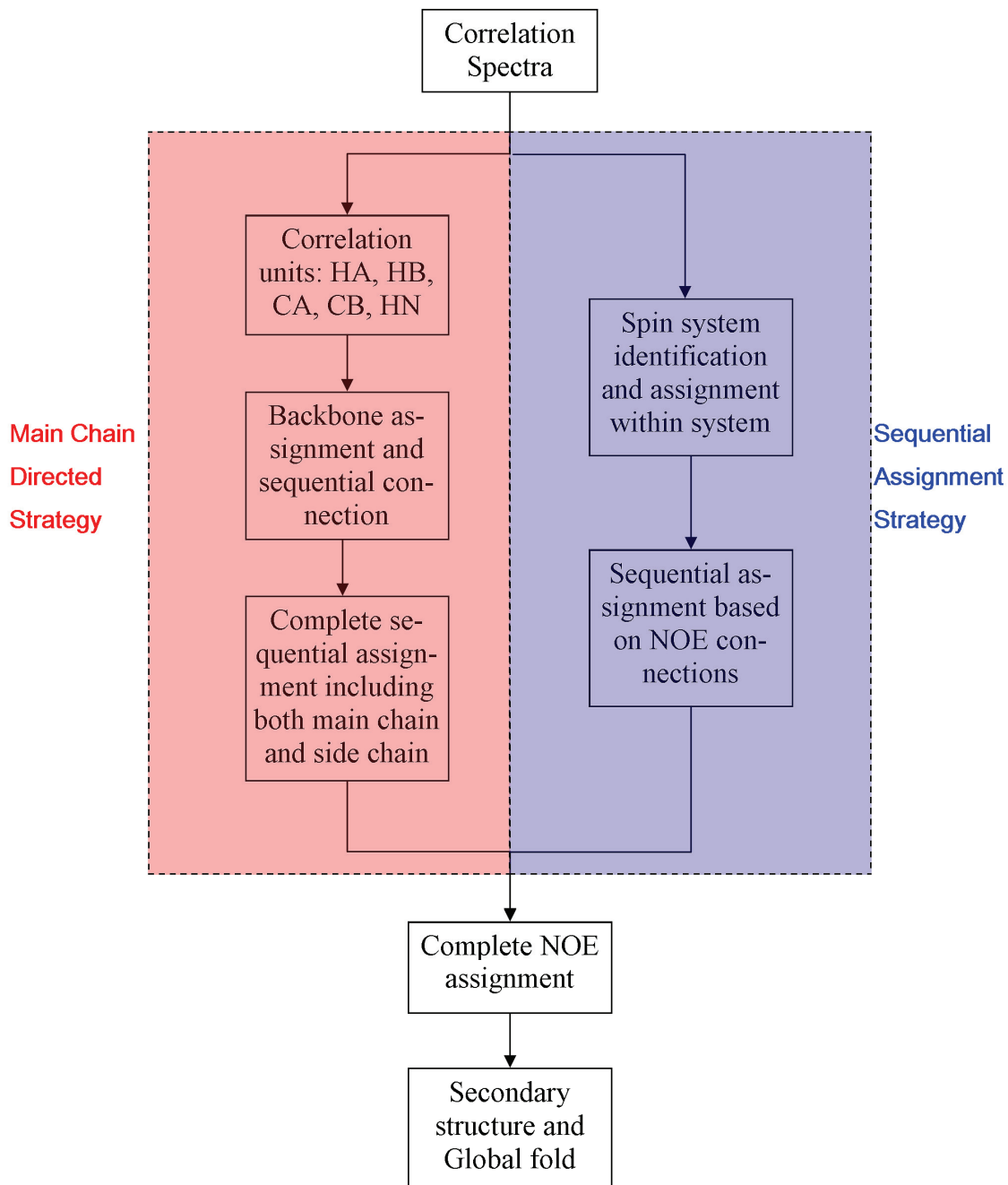
Figure 2-5 **Flow chart of structure determination by NMR.**

Spectrum peak overlapping is a major problem for unambiguous assignment. The use of three-dimensional or four-dimensional heteronuclear spectra greatly reduces peak overlapping. Comparing to the two-dimensional homonuclear experiment, the three-dimensional heteronuclear experiment resolves the resonances in an additional dimension based on the coupling from an additional atom, through bond or space. Those multi-dimensional spectra greatly simplify the assignment, but they take much more time to acquire. Thus, a more concentrated sample is required to reduce scan number for each data point. Normally, the minimum protein concentration is about 1 mM/L for multi-dimensional experiment based structure determination.

For best spectra compliance, all the experiments for one structure determination project are acquired at the same solution condition and temperature. Normally, the sample temperature is between room temperature to 37 °C. The higher temperature is preferred, because molecules move and tumble faster in higher temperatures, which results in sharper and narrower peaks. Sharper peaks are preferable because it increases signal to noise ratio and decreases the chance of overlapping.

In order to take advantage of multi-dimensional heteronuclear experiments, the protein sample should be ^{15}N and ^{13}C doubly labeled. In the case of some large proteins, deuteron labeling is also utilized in addition to the double labeling. Because of the requirement for high sample concentration and temperature, the protein solubility and stability are critical for sample preparation. Normally, a lot of effort is needed to screen different constructs and solution conditions. In the ideal condition, the protein should be a stable monomer at high concentration. There are other conditions contributing to an ideal sample solution: a low salt concentration for a better signal to noise ratio acquiredby cryo-probes; a pH value between 6.0 and 7.0, to avoid deprotonating at higher pH or destabilizing at lower pH.

Since the structure determination in this dissertation is based on the main chain directed strategy, it will be discussed in more detail below.

2.1.3.2   Backbone assignment

For the main chain directed strategy, the first step is to obtain the assignment for the backbone atoms and their sequential connections. The scheme of the assignment strategy is to obtain the backbone atom chemical shift assignment for individual spin systems and link them into a continuous polypeptide chain. With the development of NMR pulse theory, several pairs of three-dimensional experiments are currently available to provide sequential connectivity.

First, spin systems are built up based on each peak in the ^{15}N HSQC spectrum. An ^{15}N HSQC spectrum gives the correlation between each amide proton and the connected amide nitrogen atom. Except for proline, each amino acid has one amide pair in the backbone. So in principle, each and every residue in the protein should have a peak in the ^{15}N HSQC spectrum. Thus, each peak in the ^{15}N HSQC spectrum

corresponds to a residue, although the assignment is still unknown. Based on the peaks in an ^{15}N HSQC spectrum, the strips are built up for three-dimensional nitrogen correlation spectra (Figure 2-6).

For backbone assignment, the most common three-dimensional spectra include HNCACB/CNCA(CO)NH pair, HNCA/HN(CO)CA pair and HNCO/HN(CA)CO pair. In each pair, the former experiment provides the backbone chemical shift for both the current residue and the previous residue. The later experiment provides the backbone chemical shift for the previous residue only. By comparing the pairs, we should be able to differentiate the peaks from the current residue or the previous one. Now, for each and every spin system (residue), the backbone chemical shift is known for both itself and the previous residue, but the connectivity is still unknown.

To connect individual spin systems, the backbone chemical shift values are compared to find a match for its predecessor (Figure 2-7). For example, the HNCA/HNCOCA pair gives us the information that the CA chemical shift value for the predecessor of spin system X is 60.52 ppm. We then looked through all the spin systems and found that only spin system Y has CA chemical shift value of 60.52 ppm. We can link spin system Y and X as a dual-residue peptide piece. Then we can try to find the predecessor of residue Y and continue until reaching the N terminus. In the ideal situation, the entire polypeptide chain can be linked, which finishs the sequential assignment by filling in the residue type with corresponding sequence. In the real situation, the sequential linkage may not be complete from the N terminus to the C terminus, because of proline residues (no peak in the ^{15}N HSQC spectrum), or missing or overlapping peaks. Normally, fragments are first built up and their assignment may be unambiguously determined through some anchoring residues. Since glycine, serine and threonine have characteristic CA and CB chemical shift, their type may be identified independent of other information. Sometimes, there may be unique combinations of the 3 residues in the sequence, thus the assignment of the residues and the linked fragment can be uniquely anchored. From the anchored fragment, more linkage can be found toward both ends. The ultimate goal of the backbone assignment is to try to link all the spin systems and have them uniquely assigned.

What has just been described is the manual assignment, which is complicated, tedious and difficult. With the development of automatic linking algorithms, some of the backbone assignments can be easily done by computer. Among the few automatic backbone assignment softwares, Autolink is one of the best in accuracy and speed.[65] Developed by Masse and Keller, Autolink is an extension to the NMR spectrum analysis software CARA.[66] It utilizes the chemical shift value for a spin system and its predecessor to calculate a most possible sequential assignment. The central idea of the algorithm is to calculate the link possibilities and score the linkage in trying to maximize the linked residue numbers while minimize the total score. It also takes into account the secondary structure and uses the experimental database to further fine tune the scoring function. For small proteins (about 100 residues), the program can achieve very good accuracy and speed. For larger proteins, it also dramatically helps assignment with relatively good accuracy. However, the result from the calculation should always be
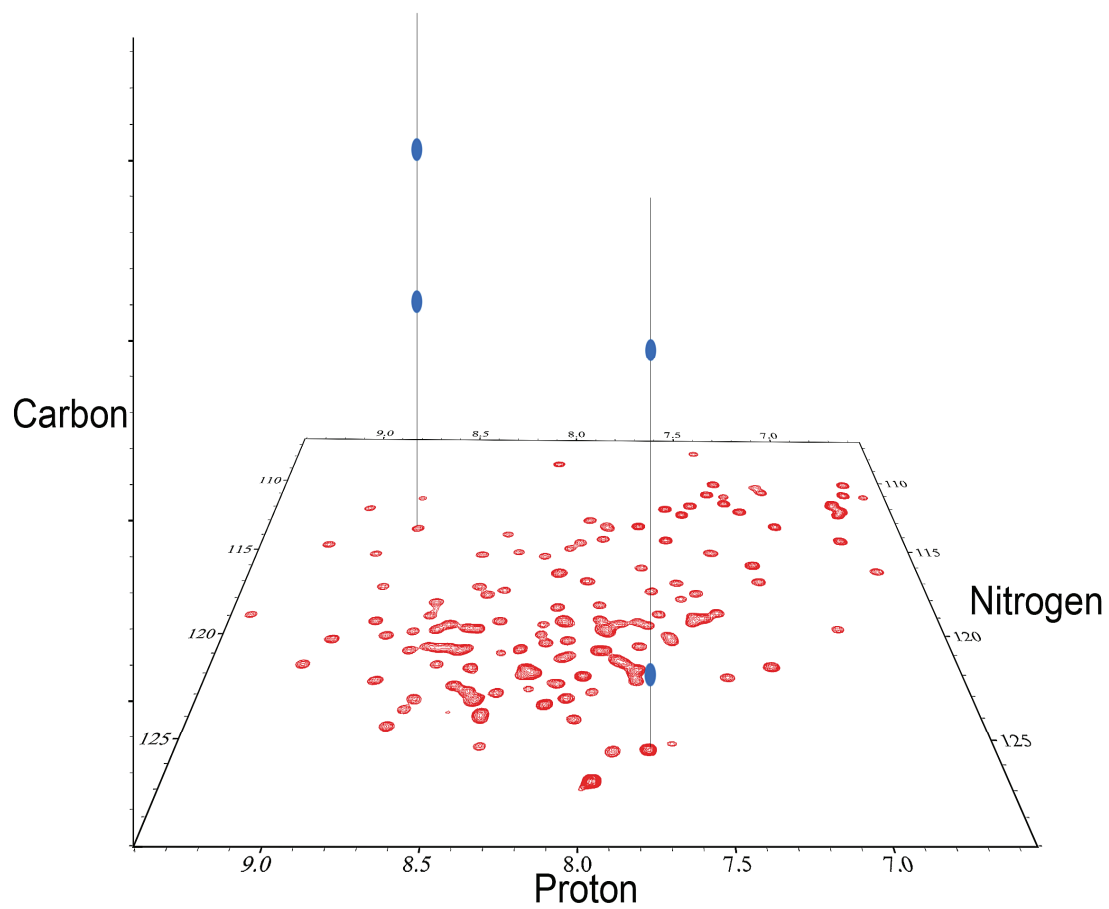
Figure 2-6   Example of 3D nitrogen correlation spectrum and 3D strips.
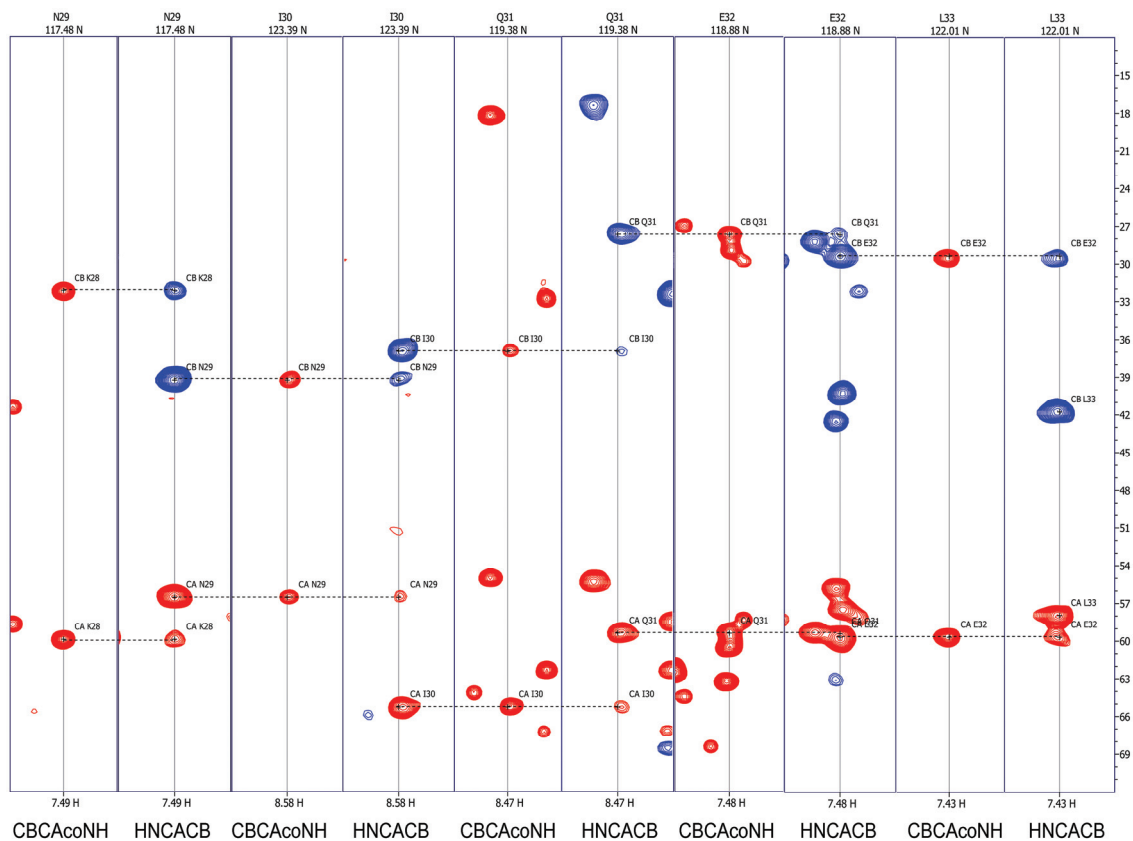
Figure 2-7   Example of backbone connection by backbone walk.
The backbone walk by the HNCACB and CBCAcoNH pair. Each CBCAcoNH strip has two peaks (red): One for the CA peak of the current residue and one for the CA peak of the predecessor. Each HNCACB strip has four peaks: two CA (positive, in red) and two CB (negative, in blue). The larger CA and CB peaks are from the current residue and the smaller CA and CB are from the predecessor.

manually verified to insure accuracy. Normally, the gaps between major fragments have to be finished manually. The best strategy is to use the program to help do most of the easy but tedious backbone assignment and manually verify the results. If used correctly, Autolink can help dramatically increase assignment efficiency and accuracy.

2.1.3.3   Side chain assignment

The next step is to obtain the side chain proton assignments for individual residues. The side chain protons include mostly methyl, methylene and aromatic protons. Normally, this is on the basis of completed backbone assignment, but sometimes the side chain assignment can also help backbone assignment by identifying residue types. Currently, no mature automatic side chain assignment software available, so it is still done manually.

For small proteins, the combination of two dimensional Total Correlation Spectroscopy (TOCSY) and Correlation Spectroscopy (COSY) should be sufficient for side chain assignment. The TOCSY provides the correlation between all the methyl and methylene protons within a spin system (Figure 2-8). The COSY provides the correlation between two protons connected by two adjacent carbon atoms. The COSY gives redundant information of TOCSY, so it is used for validation and supplement for overlapping TOCSY regions. For large proteins, the 2D spectra normally suffer from heavily overlapping methyl and methylene regions. To help reduce overlapping, the combination of 3D HCCH-TOCSY and 3D NOESY-HSQC are used. Like 2D TOCSY,
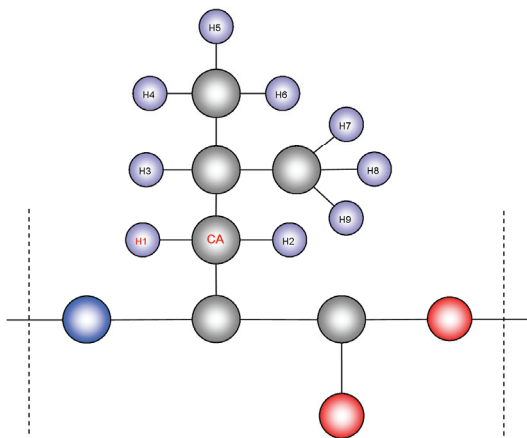
Figure 2-8   TOCSY correlation scheme for Leucine.
Take proton H1 for instance, the TOCSY can display the resonance of all the protons from H2 to H9, in principle. The correlation is mediated by CA and other carbon atoms in the residue.

the 3D HCCH-TOCSY also reveals the correlation between all the methyl and methylene protons within a spin system (Figure 2-8). It resolves the 2D TOCSY with an additional carbon dimension.

To make side chain assignment, strips are built based on the proton-carbon correlation plane which is the same as a 2D carbon HSQC spectrum. Based on the HA, CA, HB, CB chemical shift value obtained from backbone assignment, the α proton and β proton peaks can be located in the proton-carbon plane. For example, we build a strip based on the α proton shift of a particular residue, all the protons that have correlation to the α proton should resolve as a peak in the third proton dimension. Thus, by this means we should be able to assign all the proton peaks within each spin system. The 3D NOESY-HSQC and 3D HCCH-COSY are used as a supplement to the 3D HCCH-TOCSY.

2.1.3.4   NOE assignment and determination of distance restraint

To determine the distance restraints for structure calculation, the NOE peaks have to be unambiguously assigned. For a regular sized protein, the NOE spectrum can be heavily overlapped, because any proton pairs that are close enough (smaller than 5 Å) in space can have a peak. This problem can be partially solved with the 3D ^{15}N NOESY-HSQC and 3D ^{13}C NOESY-HSQC that resolve the 2D NOESY spectrum in the third nitrogen dimension or carbon dimension (Figure 2-9). Currently, besides manual assignment, some NOE assignment can be done automatically with software. Both approaches will be introduced below.
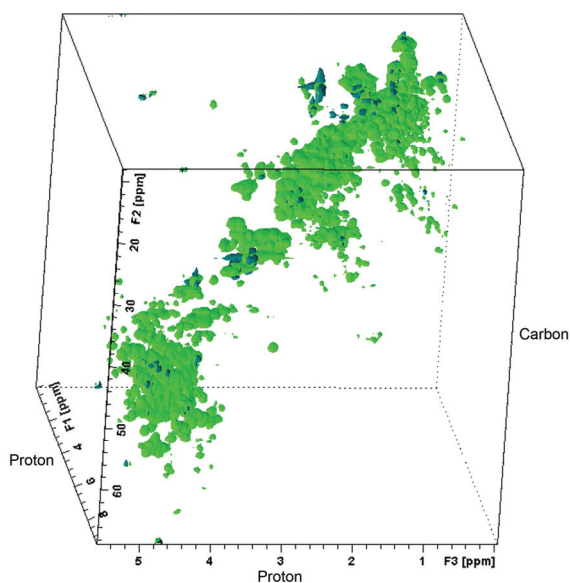
Figure 2-9   **2D NOESY is resolved in the third carbon dimension.**

32

The manual approach employs an iterative strategy. Since most aliphatic protons are dispersed in a 5 ppm chemical shift range, there are some protons having the same or very similar chemical shift values, which means one NOE peak can be attributed to more than one proton. It can be challenging to unambiguously assign the NOE peaks without any structure information. So, NOE assignment is coupled with structure calculation in several iterative steps. The initial assignment only applies to those unambiguously resolved NOE peaks and that can be determined from known structure or structure models. The initial structure will be calculated based on the partial assignment. This structure could be a very loose bundle (not enough distance restraints) with high violations (some mis-assigment).

Based on the preliminary structure, we can check the violations for incorrect assignments. We can also make additional assignments, because the unambiguity is now greatly reduced since only the protons within the 5 Å radius can give NOE cross peaks. Based on the new distance restraints, we can calculate the new structure, which is used to refine the NOE assignments in the next iteration. Thus, the NOE assignment and structure calculation are repeated several times until reaching a satisfying final structure. The criteria for a satisfying structure are defined by small values for target function and RMSD.

Similar to the manual approach, most automatic NOE assignment protocols also exploit the iterative strategy. CYANA is the most commonly used software for automatic NOE assignment. CYANA is a structure-calculation software that integrates the automatic NOE assignment module CANDID.[67,68] The input data for CANDID are the chemical shift values for all the NOE peaks as well as the chemical shift values for all the assigned protons, including the amide protons and aliphatic side chain protons. With the same iterative strategy, CANDID makes the NOE assignment and structure calculation in 7 cycles. The result is normally very reliable for most small proteins, giving enough proton chemical shift completeness. To further automate the *de novo* NOE assignment and structure calculation, CANDID was integrated with the automatic NOE peak pick program ATNOS.[69,70] The input data for ATNOS/CANDID is only the proton chemical shift list and the NOE spectra. Thus, it greatly boosts the efficiency of structure determination by NMR. For a successful automatic NOE assignment calculation, there are some prerequisites. Firstly, the proton assignment should be more than 80% completed. Secondly, a reasonable structure must be reached after the first cycle. Thirdly, the NOE spectra should be carefully calibrated to the proton list.

The purpose of NOE assignment is to identify corresponding NOE peaks and convert the peak intensity into distance restraints. For a rigid two-spin system, the NOE peak volume is proportional to inverse sixth power of distance and overall tumbling.

$$V_{ij} = \frac{1}{r_{ij}^{6}} \times f(\tau)$$

Equation 2-6

Where V_{ij} is the volume of the NOE cross peak between proton i and j; r_{ij} is the distance between the two protons. A weighting function is needed to convert the peak volume

into distance restraints. In CYANA, the module CALIBA helps calibrate the NOE intensity with distance restraints. It has different calibration functions for backbone, side chain and methyl groups. Normally, the NOE derived distance is treated as upper distance limit restraint in the calculation. While the lower distance limit restraint is set to the sum of *van der waal* radii of two proton by default.

2.1.3.5 Structure calculation

NMR can provide multiple restraints from different angles towards an accurate 3D structure. The most important restraint is the inter-proton distance restraints derived from NOE experiments. The scalar coupling constant provides torsion angle restraints and hydrogen bond distance restraints. Chemical shift provides secondary structure information and disulfide bridge possibility. Residual dipolar coupling (RDC) provides H-N and H-C bond orientation information. Although 3D structures can be derived from only the NOE distance restraints, the additional information is able to improve the structure accuracy and precision. Thus, always try to collect as many restraints as possible before structure calculation.

Although many programs are available for 3D structure calculation, they are all based on two common approaches: distance geometry [71] and restrained molecular dynamics (rMD).[72] The distance geometry algorithm samples all the possible inter-proton distances in a geometric matrix and reaches structure solutions satisfying all the experimental derived distance restraints. Software employing distance geometry include DISGEO and DG2.[73] The restrained molecular dynamics approach solves the 3D structure based on the classic mechanics and Newton's equation of motion. A standard force field defines the chemical properties of all the amino acids from experimental knowledge. NOE derived structure information is treated as pseudo force field (pseudo energy terms) that is specific to describe the structure. By rMD, a 3D structure is folded up from random conformers guided by both standard force field and the NMR derived pseudo force field. Currently, rMD can be performed in both Cartesian space and torsion angle space. Because of fixed bond length, the later is about ten times faster than the former, thus it is widely used in most structure calculation softwares, e.g. DYANA, CYANA, X-PLOR and CNS.

In practice, we use CYANA as structure calculation tool.[67] CYANA employs torsion angle dynamics simulated annealing for structure calculation. It integrates the module CANDID for automatic NOE assignment as well as automatic structure calculation from the spin peak list and a list of NOESY peak positions. The automatic structure calculation has been described in the previous section. Besides the automatic way, CYANA can calculate structures directly from various restraints too, the classic manual structure calculation. Conventional manual structure calculation is an iterative process that involves calculation of an initial structure followed by many steps of refinements and re-calculations. The refinement steps are critical because they improve structure by identifying more NOE assignments and eliminating incorrect NOE assignments based on the preliminary structure.

The current version of CYANA can not use residual dipolar coupling data. The most common restraints are the upper distance limit restraint and the torsion angle restraint. The upper distance limit restraint is calculated from the intensity of NOE peaks (Equation 2-6). The torsion angle restraint provides the range of rotation angle of each rotable bond. The backbone torsion angles can be predicted from the chemical shift value of HA, CA, CB, CO and N resonances, by the software TALOS.[74] Based on the database of chemical shift values and sequence homology, TALOS predicts the torsion angle range for the backbone phi and psi angles with good accuracy.[74] The use of torsion angle restraint greatly reduces the conformational space to sample and increases the calculation efficiency dramatically.

During NMR structure calculation, distances can not be uniquely determined because NMR derived distance restraints usually describe a range of distance instead of one fixed value. Thus, the experimental restraints describe an ensemble of structures instead of one exact structure. So structure determined by NMR is usually reported as a bundle of 20 conformers with the best target function. There are three major criteria for evaluating the quality of NMR structure. The target function measures the deviation between the structure model and the experimental data. It illustrates the accuracy of the structure. The root mean square deviation (RMSD) measures the distance deviations within the 20 conformers. It illustrates how tight the bundle is and describes the precision of the structure. The Ramachandran plot visualizes the conformation by plotting the backbone phi and psi angle value of each residue. Based on the statistics, it reports the percentage of residues that are in favorable conformation or disfavored conformation. Usually, the residues in disfavored conformation are the result of incorrect assignment. It describes the structure conformation from statistics and points out the suspicious assignment for further investigation.

2.1.4 NMR as a tool to study protein interactions

2.1.4.1 Overview

In addition to structure determination, the other powerful aspect of NMR is to study the interaction between a protein and another protein, peptide, nucleic acid and small molecules. NMR detects the spin relaxation from RF excitation and the signal is modulated by the environment of the spin. It makes NMR very sensitive to any changes in the chemical and physical environment, including temperature, solution conditions and ligand binding. The binding of a ligand can locally change the chemical environment of the interface atoms or change the tumbling rate of the protein, which are easily detectable by NMR. A simple titration experiment can provide very wealthy information regarding the interaction. In addition, NMR studies protein interactions in the solution, which is closer to the physiological condition compared to the X-ray crystallography. The advantages make NMR the prime tool to study protein interactions, especially for those weak or transient interactions.

Although very versatile, NMR still has its own weakness in studying protein interactions. One of the major disadvantages is that the system it can study is limited by the protein size, which is normally smaller than the limit of other methods. The other limitation is the requirement for backbone sequential assignment, although the titration experiment is relatively easy. Further, the chemical shift perturbation is composite changes, which may make data interpretation difficult if more than one event (e.g. two binding sites or binding plus conformational change) occurs in the mixture. And the binding is studied in high protein concentration, which may limit the value of the results.

NMR is capable of providing different levels of information regarding protein interactions. Among all the applications, the most powerful ones are binding interface mapping, binding affinity measurement and complex structure determination.

2.1.4.2   Binding interface mapping

Many NMR methods can help map the protein interaction interface. A chemical shift perturbation (CSP) titration experiment is the most powerful and easy one. It mainly monitors the environment change of individual amide protons upon titrating ligand into protein, by a heteronuclear single quantum coherence (HSQC) titration experiment. For a ^{15}N labeled protein, its ^{15}N HSQC spectrum should give a peak for each amide proton (Figure 2-10a, b). Except proline, each amino acid has one amide proton in the backbone. So each residue in the protein can be assigned to one peak, the position of which is directly modulated by the local chemical environment of the corresponding residue. Upon titration with ligand, some peaks will change position due to the perturbation of the local environment by the ligand binding. Then the shifted peaks can be easily identified by overlay of the spectra taken before and after adding the ligand (Figure 2-10c). Now if the peak assignment and protein structure is available, the binding interface can be mapped (Figure 2-10d).

Compared to other methods, CSP titration experiments are very fast and accurate. It takes only 20 minutes to acquire a ^{15}N HSQC spectrum at 400 µM/L sample concentration. CSP titration experiments are widely used also because of their capability to detect very weak and transient interactions. However, CSP titration experiments are prone to false positives. Because it only monitors the variation of the local chemical environment, anything that alters the environment should be detected as positive. This includes a change of solution pH, ionic strength or sample temperature. Even when those conditions are carefully controlled, there could still be peak shifts caused by conformational change. There are two common false positive scenarios that can be identified: 1.If shifted peaks are not clustered, it could be non-specific binding; 2. if the majority peaks shift, it could be protein unfolding or aggregation.

Besides a CSP titration experiment, cross saturation or saturation transfer (SAT) is another widely used interface mapping technique.[75] SAT is based on the idea that when two proteins have interaction, the RF excitation can be transferred from one protein to the other through the binding interface. The two proteins are differently labeled: the
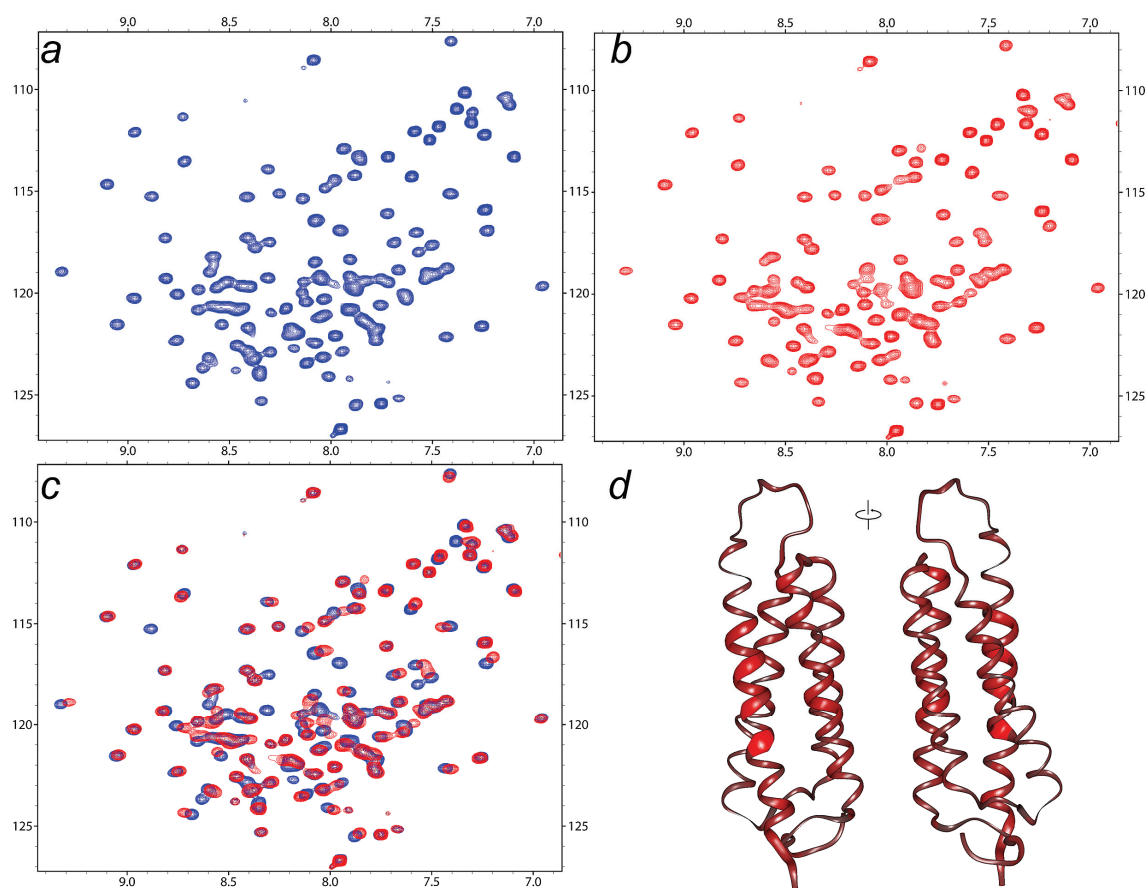
Figure 2-10   Binding site mapping by chemical shift perturbation titration.
*Example of binding site mapping by CSP titration experiment. (a) The ^{15}N HSQC
spectrum of protein alone; (b) the HSQC spectrum of protein with 1:1 unlabeled ligand;
(c) spectra overlay reveals shifted peaks; (d) on the structure, the shifted peaks mapped
to a cluster of residues, which are colored in red. The wider the tube the larger the
chemical shift perturbation.*

"donor" protein is unlabeled and the "observer" protein is perdeuteratued and ^{15}N labeled. The amide deuterons in the observer are then exchanged back to protons in the H_2O solution, in order to be observable by ^{15}N HSQC experiment. Through a cross-relaxation mechanism, the donor protein is saturated and it attenuates the signal in the interface of the observer protein. Hence, the binding interface on the observer can be identified by the signal intensity change. Compared to CSP titration, SAT is more resistant to the false positive caused by conformational change.

A Proton-Deuterium exchange experiment is capable of mapping the solvent exposed surface residues. The protein-protein interaction shields the solvent from the interface residues, so it can be identified by the variation of solvent exposure areas. For ^{15}N HSQC experiment, an amide proton is visible but an amide deuterium is not. In the solution, the amide proton can exchange with the solvent proton if exposed to solvent. If a protein is dissolved in D_2O, some of the amide protons should be substituted by deuterium through solvent exchanging. Reflected in the spectrum, some peaks should disappear over time. Since only solvent exposed residues are able to exchange with deuterium, the disappearing peaks are the solvent exposed residues. Now if some peaks are protected upon binding, the protected residues can be mapped as the binding interface.

2.1.4.3   Binding kinetics

In addition to binding site mapping, NMR is also capable of studying binding kinetics. Interactions can be categorized into fast exchange, intermediate exchange and slow exchange in the NMR timescale (Figure 2-11).
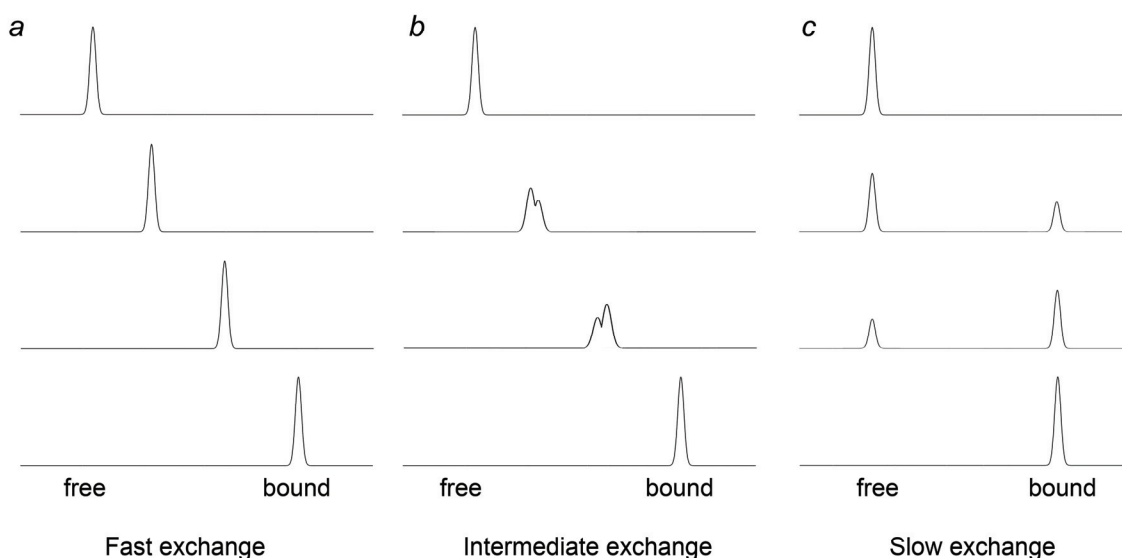
Figure 2-11   Schematic view of different binding kinetics.

Depending on the binding affinity, the receptor is in equilibrium between free and bound forms. Because of different resonance frequency, the receptor free and bound forms should correspond to two distinct peaks in the spectrum. However, when the two states are interchanging too fast to be distinguished by NMR, the two peaks average into one single peak. The position of the average peak is between the two component peaks and determined by the equilibrium. This situation is often typical for weak binding where the ligand associates and disassociates from the receptor very fast, usually in submilliseconds. It is called fast exchange in the NMR timescale (around a second). In contrast, if the interaction is so strong that the ligand disassociates in more than a second, two distinct peaks should exist simultaneously. The ratio of the peak integral is determined by the ratio of the two species in the equilibrium. This situation is called slow exchange in the NMR timescale. Interactions with intermediate disassociate rates exhibit a broad peak at the average position, which is called intermediate exchange in the NMR time scale.

Thus, binding kinetics is observable by resonance behavior upon ligand titration. The binding affinity can also be compared, since binding affinity is directly correlated with exchange rate.

2.1.4.4 Binding affinity

Besides binding affinity comparison by kinetics behavior, NMR is able to directly measure the binding affinity too. As shown in Figure 2-12, during a series of titration experiments, peaks in fast exchange should have a series of movement until they reach an ultimate position, which corresponds to the complete bound form. Each peak is the average resonance of free and bound form at that particular ligand receptor ratio. The plot of chemical shift difference against ligand receptor ratio should result in a standard binding curve that approaches a plateau when ligand/receptor ratio reaches maximum. It can be fitted with standard two parameter nonlinear least square fit one site binding function or this function if considering the dilution effect during titration.[76]

$$\Delta\delta_{composite} = \tfrac{1}{2}\Delta\delta_{max}\left[\left(1+R+\tfrac{P_0R+L_0}{P_0L_0K_d}\right)-\sqrt{\left(1+R+\tfrac{P_0R+L_0}{P_0L_0K_d}\right)^2-4R}\,\right] \qquad \text{Equation 2-7}$$

Where $\Delta\delta_{composite}$ is the composite chemical shift distance for each titration point; R is the ligand to protein concentration ratio; P_0 and L_0 are the initial concentration for protein and ligand stock. The fitting should derive the value for $\Delta\delta_{max}$ and K_d.

For ^{15}N HSQC based CSP titration experiment, the chemical shift difference is normally calculated as composite chemical shift difference by the equation:

$$\Delta\delta_{composite} = \sqrt{\frac{\Delta\delta_{proton}^{2}+(0.2\times\Delta\delta_{nitrogen})^{2}}{2}} \qquad \text{Equation 2-8}$$
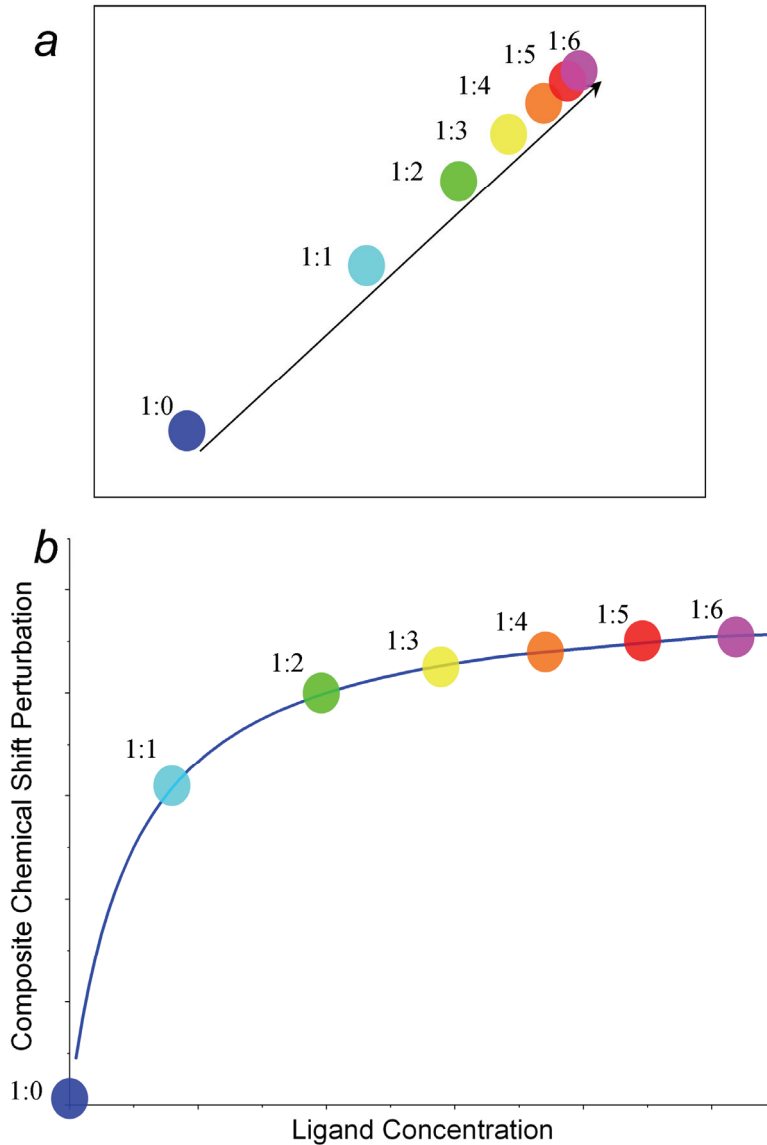
Figure 2-12   Example of binding affinity determination by NMR CSP titration.
(a) For a fast exchange binding, peak moves continuously until reaching an ultimate position, which corresponds to completely occupied receptors. (b) The composite chemical shift perturbation can be plotted with ligand concentration to fit a standard binding curve, from which the Kd and Ymax are derived.

Where $\Delta\delta_{proton}$ is the chemical shift difference in the proton dimension and $\Delta\delta_{nitrogen}$ is the chemical shift difference in the nitrogen dimension.

Since normally there is more than one peak experiencing chemical shift perturbation, the overall binding affinity is usually derived as the average of several peaks in the binding site. The overall binding affinity can also be globally fitted from several individual data sets. Compared to the Kd value measured by isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR), NMR derived Kd values are usually larger (weaker) by about one to two orders of magnitude. The reason for this might be that NMR always measures the average position, which dilutes the effect of binding. Nevertheless, the relative binding affinity value should be correct and thus can be compared to other Kd values measured by NMR.

2.1.4.5   Complex structure determination

Complex structures can provide very rich information regarding protein-protein interactions. It is the most direct way to study protein interactions, especially in studying signal transduction, regulation and enzymatic functions. Currently, most complex structures are still determined by X-ray crystallography. However, NMR is playing a more and more important role in studying weak and/or transient interactions, which are very difficult for X-ray crystallography. There are two categories of complex structures that NMR contributes to: 1. full complex structure determination based on NOEs; 2. NMR-driving docking or modeling.

For full complex structure determination, different strategies were developed over the past decade. The most common one is using intermolecular NOEs detected by $^{15}N/^{13}C$-filtered/edited NOESY experiments. This method works best when the interaction is strong or under condition of slow exchange. If possible, residual dipolar coupling data is also used to provide relative orientations. If the binding affinity is weak or under condition of fast exchange, a transfer NOE experiment is able to provide intermolecular NOEs. Also, the RDC data can be combined for structure determination. When the binding site and relative orientation is known for a protein-protein complex, sometimes it is possible to combine the two components by designing a flexible amino acid linker. This way, the complex can be treated as one single protein with a flexible region. It is usually necessary to check if the linking introduces conformational changes or altered interactions.

For NMR-driving docking or modeling, there are some prerequisites: 1. structures of the individual components are known; 2. the binding interface is identified by NMR; 3. no major conformational change upon binding. NMR is very capable of mapping a binding interface, as introduced in the previous section. However, the interface information can be obtained not only by NMR, but also by mutagenesis, or bioinformatics analysis. The docking integrates the interface information with known individual structures to achieve the complex structure with atom level resolution. Normally, the complex structure will be validated by independent data for reliability.

The validating data includes NMR cross saturation data, H/D exchange data, chemical cross-linking, RDC data and more mutagenesis. Of all the NMR-driving docking software, HADDOCK is commonly used because of its accuracy and efficiency. It will be discussed in more details in the next section.

2.2   Data driving complex modeling by HADDOCK

2.2.1   Overview

Protein-protein complex structure is critical in understanding protein interaction, regulation and enzymatic functions. Despite its importance, there are only a small number of protein complex structures available. The main reason for the limited number is that the complex structure determination is much more difficult than solving protein structure alone. At the same time, more and more structures of binding partners are determined and their interactions are characterized by various biochemical and biophysical approaches. To solve the problem, Dominguez et al developed the docking software HADDOCK (High Ambiguity Driven biomolecular DOCKing).[77]

In general, computational docking uses a similar protocol – translate and rotate one binding partner around the other partner and each pose is scored by various terms, which include electrostatic interaction, van der Waals and surface complementarities. The problem is that the conformational space to sample becomes very large with the increase of target size. As a result, those docking protocols may not be able to find the correct pose or provide unique resolution. Different from other docking software, HADDOCK makes use of known binding data to help maximize docking accuracy and efficiency. Any information describing the binding site on any binding partner can be used as input data, which includes NMR chemical shift perturbation data, NMR H/D exchange data, NMR cross saturation data, biochemical mutation binding study and bioinformatics sequence analysis. Described as AIRs (Ambiguous Interaction Restraints), the information is defined as ambiguous distance restraints used to drive the docking process.

HADDOCK integrates ARIA script with CNS.[78,79] The core structure calculation is executed by CNS, but the accuracy and efficiency is greatly improved with experimentally derived restraints. Because of the accuracy, more and more studies have been carried out with HADDOCK and the targets were also extended from protein-protein complex to protein-DNA, protein-RNA and protein-polysaccharide complexes.

2.2.2   Definition of AIRs

In HADDOCK, the binding site residues are grouped into "active" and "passive" residues based on NMR titration experiments or mutation experiments or any other experimental information. For NMR titration data, the "active" residues are defined as

significantly shifted residues with high solvent accessibility (>50%). The "passive" residues are defined as all other shifted residues and/or the surface neighbors of the "active" residues, with high solvent accessibility (>50%).

The AIR is defined as the inter-atomic distance restraint that has a maximal value of 3.0 Å. For a complex formed with A and B, AIRs consider all the distances between any atom in the "active" residues in A to all atom in the "active" and "passive" residues in B; and any atom in the "active" residues in B to all atom in the "active" and "passive" residues in A. In this definition, the "passive" residues do not participate in the AIR to the partner, but can satisfy the AIR from the "active" residues in the partner. The 3 Å maximal distance value is the best representation for hydrogen/hydrogen and heavy atom/heavy atom van der Waals distance.

The great advantage for introducing AIR is that the number of possible conformational spaces to sample is dramatically reduced during docking procedure. This simplification is guided by experimental information describing the binding interface, so the chance of introducing error is minimized. With greatly reduced conformational space to sample, the docking can afford more time consuming but accurate energy minimization steps. This also helps promote the unique docking solution.

2.2.3   Docking protocols

HADDOCK uses CNS as the core module for docking and structure calculation. There are three steps in the docking protocol.

1. *Randomize the orientation and rigid body energy minimization.* The two binding partners are initially placed 150 Å away in space. They are allowed to translate and rotate while rigid body energy minimization is performed. Under the restraint of AIRs, the orientations of the binding partners are sampled to optimize the contacts. This step determines the pose and orientation for the complex, leaving local energy minimization to the next step. Normally, this step is relatively fast.

2. *Semi-rigid body simulated annealing in torsion angle space.* Residues in the binding interface are allowed to move in order to optimize side chain contact. The binding interface residues are defined as all the "active" and "passive" residues and 2 residues in extension to both ends. Three stages of simulated annealing refinements are performed, in which the flexibility in the interface increases gradually. This step provides the basis for final water refinement and scoring. In general, it is the most time consuming step among all three steps.

3. *Refinement in Cartesian space with explicit solvent.* An 8 Å shell of TIP3P water box is added before actual molecular dynamics refinement. This is a rather gentle refinement. No significant movement will occur in this step. Although small, the movement is necessary to optimize the energetics in the binding interface. Then, all the poses are scored with average interaction energy and average buried

2.2.4 Results analysis

Current scoring functions can rank the binding energy more accurately if all the binding poses are similar. However, HADDOCK is driven by highly ambiguous restraint, so there will be more than one docking solution generated. Thus, the data analysis strategy is to cluster and compare scores within the cluster.

HADDOCK provides scripts for complex structure clustering. The complex RMSDs are defined by superimposing the interface backbone atoms of protein A and calculating the RMSD of the interface backbone atoms of protein B. The complex structures are then clustered by this complex RMSDs based on a user defined cut off value. By fine tuning the cut off value, all the complexes can be clustered into several groups (less than 5, ideally). The top ranked complex solutions in each group should be carefully inspected. The best model should satisfy the following conditions: 1. Having one of the lowest binding energies; 2. Clustered in one of the largest groups; 3. satisfying all the experimental data.

For most cases, this docking and analysis protocol should yield a reasonable complex structure. However, in some difficult situations, a good model may be difficult to obtain. For trouble shooting and re-docking, careful analysis is needed to fix the problem. If there are too many clusters even for a small RMSD cut off value, the complexes are too diverse. The main reason for the large diversity is insufficient AIRs to restrain the binding into a limited number of poses. More experimental data is needed to gather enough restraints. If docking results in a positive binding energy or is not be able to reach a solution, there might be some error in the structure or the AIRs or both. Careful experiments are needed to map the binding interface and rule out the possibility of binding induced conformational change.

CHAPTER 3. THE GIT1 PAXILLIN BINDING DOMAIN IS A FOUR-HELIX BUNDLE AND IT BINDS TO BOTH PAXILLIN LD2 AND LD4 MOTIFS *

3.1   Introduction

Cells attach and communicate with the extracellular matrix (ECM) through membrane peripheral proteins that form focal adhesions (FAs).[80,81] Cell motility is regulated through the alternating assembly and disassembly of FAs and cytoskeletal proteins.[82] The dynamics of FAs is controlled by the signaling of different adhesion molecules such as focal adhesion kinase (FAK), paxillin, the G protein-coupled kinase (GRK)–interacting (GIT) protein, and p21-activated kinase (PAK),[52,83-85] whereas cytoskeletal remodeling is regulated by small GTPases of the Ras and Rho family, such as Rac1, Cdc42, and RhoA.[86-88]

GIT proteins play an important role in initiating the disassembly of FAs.[12,89] Both members of the GIT protein family, GIT1 and GIT2/p95-PKL, have an N-terminal Arf GTPase-activating protein domain (Arf-GAP), a Spa2-homology domain (SHD), a coiled-coil (CC) domain, and a C-terminal paxillin-binding domain (PBD). Previous studies have shown that the SHD binds the PAK–PIX complex and the PBD binds paxillin.[12,38,61,62] GITs, functioning as scaffold proteins, target the PAK complex into FAs by binding with paxillin via their PBDs.[12,38,56] GIT PBDs span about 130 amino acids and are highly conserved among species.[49] Recently, a low-resolution structural model for the GIT1 PBD was derived from small-angle X-ray scattering (SAXS) and homology modeling, based on the structure of FAK focal adhesion targeting (FAT) domain, which is a four helix bundle protein.[62] Mutagenesis studies suggest that paxillin binds the GIT1 PBD through the putative helices 1 and 4.[62] However, a high-resolution structure of the GIT1 PBD is unavailable and the mechanism of paxillin–GIT interaction remains unclear.

Paxillin is one of the major binding partners of GIT proteins in FAs, functioning as an elongated adaptor protein that links actin filaments with integrin.[90] It contains multiple docking sites for different signaling and structural proteins in FAs and comprises five N-terminal LD motifs and four C-terminal LIM domains.[29,30] The LD motifs are named after their consensus sequence LDXLLXXL and bind with multiple proteins, including FAK, GIT, vinculin, actopaxin, and integrin-linked kinase (ILK).[33,36-38] Structural studies have shown that the LD2 motif and the bound form of the LD4 motif form amphipathic α-helices, with several leucines forming a large hydrophobic patch.[31,91] FAK interacts with paxillin on both LD2 and LD4 motifs.[33,92] Unlike FAK, the interaction between GIT1 and paxillin is reported to be mediated by the LD4 motif only.[38,61]

In this study, we determined the solution structure of the PBD of rat GIT1 (residues 640–770) by nuclear magnetic resonance (NMR) spectroscopy. Using synthesized LD peptides, we also studied the interaction between the GIT1 PBD and paxillin LD motifs. Our finding reconciles some controversial observations of earlier studies and provides a clearer picture of the role of GIT proteins in focal adhesion regulation.

3.2 Experiment procedures

3.2.1 *Protein purification and peptide synthesis*

Full-length rat GIT1cDNA was a gift from Dr. Edward Manser, Institute of Molecular and Cell Biology, Singapore. The sequence of rat GIT1 is identical to human GIT1 sequence except one substitution (P644L) in the N terminal unstructured region. We subcloned the GIT1 PBD (residues 640–770) into the pET28 vector (EMD Biosciences, San Diego, CA). The protein was over expressed in Escherichie coli strain BL21(DE3). ^{13}C/^{15}N labeled sample was prepared by growing the cells in 3-(N-morpholino)propanesulfonic acid (MOPS) buffered medium containing ^{15}NH4Cl (1 g/L) and ^{13}C6-glucose (3.6 g/L).[93] Proteins were purified using a Ni2+-charged His-Bind Resin (EMD Biosciences, San Diego, CA) according to the manufacturer's protocol. The 6XHis tag was removed by thrombin cleavage (EMD Biosciences, San Diego, CA) at room temperature for 4 hours. The digested protein was dialyzed in 20 mM potassium phosphate (pH 6.5), 5 mM 1,4-dithiothreitol (DTT), and 5 mM EDTA then further purified by gel filtration on a HiLoad 26/60 Superdex 75 HR column (GE Healthcare, UK). Human paxillin LD2 (residues 140–161), LD4 (residues 261–282) and the LD4 peptide phosphorylated at the S272 position were chemically synthesized and purified by HPLC at the Hartwell Center of Bioinformatics and Biotechnology, St. Jude Children's Research Hospital, Memphis, TN.

3.2.2 *NMR spectroscopy*

The ^{15}N/^{13}C double-labeled GIT1 PBD sample used for structure determination was approximately 400 μl at a concentration of 1.2 mM in 20 mM potassium phosphate buffer (pH 6.5), 5 mM deuterated EDTA, 5 mM deuterated DTT, and 5% (v/v) D2O. All NMR spectra for structure determination were acquired at 37°C with either a Bruker Avance 600 MHz spectrometer or a Bruker Avance 800 MHz spectrometer, both equipped with a cryoprobe. To determine the residues exposed to the solvent, a ^{15}N-labeled GIT1 PBD sample was first lyophilized in a Labconco FreeZone Plus 6 freeze dry system (Labconco Corp., Kansas City, MO) and dissolved in the same volume (400 μl) of D2O. ^{15}N heteronuclear single-quantum coherence (HSQC) spectra were collected at 25°C every 10 min by the Bruker 600 MHz spectrometer. Spectra were analyzed by the software Sparky.[94]

3.2.3 NMR data analysis and structure calculation

All NMR spectra were processed with NMRPipe [95] and analyzed with CARA.[66] Backbone assignments were obtained based on CBCA(CO)NH, HNCACB, HNCA, and HN(CO)CA experiments. Side-chain assignments were obtained by using HCCH-TOCSY, HCCH-COSY, HBHA(CO)NH, CC(CO)NH, and HCC(CO)NH spectra. Dihedral angle range was predicted by TALOS.[74] For structural determination, the NOE distance restraints were derived from ^{13}C-NOESY and ^{15}N-NOESY spectra. NOE spectra were first analyzed by an automatic algorithm with the software ATNOS/CANDID [69,70] and the results were then manually inspected and modified. Structure calculation and refinement were carried out by using the software CYANA 2.1.[68] The final structures were checked and validated by PROCHECK.[96] Structure figures in this paper are generated by Molmol.[97] For electrostatic potential map, the figure was generated by Pymol, based on the electrostatic potential generated by GRASP.[98] Structure superposition and structure based sequence alignment are generated by the Swiss-Pdb viewer.[99]

3.2.4 Chemical shift perturbation titration

All NMR titration experiments were performed at 25°C under the same conditions: ^{15}N-labeled GIT1 PBD at a concentration of 400–450 μM in 20 mM potassium phosphate (pH 6.5), 5 mM deuterated EDTA, 5 mM deuterated DTT, and 5% (v/v) D_2O. All peptide stocks were prepared in the same buffer and their pH was readjusted to 6.5 before titration. Titrations were made by adding ligand at the following ratios: 1:0.5, 1:1, 1:2 and 1:5. A series of ^{15}N HSQC spectra were taken on a Bruker Avance 600 MHz spectrometer with cryoprobe and data were analyzed with Sparky.[94]

3.2.5 Biacore binding assay

Synthesized LD peptides with the N-terminal biotin tag were attached to a NeutrAvidin-covered gold surface (CM5 chip; Biacore, GE Healthcare, UK). Kinetic studies were performed at 20°C with a BIACORE 3000 (Biacore, GE Healthcare, UK). The purified GIT1 protein (in 20 mM phosphate buffer (pH 6.5), 150 mM NaCl, 5 mM EDTA, 0.1 mg/mL BSA, and 0.005% P20 surfactant) was injected to flow through the chip and Kd was derived by fitting the data from three injections. The association and dissociation were observed at a flow rate of 50 μL/min in a concentration range of 1.4–110 μM. Binding affinities were determined by the program Scrubber 2 (Version 2.0b, BioLogic Software). The experiment was repeated with chips covered with both high- and low-density peptides and the results were found compatible.

3.2.6 Circular dichroism (CD) spectroscopy

All CD spectra were obtained with an Aviv 62DS CD spectrometer (Aviv, Lakewood, NJ). CD spectrum of the PBD was taken at 25°C in 50 mM potassium phosphate (pH 6.5) and 1 mM EDTA. The protein thermo stability was determined by monitoring the CD signal of the PBD sample at 222 nm while increasing the sample temperature.

3.3   Results

3.3.1   Solution structure of the GIT1 PBD is a four-helix bundle

The PBD of rat GIT1 (residues 640–770, Figure 3-1a) was expressed in E. coli as a soluble protein, and CD spectroscopy showed that the protein is predominantly helical; and the PBD started to denature only at about 45 ºC (Figure 3-2). The solution structure of the PBD was determined by heteronuclear multidimensional NMR spectroscopy. After backbone and side-chain assignments, the Nuclear Overhauser Effect (NOE) constraints were obtained from both ^{15}N-edited 3D-NOESY and ^{15}N/^{13}C-edited 3D-NOESY spectra. The structure of the PBD was determined based on 2440 NOE distance constraints and 189 dihedral angle constraints. The structures were calculated by CYANA, and 20 structures with the lowest target function value were selected and superimposed (Figure 3-1b). Table 3-1 gives the final structural statistics. All experimental NMR constraints are satisfied, the average target function of the 20 structures selected is 0.7, and no NOE constraint violation is more than 0.2 Å. Most residues (81%) have dihedral angles in the most favored region of the Ramachandran plot, and 10.9% residues are in the additionally allowed region. The only residue in the disallowed region is D643, which is located in the unstructured N terminus. The solution structure is of high precision. The average root mean square deviations (RMSDs) of the 20 structures from the average structure for backbone atoms and all heavy atoms of residues 643–767 are 0.39 Å and 0.77 Å, respectively. The Coordinates and the structure factors of the GIT1 PBD have been deposited in the Protein Data Bank (PDB) with accession codes 2JX0.

The GIT1 PBD forms a compact rod-shaped protein with a well-defined C terminus and a slightly floppy N terminus. The C terminal end is in proximity to the N terminus, a feature common to many functional independent domains. The PBD structure forms a four-helix bundle with a right-handed up-and-down topology, about 58 Å in height and 26 Å in diameter (Figure 3-1c). There are many NOEs observed between helix turns. The four helices form an antiparallel bundle, with an average axis angle of 19.1º between neighboring helices. The hydrophobic side chains from all helices interlace to form a network of hydrophobic interactions that dramatically stabilize the bundle (Figure 3-1b and c). Extensive interhelical NOEs were observed, resulting in well defined side chains of the residues located in the interhelical regions (Figure 3-1c).
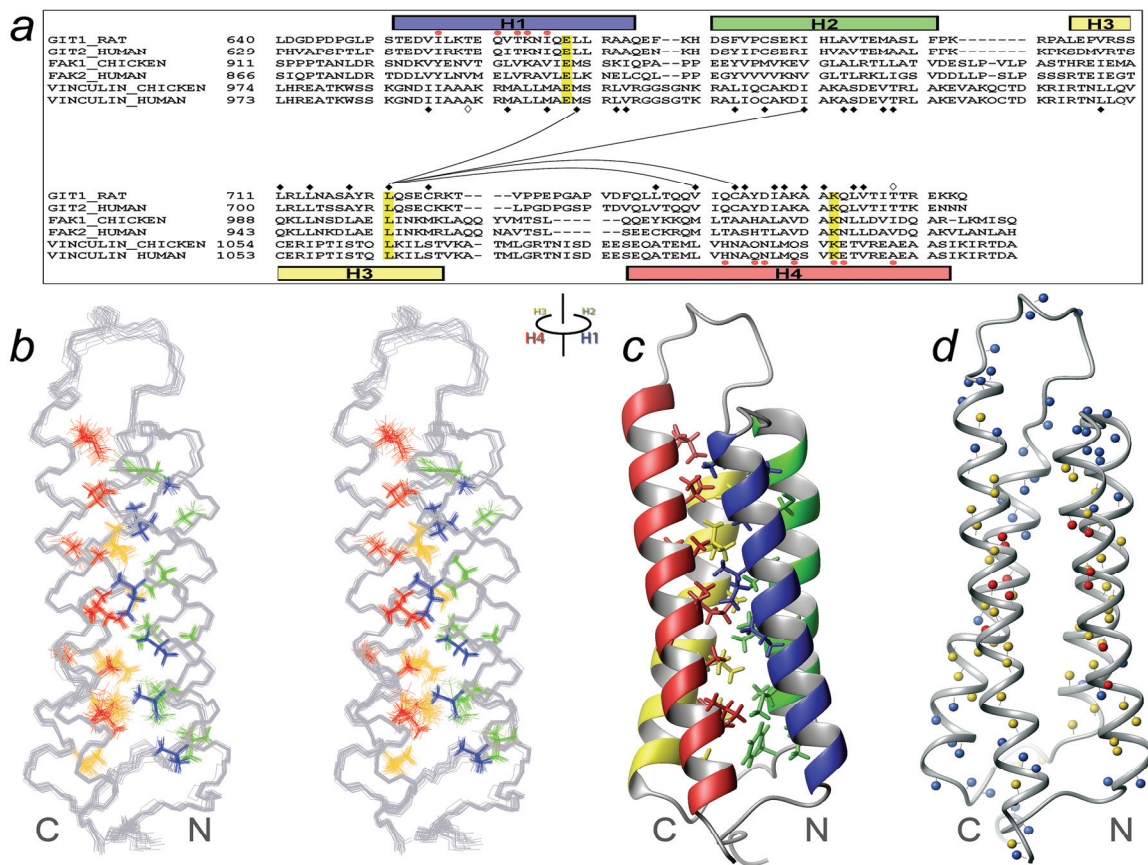
Figure 3-1   Structure of the paxillin-binding domain (PBD) of rat GIT1.
(a) Sequence alignment based on structures of the rat GIT1 PBD, the chicken FAK FAT domain (1KTM), and the H2 to H5 of chicken vinculin Vt domain (1QKR). The sequence of human GIT2, FAK2/PYK2, and vinculin are also aligned based on the structure-based sequence alignment. Residues that are significantly perturbed on binding of LD peptides are indicated by red dots. The hydrophobic core is composed of residues marked by black diamonds. Hydrophobic residues are marked by filled diamonds and the two polar Ts are marked by empty diamonds. The absolutely conserved residues are highlighted in yellow. The example of the hydrophobic network in Figure 3-3 is represented by black-line connections. (b) Stereo view of the backbone trace of the ensemble of the 20 best structures. The side chain of the hydrophobic core is shown.   (c) Ribbon representation of the GIT1 PBD structure. H1 through H4 are shown in blue, green, yellow, and red, respectively. Side chains of the interlaced core are also presented. (d) Representation of the H/D exchange experiments. The unambiguously observed non-proline amide proton atoms are represented as small balls in the structure. After the first 15 min, the solvent-exchanged amide proton atoms are colored in blue whereas the protected amide protons are colored in yellow. The atoms that remain protected after 380 min are colored in red.

49
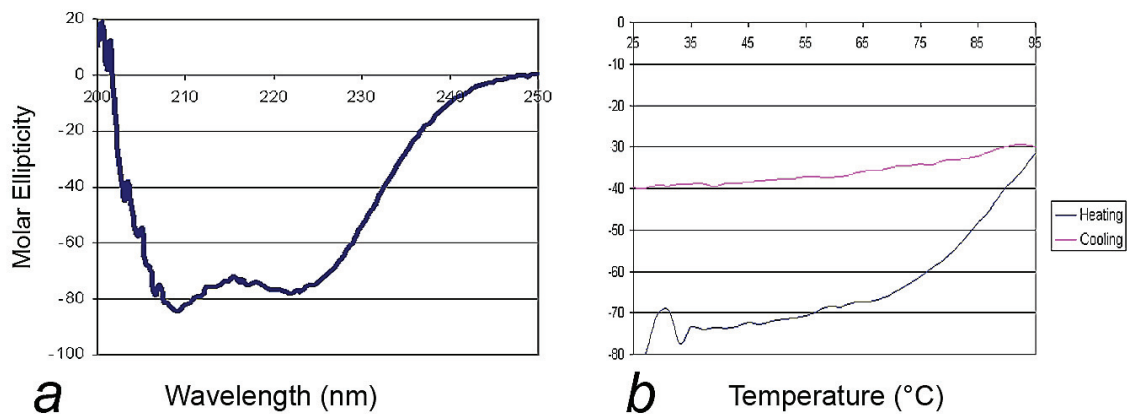
Figure 3-2   CD spectra of purified GIT1 PBD.
(a) CD spectrum of the GIT1 PBD at 25°C in 50 mM potassium phosphate (pH 6.5) and 1 mM EDTA. (b) Molar ellipticity measured at 222 nm while temperature was changed from 25°C to 95°C and back to 25°C.

Table 3-1   Statistics of the GIT1 PBD structure determined by nuclear magnetic resonance (NMR) spectroscopy.

Distance restraints	Numbers
Total number of NOE restraints	2440
Intraresidue	574
Sequential	750
Medium range	718
Long range	398
Restraints per residue	18.07

Dihedral angle restraints	Numbers
Total number of restraints	189
Phi angle restraints	93
Psi angle restraints	96

Ramachandran plot	Percentage
Residues in most favored regions	88.1%
Residues in additional allowed regions	10.9%
Residues in generously allowed regions	0.5%
Residues in disallowed regions	0.4%

RMSD (residue 643-767)	Values
Average backbone RMSD to mean	0.39 ± 0.09 Å
Average heavy atom RMSD to mean	0.77 ± 0.07 Å

Target function	
	0.70 ± 0.04

The PBD fold was confirmed by the solvent-exposed surface mapped by the hydrogen/deuterium (H/D) exchange experiments. Fast ^{15}N HSQC spectra were taken immediately after dissolving the lyophilized ^{15}N-labeled PBD in D2O. After the first 15 min, approximately 57% of amide peaks had disappeared because of H/D exchange. All the protected peaks are in the helical regions, whereas most peaks that vanished belong to the residues in the loop regions and the termini (Figure 3-1d). Some peaks were protected even after 6 hours; they are all within the core of the four-helix bundle. The H/D exchange data validates the 81 pairs of H-bonds in the structure, most of which were between O_i and HN_{i+4} from individual helices.

Of the four helices in the PBD structure, H4 (30 residues) is slightly longer than helices H1 (23 residues), H2 (21 residues) and H3 (22 residues). The H3/H4 loop is longer than the H1/H2 loop, which may explain why H3 and H4 appears longer than H1 and H2 in the low resolution SAXS model.[62] The GIT1 PBD had 11 prolines; all locate in the loops or at the end of helices. The rigid backbone and hydrophobic side chain of proline make the loops less flexible, allowing us to observe extensive NOE connections between the loops and the rest of the protein. The PBD also has three cystines; the chemical shifts of Cβ of all three cystines are less than 32 ppm, indicating that they are in the reduced states.[100] This might be caused by the presence of 5 mM DTT in the NMR sample buffer. In the solution structure, C683 and C725 are closely located in helices H2 and H3 and the distance between the Cα atoms of the two residues is 8.7 Å. A small degree of rotation of the C683 side chain or a slight change in the packing of the helices can bring the two cysteine side chains close enough to form a disulfide bond. Considering the high thermo stability of the PBD, we cannot rule out the possibility of a disulfide bond bridge between the two cysteines.

A sequence comparison of the GIT1 PBD with those of two other four-helix bundle domains – the FAT domain of FAK and the vinculin Vt domain – revealed that the hydrophobic residues in the core regions are well conserved among all the three sequences (Figure 3-1a). The best example of the hydrophobic residues is the L721 in helix H3, wherein the side chain of the residue makes hydrophobic contacts with side chains of L668, I687, V746, and C749 (Figure 3-3). In the structural-based sequence alignment, not only is L721 conserved among the three domains, but the residues that had hydrophobic contacts with this leucine are well conserved too. Although different in surface residues, the three domains share a conserved central hydrophobic network that holds the bundle together. The overall folding of the GIT1 PBD is very similar to that of the FAT domain of FAK (RMSD value of 1.65 Å) and helices H2 to H5 of the vinculin Vt domain (RMSD value of 1.68 Å calculated and averaged on Cα pairs; Figure 3-4). The vinculin Vt domain is a five-helix bundle, with the first helix attached to the groove between helices H2 and H5. The distance between Vt H2 and H5 is slightly more than that between GIT1 H1 and H4, probably because Vt H1 is attached to the groove.

It has been demonstrated that FAK FAT domain features helix H1 swapping dynamics and dimer formation.[91,101] However, it was not determined whether the dimer is of biological importance in the context of full length FAK. Different from FAK FAT domain, no evidence showed similar H1 swapping dynamics or dimerization
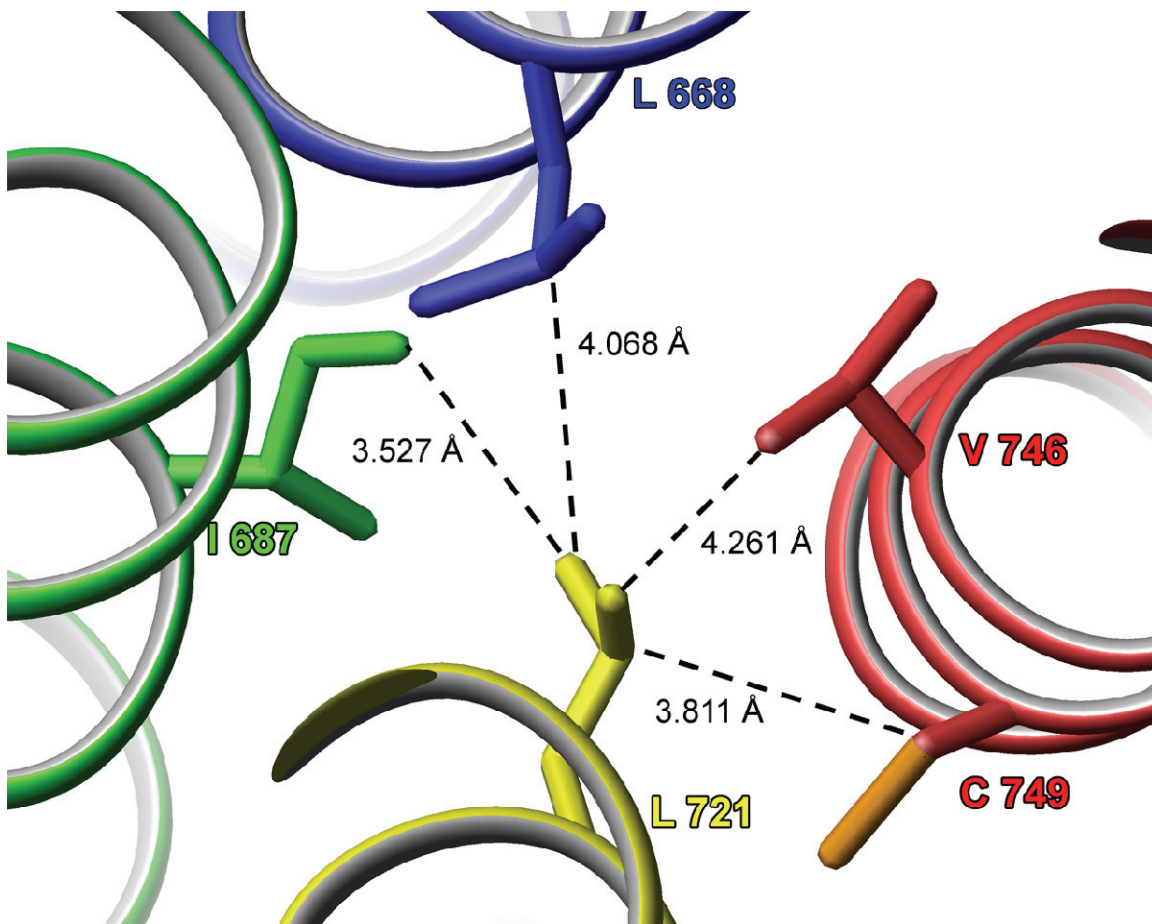
Figure 3-3   The conserved hydrophobic core of the PBD.
The L721 side chain is enclosed by the side chains of L668, I687, V746, and C749. See the Figure 3-1 legend for color coding details.
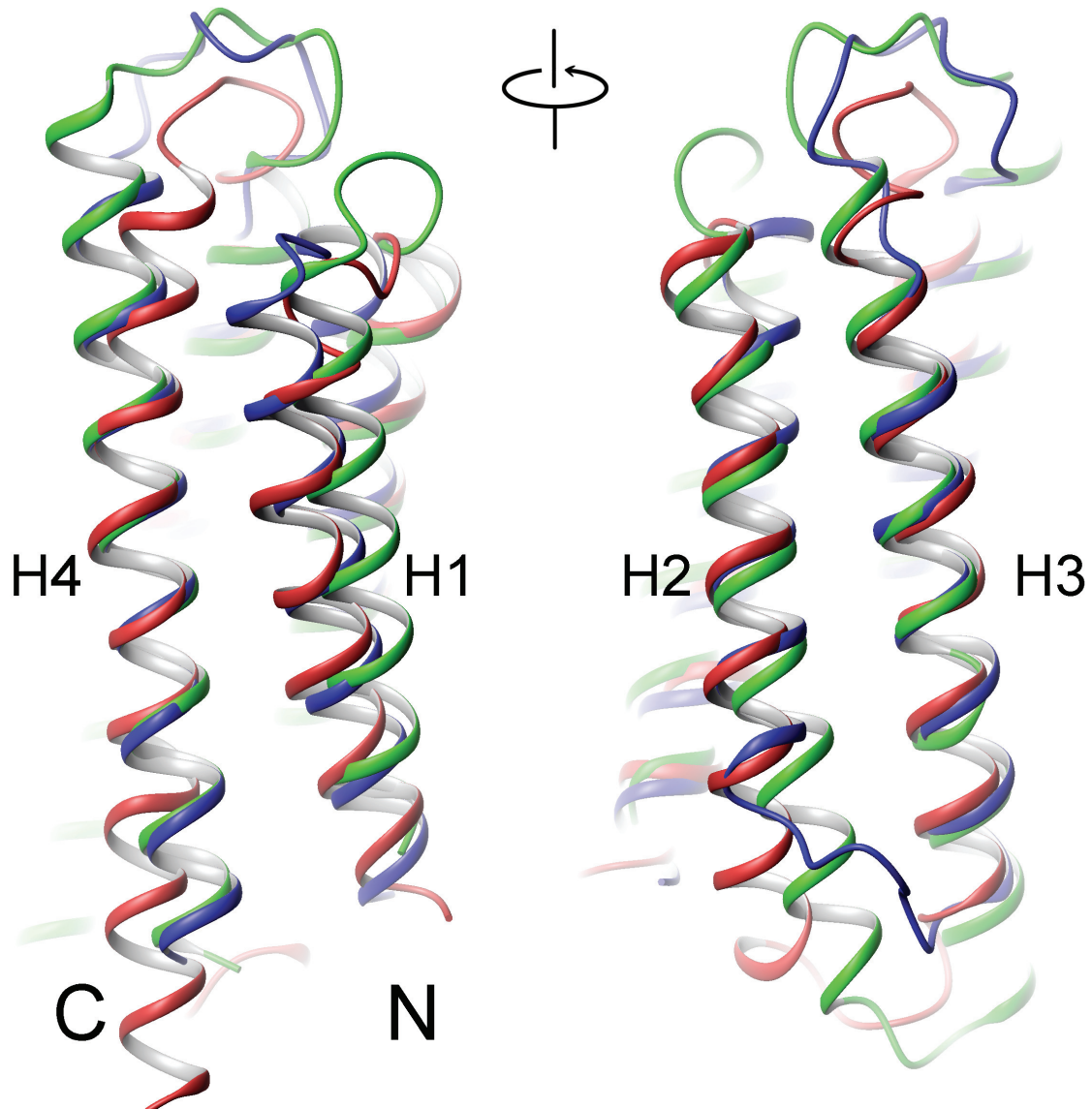
Figure 3-4   Superposition of the structures of GIT1 PBD, FAK FAT, and vinculin Vt domains.

Superimposition of GIT1 PBD, FAK FAT (1KTM), and vinculin Vt domains (H2 to H5) (1OKR) are shown in blue, red, and green respectively. H2, H3, and H4 are relatively well overlaid whereas Vt H2 is more apart from the bundle. Structures are aligned by the Swiss-Pdb viewer and the ribbon represented in MolMol.

phenomenon for GIT1 PBD. The change of sample temperature from 37°C to 13°C resulted in a series of spectra showing all peaks with unchanged intensities. It suggested there is no local line broadening in the protein, including the loop between H1 and H2 (Figure 3-5).   The GIT1 PBD seems to be more resistant to H1 dynamics than FAK FAT domain. The reason might be that the loop between H1 and H2 has a sequence of EFKH for the PBD and PAPP for the FAK FAT domain. The three prolines in the FAT domain make the loop more rigid, probably introducing the conformational strains that cause FAT H1 swapping.

3.3.2   The PBD and the FAT domain are similar on the H1/H4 surface but different on the H2/H3 surface

Forming four amphipathic helices, the GIT1 PBD buries most of its hydrophobic side chains in the bundle core. There is only one major hydrophobic patch on the surface. As shown in Figure 3-6, side chains from H1 and H4 form two positive-charged ridges sandwiching an elongated hydrophobic groove, which covers about two thirds of the bundle length. To the center of the H1/H4 surface, positively charged K663, K755, K758, and polar residue T662, Q666, Y751 enclose a well-defined hydrophobic pocket composed of I665, A754, and A757 at the bottom. These residues are conserved between GIT1 and GIT2 in all organisms for whom sequences are available. Among them, K663, K758, A754, and A757 are also conserved with the FAK FAT domain, with a conserved I665/V936 substitution (Figure 3-1a). This surface makes an ideal binding site for paxillin, as confirmed by chemical shift perturbation titration experiments.

When the GIT1 PBD was compared with the FAT domain of FAK, the latter has two major hydrophobic grooves located at both the H1/H4 and the H2/H3 faces, which correspond to the LD2 and LD4 motif binding sites, respectively (Figure 3-6).[32] Both grooves have a surface setting similar to that of the H1/H4 site of the PBD: a hydrophobic groove with surrounding polar and positive-charged ridges. Superimposition of the structures of the PBD and the FAT domain revealed that K663GIT1/K934FAK and K758GIT1/K1033FAK were in the same position, which is another indication of functional similarity arising from structural similarity. The electrostatic interaction between the side chain of FAK K1033 and the conserved D146 in the LD2 motif might be important in determining binding specificity. Thus, we speculated that the analogous K663 and K758 in GIT1 might also greatly contribute to paxillin binding; this prediction was confirmed by the studies described in the next section.

The FAT surface has two hydrophobic surfaces that serve as LD motif binding sites, but the GIT1 PBD has only one such surface (Figure 3-6). In the FAT domain, helices H2 and H3 form a second hydrophobic surface, and at the center of the FAT H2/H3 site, the bottom of the hydrophobic pocket is composed of G959 and A996 on H2 and H3, respectively. However, the equivalent position on GIT1 PBD is occupied by bulky H688 and Y719 which effectively close the pocket. Moreover, the positively charged H2 ridge in the FAT domain is also eliminated by the substitutions of K956 and R963 in FAK by E685 and T692 in GIT1. Interestingly, we found that the vinculin Vt

55
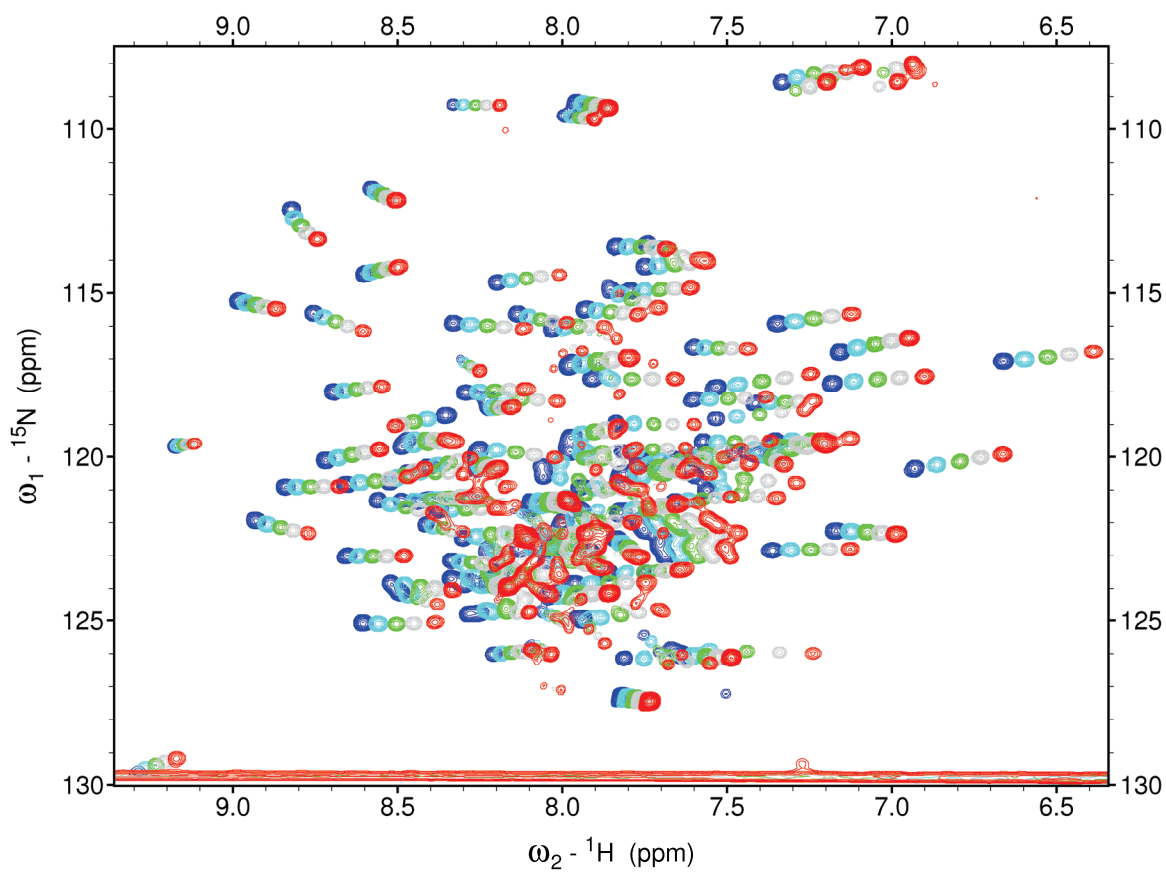
Figure 3-5 **^{15}N HSQC spectra of the GIT1 PBD measured at different temperatures.**
Blue: 37°C; cyan: 31°C; green: 25°C; gray: 19°C; red: 13°C. All peaks in the spectra showed similar intensities at different temperatures, indicating that there was no floppy region in the PBD.
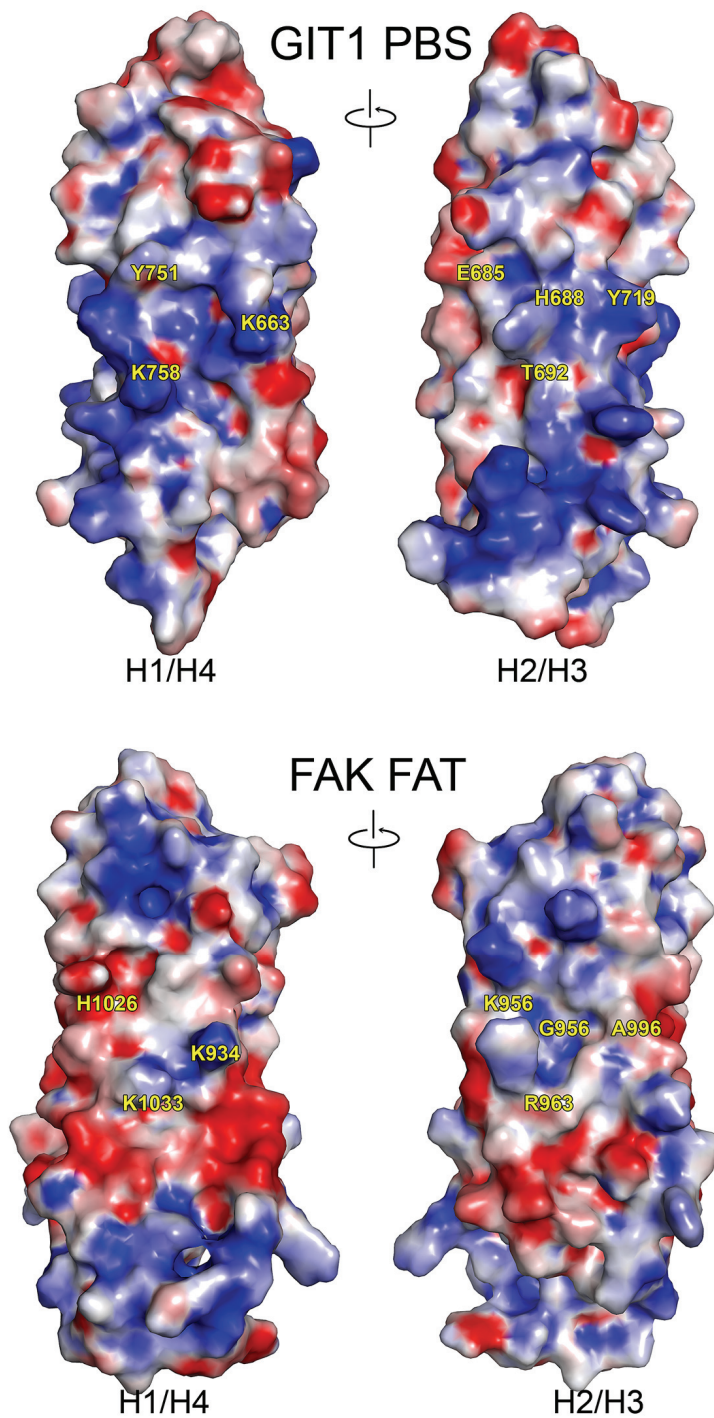
Figure 3-6   Electrostatic surfaces of PBD and FAT domains.
Positive charge is shown in blue and negative charge in red. Equivalent residues are marked in both structures. The figure was generated by Pymol, based on the electrostatic potential generated by GRASP.

domain is more similar to FAK; for example, FAK K956 is conserved as K1021 in the Vt domain, whereas FAK G959 is substituted with a small residue A1024 in Vt. This suggests a possible paxillin-binding site on the H2/H3 surface of the vinculin Vt domain.

3.3.3 Paxillin LD4 motif binds to the PBD at the H1/H4 face

The binding between the GIT1 PBD and paxillin is essential to localize the PAK–PIX–GIT complex to the focal adhesion.[60] Although the interaction between the paxillin LD4 motif and GIT proteins had been proposed,[12,38,61] the details about the binding remained unknown. Therefore, we analyzed the interaction between the GIT1 PBD and the paxillin LD4 motif by chemical shift perturbation experiments. In the experiments, we used a peptide comprising the LD4 motif of paxillin (residues 261–282, designated LD4 peptide, Figure 3-7c) to titrate the ^{15}N-labeled PBD, and found that the LD4 peptide binds to the PBD on the H1/H4 surface. The binding site locates to the central region of the H1 and H4 solvent-exposed surface, corresponding to the two clusters of largely shifted peaks in the perturbation plot (Figure 3-8a). Although the two clusters seem to be remote in the sequence, they are actually in close contact in the 3D structure, forming the two ridges and the hydrophobic pocket mentioned previously (Figure 3-6a).

Most of the shifted peaks affected by the bound LD4 peptide are in the intermediate exchange time scale. Additionally, the cluster in H4 is higher than that in H1 in average, indicating a slightly stronger association with H4 than H1 (Figure 3-8a). The solvent-exposed T662, K663, Q748, Y751, D752, K755, and K758 exhibited significant perturbations (K633, Y751, K758 had the largest perturbations as shown in Figure 3-7a and d). The residues between them, which are buried between the helices, showed only moderate changes. This indicates that the large shift is not due to a binding-induced conformational change. The cluster in H1 corresponds to the paxillin binding sequence 2 (PBS2) proposed by sequence analysis.[38] Compared with H1/H4, residues in the H2 and H3 had only minimal or moderate shifts on LD4 peptide titration. A closer inspection of the shifted peaks in H2 and H3 showed that most peaks are residues at the interface with H1 or H4, and among those, buried residues showed more perturbation than exposed ones, indicating that the perturbation was due to proximity to the H1/H4 binding site. Similarly, recent mutagenesis data showed that mutations on H2/H3 did not influence paxillin binding, while mutations on H1/H4 decreased paxillin binding, especially the K663 and K758 mutations substantially decreased paxillin association.[62]

3.3.4 The effect of paxillin S272 phosphorylation on the binding of GIT1 PBD

It was reported that paxillin S272 (S273 in the chicken sequence) phosphorylation promoted GIT1 focal adhesion localization in cells,[89] and so we wanted to investigate the interaction with GIT1 and phosphorylated paxillin. We studied the interaction between purified GIT1 PBD and synthesized LD4 peptide with S272 phosphorylation by NMR chemical shift perturbation experiments. Different from the previous ITC data,[62] we found the GIT1 PBD spectra titrated with phosphorylated LD4 peptide (LD4p) was very
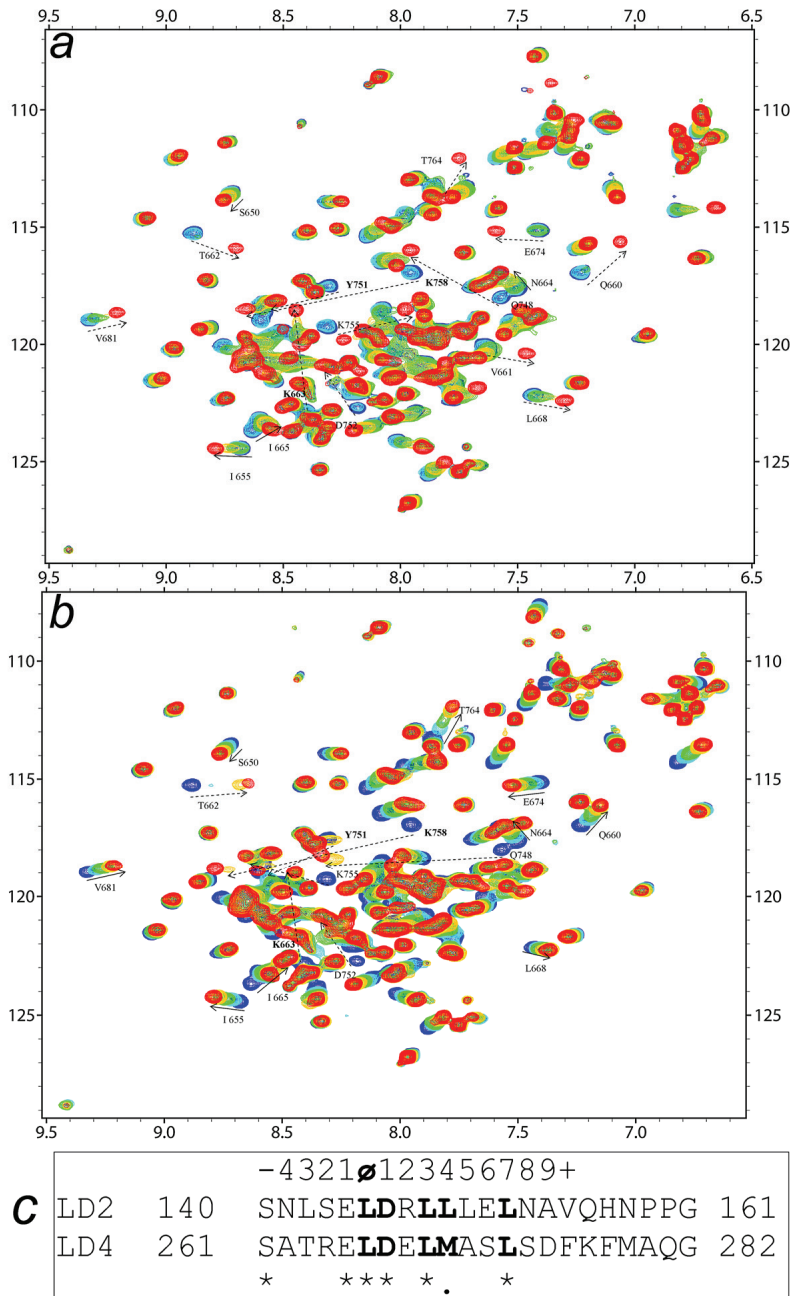
Figure 3-7   The interaction between the PBD and LD peptides.
*(a,b) Overlay of GIT1 PBD spectra titrated with LD4 and LD2 peptides, respectively
(Blue: GIT1 PBD alone; Cyan: adding 0.5 eq. of ligand; Green: adding 1 eq. of ligand;
Gold: adding 2 eq. of ligand; Red: adding 5 eq. of ligand. Arrows mark significantly
perturbed resonance and dashed arrows mark perturbation with intermediate exchange
rate. (c) Sequence alignment of human LD2 and LD4 peptides used in this study.
Identical residues are marked with asterisks and the LDXLLXXL motif is showing in bold.
Conventional nomenclature is used; the first L is Φ0 and other residues are named as in
the figure.*
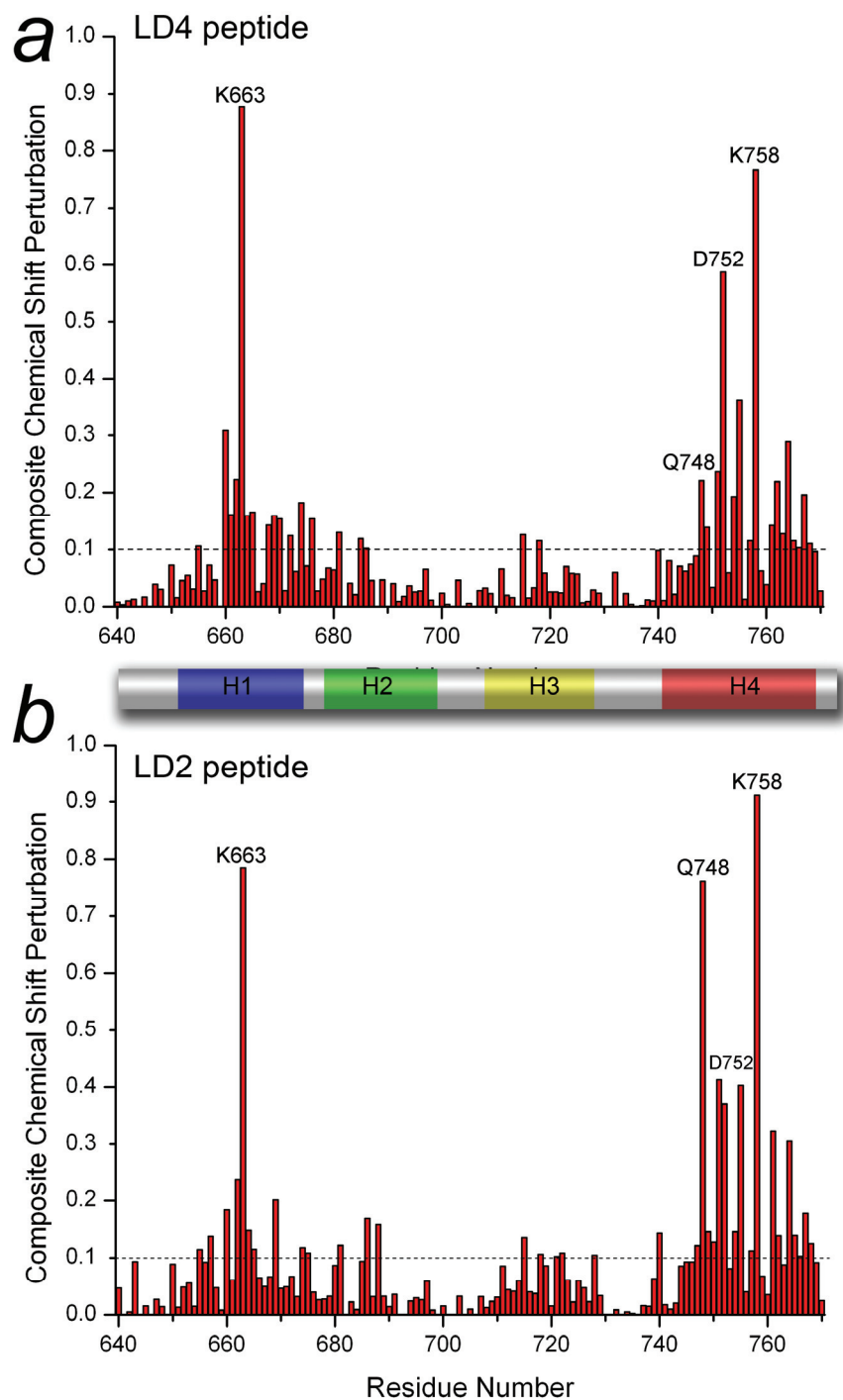
Figure 3-8   Analysis of the interaction between the PBD and LD peptides.
(a,b) Composite chemical shift perturbation plotted against residue number for LD4 and LD2, respectively. Composite chemical shift perturbation was calculated by the equation $\Delta\delta_{composite} = (\Delta\delta_{NH}^2 + \Delta\delta_N^2 /25)^{\frac{1}{2}}$. K663, K758, Y751 and Q748 are labeled and they are among those dramatically shifted peaks.

similar to that of the PBD titrated with the same ratio of LD4 peptide (Figure 3-9), indicating that the two peptides bound to the PBD with similar affinities. Superimposing the ^{15}N-HSQC spectra of the PBD mixed with 5 equivalents of the two peptides revealed that most perturbed peaks could be overlaid (Figure 3-10a), suggesting the two peptides bind to the PBD in similar fashions. However, there were clear differences between the two spectra. For example, comparison with the peaks in the spectrum of apo-PBD showed that some peaks were perturbed more by the LD4 binding (Figure 3-10b and e), while others were perturbed more by the LD4p peptide (Figure 3-10d). In addition, the two peptides also shifted some peaks in different directions (Figure 3-10c and f). In both cases, neither peptide produced a consistently larger chemical shift perturbation throughout the PBD binding site than the other, further indicating that the two peptides, LD4 and LD4p, have similar binding affinities for the PBD. These observations suggest that unlike the interaction between the FAT domain of FAK and LD4p,[45] in the complex of PBD bound to LD4p peptide, the phosphate group in the peptide has some contacts with the PBD, and those contacts might compensate for the entropy lost due to the phosphorylation.[45]

3.3.5   *The binding between the paxillin LD2 peptide and the GIT1 PBD*

Because the only hydrophobic patch on the GIT1 PBD corresponds to the LD2 motif binding site on the FAT domain of FAK, we investigated whether the LD2 motif also participates in GIT1 binding. Synthesized LD2 peptide (human paxillin 140-161, Figure 3-7c) was titrated into uniformly ^{15}N-labeled GIT1 PBD. Surprisingly, we found that the LD2 peptide also dramatically perturbed the PBD spectrum (Figure 3-7b). The LD2 and the LD4 binding perturbed the same set of residues and they shifted in similar directions and distances (Figure 3-7). Like LD4 peptide, when the LD2 peptide was titrated into the solution of GIT1 PBD, we found that while some peaks remained at the same position, many peaks shifted during the titration; some peaks had large shifts and some even showed intermediate exchange phenomena. Thus, the LD2 peptide perturbed the GIT1 PBD spectrum to a similar extent as the LD4 peptide. Additionally, both peptides perturbed the same set of peaks, and their binding sites overlapped on the H1/H4 surface (Figure 3-6a).

To determine how the binding occurs with both LD motifs present, we performed competitive titration experiments, in which up to 5 equivalents of LD2 peptide was titrated into ^{15}N-labeled PBD in the presence of 5 equivalents of LD4 peptide and vice versa. The competitive experiments agreed with the single-peptide titration data, confirming that each peptide was able to further perturb the spectrum in the presence of equal concentration of the other peptide (Figure 3-11). The two final spectra overlaid well, indicating that the binding reaches equilibrium regardless of the order in which the peptides were added. It is likely that both peptides compete for the same binding site with similar affinity. The concentration of available LD motifs may dynamically determine the equilibrium state.
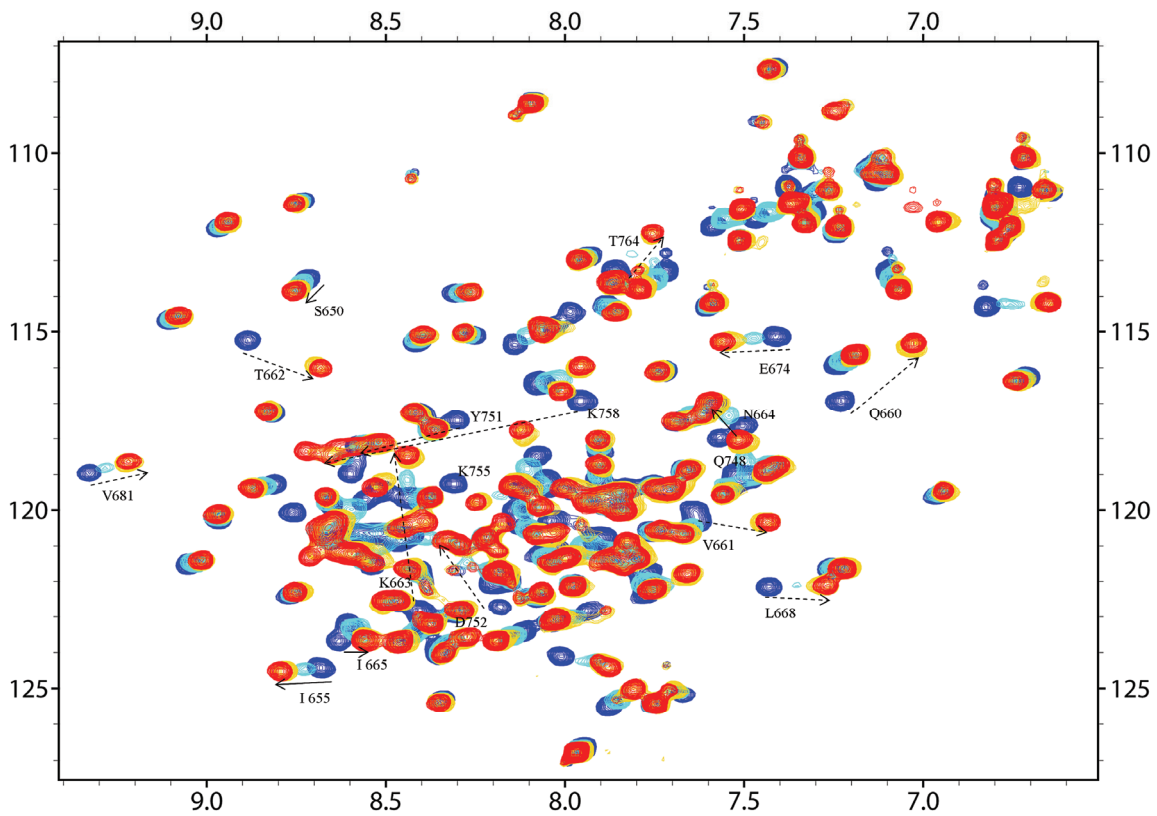
Figure 3-9   Binding of the GIT1 PBD to the LD4p peptide.
Overlay of GIT1 PBD spectra titrated with the LD4(S272)p peptide. Blue: GIT1 PBD alone; Cyan: adding 0.5 eq. of ligand; Gold: adding 2 eq. of ligand; Red: adding 5 eq. of ligand). The result was very similar to the LD4 peptide titration, indicating the S272 phosphorylation does not greatly influence the binding.
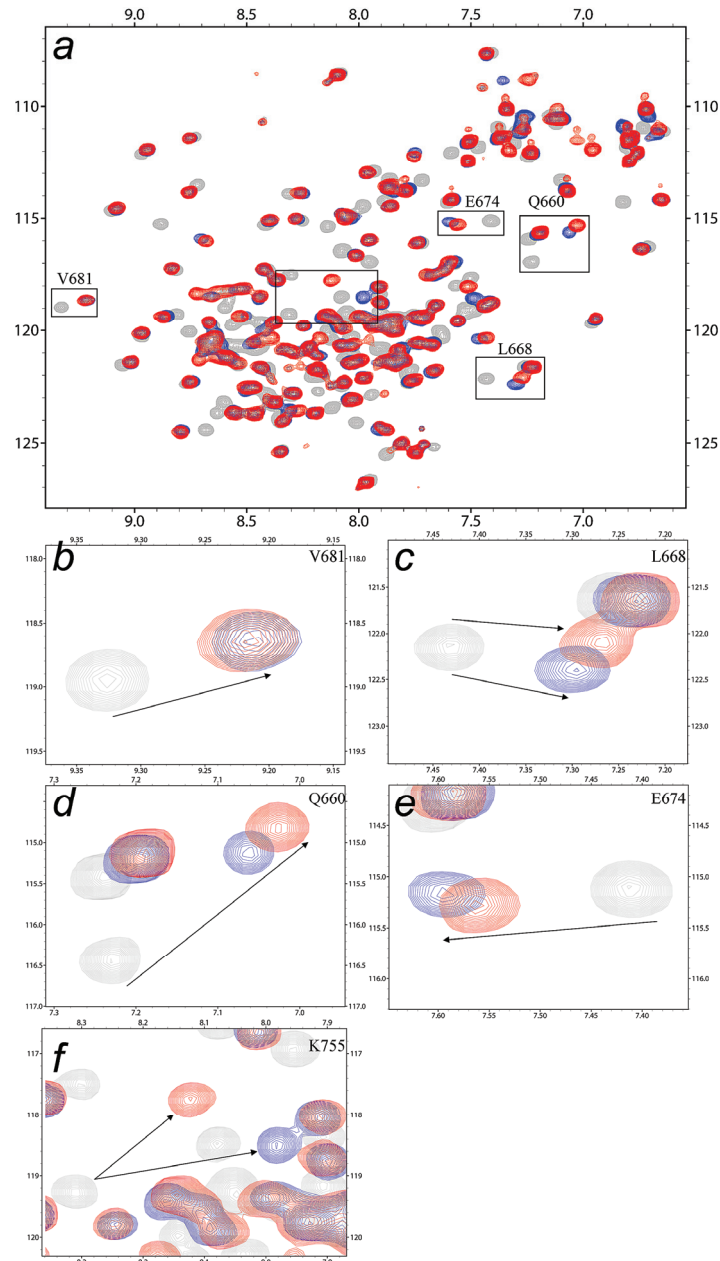
Figure 3-10　Comparison of PBD spectra titrated with 5 eq. of LD4 or LD4p peptides.
(a) Full spectra overlay. Gray: GIT1 alone; Blue:GIT1 PBD+LD4; Red:GIT1 PBD+LD4p. While most peaks in the spectra of PBD bound with the two peptides could be overlaid well, some peaks exhibited difference. (b)-(f) examples of different peaks in the spectra of PBD bound with the two peptides: (b) like the example showed, the bindings of the two peptides had similar effects to the most peaks in the spectra of PBD; (c) and (f) some peaks moved to different directions due to the bindings; (d) and (e) other peaks moved to the same direction, but the magnitudes of shift were different. Some shifted more due to the LD4 peptide binding while others shifted more due to the LD4p binding.
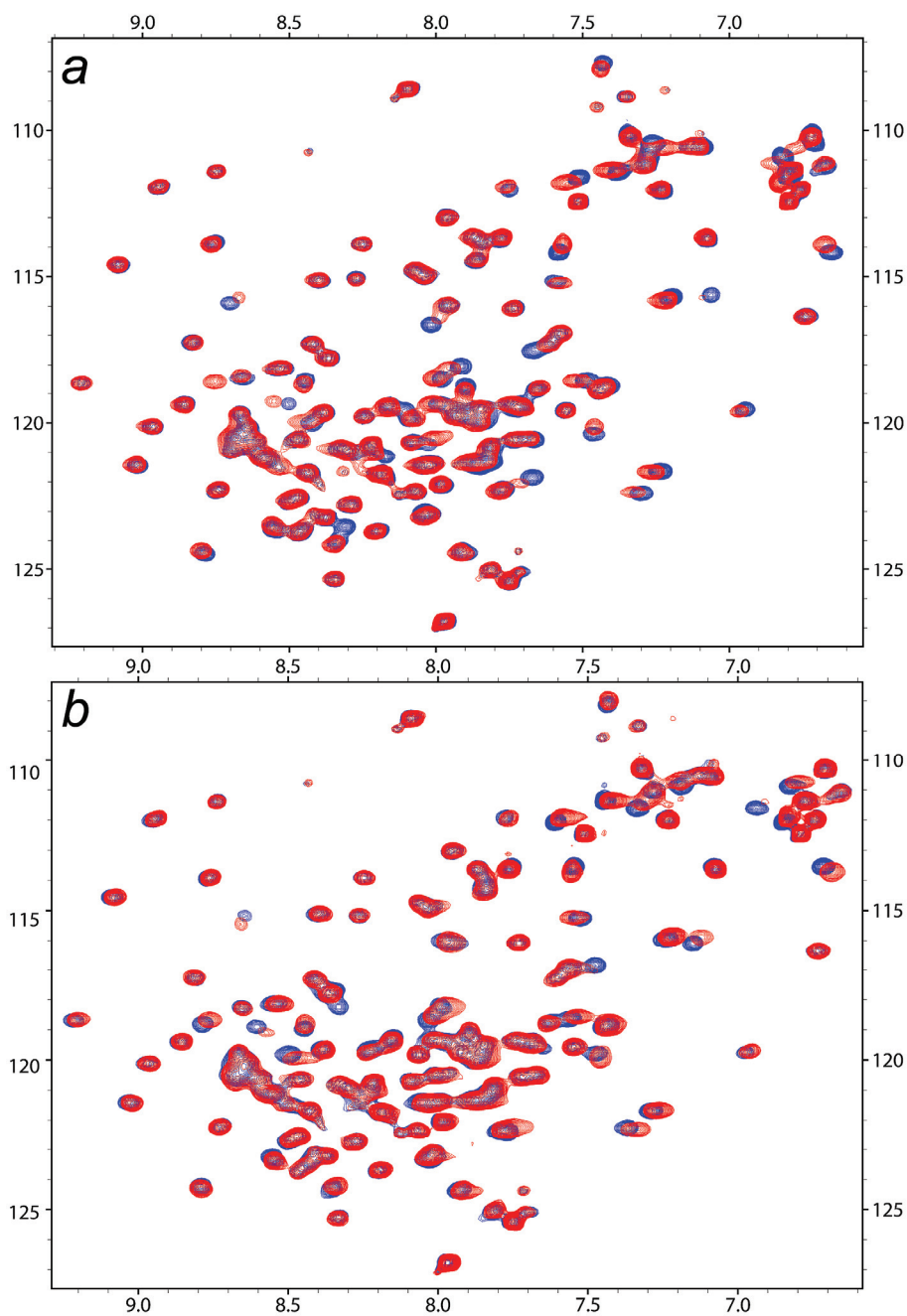
Figure 3-11   Competitive binding study of the GIT1 PBD to LD2 or LD4 peptide.
(a) ^{15}N HSQC spectra of GIT1 PBD with 5 eq. of the LD4 peptide (blue) and of the same sample with another 5 eq. of the LD2 peptides were added (red). (b) ^{15}N HSQC spectra of GIT1 PBD with 5 eq. of the LD2 peptide (blue) and of the same sample with another 5 eq. of the LD2 peptides were addes (red). The titration data showed that each peptide could further perturb the spectrum of the PBD in the presence of the same concentration of the other peptide, indicating that the two peptides had similar binding affinities to the GIT1 PBD.

To test the binding between PBD and paxillin, we titrated the PBD with up to 5 equivalents of chicken paxillin 133-288, which contains both LD2 and LD4 motifs. The spectrum of PBD titrated with 5 equivalents of paxillin 133-288 overlaid well with the PBD spectrum titrated with 5 equivalents of both LD2 and LD4 peptides, exhibiting an average of binding from both LD motifs (Figure 3-12). It also showed that the LD3 motif, which was also present in the paxillin construct, does not interact with GIT1 PBD.

3.3.6 *Examining the interactions between LD peptides and GIT1 PBD by Biacore binding studies*

To further characterize the binding properties of the two paxillin peptides to the PBD, we measured the binding affinities between the LD peptides and the PBD by the Biacore binding assay. In the experiments, the LD peptides were synthesized with an N-terminal biotin tag connected with a flexible GGSG linker. The peptides were immobilized on a NeutrAvidin coated chips and run through the purified GIT1 PBD protein. As summarized in Table 3-2, the PBD binds to the LD4 peptide with a Kd of 7.2 µM, which is close to the reported Kd of 10 µM.[62] The LD2 peptide showed a slightly higher Kd of 25.1 µM. Because the LD2 peptide binds to GIT1's PBD with a Kd only about three times weaker than the LD4 peptide, the two peptides should be able to compete for the ligand binding site on the surface of the PBD, and indeed this conclusion is in agreement with our NMR experiments. Also consistent with our NMR studies, the LD4p peptide exhibited only a slight increase in Kd value (Kd=10.2 µM).

3.4 Discussion

The solution structure of the GIT1 PBD is a stable four-helix bundle, which is similar to the structures of the FAT domain of FAK and the vinculin Vt domain. The structure was validated by the solvent exposure surface mapped by the H/D exchange experiment. The PBD forms an up-and-down helix bundle with an average axis angle of 19.1º, which is close to the ideal antiparallel helix packing angle of 20º.[102] The PBD, the FAT domain of FAK and the vinculin Vt domain share a well conserved hydrophobic core, but their surface residues exhibit more diversity. This explains the very similar overall folding of the three domains as well as their distinct binding specificity.

Both the PBD and the FAT domain interact with the LD2 and LD4 motifs of paxillin and are responsible for targeting GIT1 and FAK to focal adhesions. However, there are distinct differences between the two domains. First, in solution, the FAT domain forms a dimer by swapping the H1 between the two protomers,[91] whereas the PBD is a well-folded monomer. Second, the LD2 and LD4 motifs of paxillin interact with the FAT domain simultaneously at two opposite faces of the four-helix bundle,[44] whereas the LD2 and LD4 motifs of paxillin bind to the PBD at the same site, the H1 and H4 site. These functional differences between the FAT domain and the PBD are rooted in the difference of their sequences.
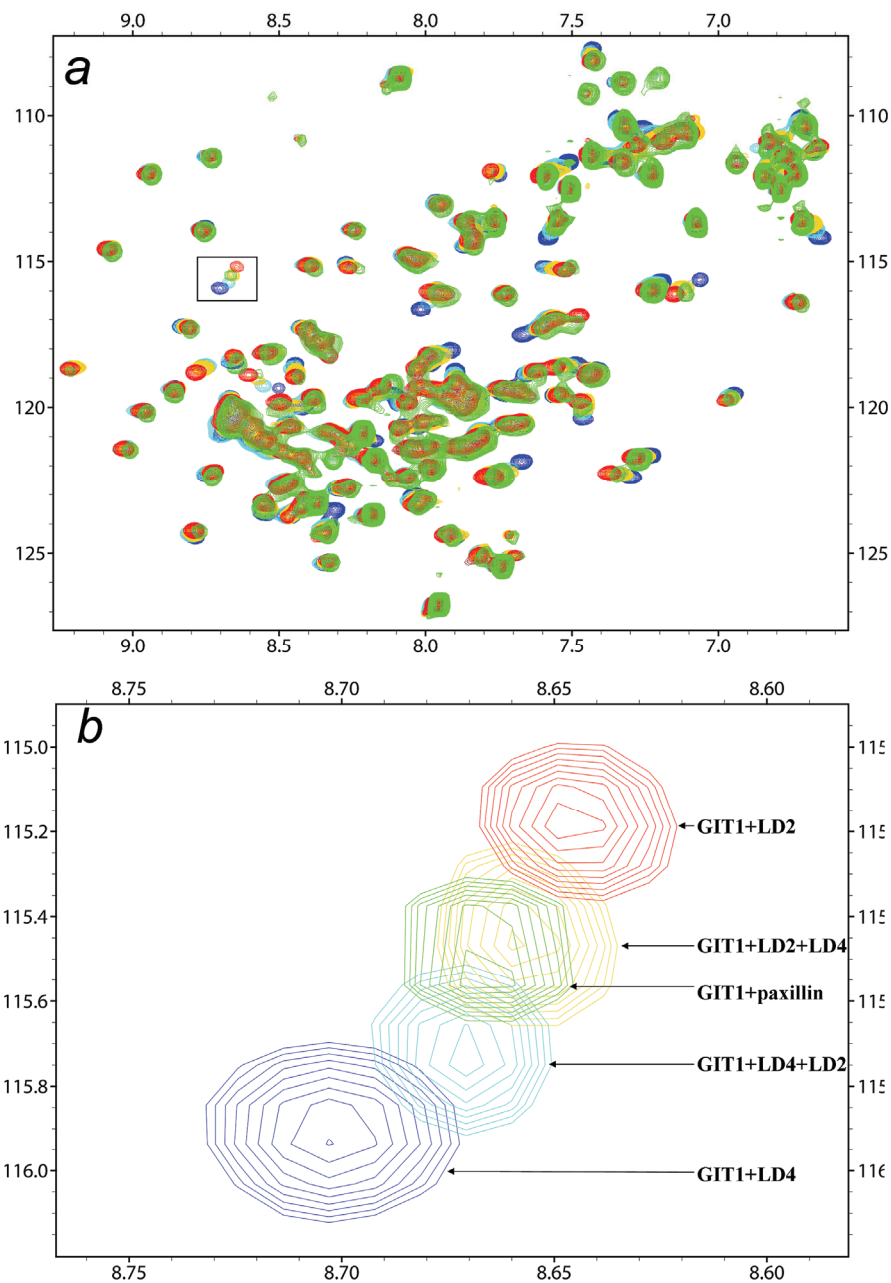
Figure 3-12   Spectrum of GIT1 PBD titrated with paxillin 133-288.
Overlay of the spectra of the GIT1 PBD in the presences of 5 eq. of paxillin 133-288 and 5 eq. of both the LD2 and LD4 peptides respectively. Blue: GIT1+LD4; Cyan: GIT1+LD4+LD2 (in order); Gold: GIT1+LD2+LD4 (in order); Red: GIT1+LD2; Green: GIT1+paxillin 133-288. (a) The whole spectrum of GIT1. The black box showed the peak of T662, which was magnified in (b). The paxillin 133-288 titration shifted the peaks of PBD to the average positions of the spectra of the PBD bound to the LD2 and LD4 peptide respectively.

Table 3-2   Binding affinities between GIT1 PBD and LD peptides measured by Biacore binding assay.

Interaction	K_D (µM)	Rmax (RU)
GIT + LD2	25.1 ($\pm$0.2)	91.0 ($\pm$0.2)
GIT + LD4	7.0 ($\pm$0.2)	53.3 ($\pm$0.4)
GIT + LD4p	9.4 ($\pm$0.2)	87.7 ($\pm$0.4)

The previously reported interaction between the LD4 motif of paxillin and the GIT1 PBD was confirmed in our studies.[12,38,61] The binding site mapped by NMR was also in agreement with the recent mutation study,[62] in which all residues critical for LD4 binding exhibited strong chemical shift perturbation. The LD4 motif binds to the H1 and H4 site of the PBD, unlike the complex of LD4 with the FAT domain of FAK, where it binds to the H2 and H3 surface. This variation may also explain another difference between the two complexes: while S272 phosphorylation within the LD4 motif decreases the LD4 binding to the FAT domain,[45] the LD4p peptide interacts with the PBD similarly to the LD4 peptide.

In this study, the LD4 S272 phosphorylation interacted with PBD similarly to LD4 peptide. The NMR showed that the spectrum of PBD titrated with LD4 S272p peptide overlaid well with the spectrum titrated with LD4 peptide. The Biacore also showed that there is only a slight decrease in binding affinity by S272 phosphorylation. This result is different from the recent report, in which S272 phosphorylation decreased LD4 binding by 12 folds.[62] The reason for the difference is still under investigation. One possibility we noticed is that the author might have used a shorter LD4 peptide, since we found that GIT1 PBD binds very weakly to the phosphorylated LD4 peptide if it ends at the D275 (Figure 3-13).

It has been proposed that paxillin S272 phosphorylation is an important event in focal adhesion regulation.[12,45,89] In a working model, FAK and GIT1 compete for paxillin binding, and, in turn, the two proteins compete for FA localization. When paxillin S272 is dephosphorylated, paxillin prefers binding to FAK, because FAK binds both the paxillin LD2 and LD4 motifs cooperatively with a higher affinity. Once S272 is phosphorylated, the affinity between LD4 and FAK is weakened, causing the disassociation of FAK–LD4.[45] However, the phosphorylation of S272 in LD4 does not affect the interaction between paxillin and GIT1. Therefore, paxillin S272 phosphorylation shifts the balance of the paxillin interaction with FAK versus GIT1 in favor of GIT1 and causes an increase in the paxillin–GIT1 complex. This would decrease FAK while increasing PAK (through GIT1) in the FAs. Since FAK has been linked to FA assembly and PAK can promote FA disassembly,[12,45,89] this balance shift may promote regulated FA assembly and the disassembly cycle.

The FAT domain of FAK binds to paxillin by interacting with both the paxillin LD2 and LD4 motifs simultaneously using its H1/H4 and H2/H3 sites.[44,103] With two binding sites, FAT should have a much stronger binding affinity with paxillin than the GIT1 PBD since the PBD has only one LD binding site, the H1/H4 site. This concern can be alleviated by another novel finding in this work – the paxillin LD2 motif also interacted with the GIT1 PBD. Since the LD2 motif binds to the PDB as well, GIT1 might achieve a binding affinity comparable to FAK by interacting with both the LD2 and LD4 motifs of paxillin. Indeed, in our studies, we not only showed that the LD2 motif of paxillin could bind to the GIT1 PBD, but in one experiment we also found that the spectrum of the PBD bound with paxillin 133-288, which contained both the LD2 and LD4 motifs, was an average of the spectra of the PBD bound with the LD2 and LD4 peptides (Figure 3-12). Furthermore, such observation did not change even when the
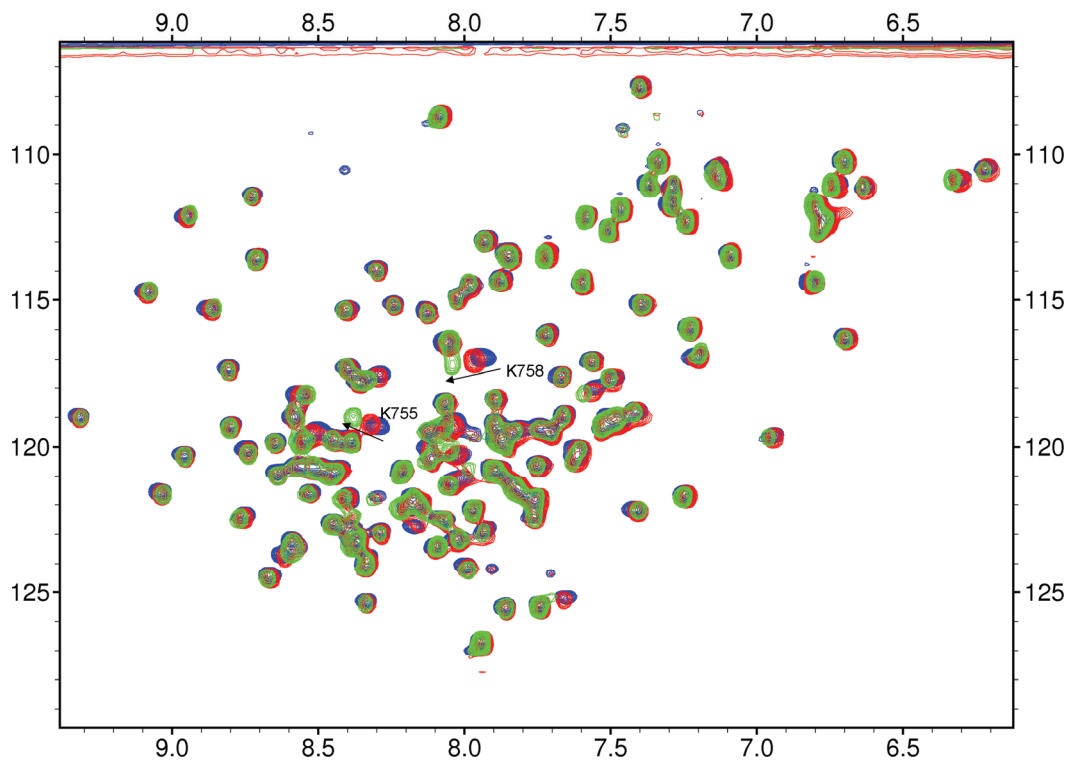
Figure 3-13   GIT1 PBD spectra titrated with shorter LD4p peptide.
PBD titrated with up to 10 eq. of the short LD4p peptide. The peptide is 7 residues short at the C terminus and has a sequence of SATRELDELMASpLSD. Blue: GIT1 PBD alone; Red: adding 2 eq. of ligand; Green: adding 10 eq. of ligand. Different from the longer LD4p, the shorter LD4p bound to PBD very weakly, with only small perturbation at K755 and K758.

paxillin concentration was increased to five times that of the PBD, indicating that the two LD motifs have similar affinity to the PBD and bind GIT1 simultaneously. In other words, it suggests that one paxillin can bind two GIT1 proteins. In consistent with the observation, GIT1 can dimerize in solution through its central coiled-coil domain.[104] Considering the length of the flexible linker between the LD2 and LD4 motifs, it is possible that two GIT1 proteins in a dimer can bind to one paxillin molecule simultaneously, through both the LD2 and LD4 motifs. Indeed, GIT1 is found primarily as an oligomer together with PIX,[59,104] and the oligomerization is essential for GIT1 FA localization.[105,106]

CHAPTER 4. AN NMR STUDY OF GIT1-PAXILLIN INTERACTION AND THE DATA DRIVING COMPLEX STRUCTURE MODELING

4.1   Introduction

Cell attachment and migration are critical for embryogenesis, tissue repair, inflammatory response, tumor invasion and metastasis.[1] Cell peripheral protein complex focal adhesion (FA) plays central role in cell attachment and migration. FA mechanically couples cell membrane attachments with cytoskeletons and integrates signals in controlling the assembly and disassembly of adhesions. The temporal and spatial distribution of FA is regulated by multiple proteins, e.g. Focal Adhesion Kinase (FAK), G Protein Coupled Receptor (GPCR) Kinase (GRK) Interacting proteins (GITs), p21-activated kinase (PAK), paxillin and small GTPases.[52,83-88]

GIT family proteins are multi-domain adaptors that consistent of an N terminal Arf GTPase-activating protein (Arf-GAP) domain, a central Spa2-homology domain (SHD), and a C terminal paxillin-binding domain (PBD). GIT SHD interacts with PAK through the p21-activated kinase interacting exchange factor (PIX).[53,56] GIT and PIX can dimerize by themselves and oligomerize to form high molecular weigh complexes.[12,59,61,104,106,107] GIT PBD binds to paxillin and recruits the GIT complex into FA.[12,38,54,60] GIT1 PBD forms a four-helix bundle that resembles the structure of FAK focal adhesion targeting domain (FAT).[62,108] It interacts with both paxillin LD2 and LD4 motifs through the same binding site on the H1 and H4 surface.[108] Although the GIT1 PBD structure and its interaction with paxillin were characterized,[108] the atomic complex structure remained undetermined.

Paxillin is a scaffold protein that binds multiple targets through its five N terminal LDXLLXXL motifs (LD motifs) and four C terminal LIM domains.[90,109] Previous studies indicated the LD4 motif was the GIT1 binding site,[12,38] but we demonstrated that both LD2 and LD4 motifs can bind GIT1 with comparable binding affinities.[108] Although the GIT1 titration experiment with unlabeled paxillin 133-288 suggested a simultaneous interaction between paxillin LD2, LD4 motifs and a PBD dimer,[108] there was no study directly observing the binding of both LD motifs.

There are three splice variants for human paxillin. The only sequence difference is in the central region right after the LD4 motif. Isoform β and γ differ from the isoform α with an insertion of 34 or 48 residues after the K277.   It has been shown that the splice isoforms bind differently to FA proteins, for example, isoform β binds stronger to FAK than to vinculin; however isoform γ binds stronger to vinculin than to FAK.[35]   Since the binding to FAK or vinculin are mediated by the LD4 motif, the insertion might help determine the specificities to FAK or vinculin. One hypothesis is that the different sequences after the LD4 motif may influence the interaction between the LD4 motif and its binding partners, which may play a regulatory role. In our recent study, we demonstrated that the GIT1 PBD was able to interact with paxillin isoform α LD4 peptide (human paxillin α 261-282). Since the splice isoforms are different after K277,

their interactions to GIT1 PBD are still unknown for the isoform β and γ. The function of the splice variants right after the LD4 motif remains unidentified.

To understand the function of the paxillin splice isoforms, we tested GIT1 PBD interaction with different LD4 peptides that derived from the isoforms. We also tested GIT1-paxillin binding by directly observing the ^{15}N labeled paxillin 133-288 construct. The complex structures determined by data driving HADDOCK modeling reconciled existing data and explained the binding specificities of paxillin-GIT1 interaction versus paxillin-FAK interaction.

4.2 Materials and methods

4.2.1 Protein purification and peptide synthesis

^{15}N labeled rat GIT1 PBD (640-770) was expressed and purified as described before.[108] The sample condition for all the NMR titration experiments were: 400 uM of purified GIT1 PBD in the solution of 20mM potassium phosphate (pH 6.5), 5 mM 1,4-dithiothreitol (DTT), 5 mM EDTA and 5% (v/v) D_2O. Chicken paxillin 133-288, which covered both the LD2 motif and the LD4 motif, was expressed with pET28b vector in *E. coli* strain BL21(DE3). It was purified with the same Ni-affinity chromatography protocol and further purified with HPLC. The paxillin NMR sample was prepared in the same buffer condition: 20mM potassium phosphate (pH 6.5), 5 mM 1,4-dithiothreitol (DTT), 5 mM EDTA and 5% (v/v) D_2O. All the human LD peptides were chemically synthesized and purified by HPLC at the Hartwell Center of Bioinformatics and Biotechnology, St. Jude Children's Research Hospital, Memphis, TN. The corresponding sequences were presented in Table 4-1.

4.2.2 NMR chemical shift perturbation titration experiments

All the titration experiments were performed at 25 °C on either a Bruker Avance 600 MHz NMR spectrometer or a Bruker Avance 800 MHz NMR spectrometer, both equipped with cryoprobes. All the peptides were dissolved in the same buffer as the protein: 20mM potassium phosphate (pH 6.5), 5 mM 1,4-dithiothreitol (DTT), 5 mM EDTA and 5% (v/v) D_2O. The pH was adjusted back to 6.5 before titration. For the natural abundance ^{15}N HSQC of the LD peptides, unlabeled LD2 peptide and LD4_S1 peptide (used LD4_S1 instead of LD4α for solubility reason) were dissolved in the above buffer with final concentration of 8mM. 1024 scans on a Bruker Avance 600 MHz NMR spectrometer yielded good signal to noise ratio. All the NMR spectra were processed with NMRPipe and analyzed with Sparky.[94,95] Structural analysis and all the structure figures in this paper were generated with MolMol and Pymol.[78,97]

Table 4-1   The peptide used in the study and Kd value derived from Biacore SPR.

Peptide	Sequence	Kd by Biacore SPR (µm)	Rmax (RU)
LD4α	SATRELDELMASLSDFKFMAQG	7.0 ± 0.2	53.3 ± 0.4
LD4β	SATRELDELMASLSDFKIQDLE	6.5 ± 0.2	50.3 ± 0.5
LD4γ	SATRELDELMASLSDFKGSWPL	6.5 ± 0.1	78.0 ± 0.3
LD4_S1	SATRELDELMASLSDFK	12.8 ± 0.2	28.4 ± 0.1
LD4_S2	SATRELDELMASLSD	Not detectable	/
LD4_S2p	SATRELDELMAJLSD	Not detectable	/

4.2.3 Biacore surface plasmon resonance (SPR) binding assay

LD peptides used in this study were the same sequence to the ones for NMR CSP titrations except that a GGSG linker joined a biotin tag at the N terminus. The peptides were attached to a NeutrAvidin-covered (Pierce) gold surface (CM4 chip; Biacore, GE Healthcare, UK). Kinetic studies were performed at 20°C with a BIACORE 3000 surface plasmon resonance instrument (Biacore, GE Healthcare, UK). The purified GIT1 protein (in 20 mM potassium phosphate buffer, pH 6.5, 150 mM NaCl, 5 mM EDTA, 0.1 mg/mL BSA, and 0.005% P20 surfactant) was inject to flow through the chip and the K_d was derived by fitting the data from three injections. The association and dissociation were observed at a flow rate of 50 μL/min in a concentration range of 1.2–100 μM. Binding affinities were determined by the program Scrubber 2 (Version 2.0b, BioLogic Software). Triplicate injections were made and the average was reported.

4.2.4 Circular dichroism spectroscopy

All CD spectra were obtained with an Aviv 62DS CD spectrometer (Aviv, Lakewood, NJ). The peptide samples were prepared so that the final concentrations were about 50 μM in a solution of 10mM potassium phosphate at pH 6.5. Experiments were carried at 25 °C and a blank was used to eliminate solution influence.

4.2.5 Complex structure modeling

Complex structures were modeled with the software HADDOCK V1.3.[77] The top 20 GIT1 PBD structures with lowest target function were taken for complex modeling. The structure of LD2 and LD4 peptides were made as an ideal α-helix by Swiss-Pdbviewer,[99] according to the CD data and the LD2 structure solved before.[31] The peptide structures were further energy minimized by 10000 steps of conjugated gradient method, in Sybyl V7.3 (Tripos, St Louis, MO). The AIR restraint files were generated according to the previous LD2 and LD4 peptide titration data.[108] The docking was performed on a Linux cluster with 16 CPUs. The top 200 best models were clustered and the complex with best intermolecular energy was chosen as our model for further analysis.

4.3 Results

4.3.1 Comparison of paxillin splice variants for GIT1 PBD interaction.

Using magnetic resonance spectroscopy (NMR) and Biacore binding assay, we tested the interactions between GIT1 PBD and three LD4 peptides derived from the splice isoforms (For sequences refer to Table 4-1).

NMR chemical shift perturbation (CSP) titration experiments were performed so that the binding interface and binding affinity can be compared. The three LD4 peptides were chemically synthesized and titrated into ^{15}N labeled GIT1 PBD. The peptides were in the same length and only different in the C terminus. Similar to the GIT1 PBD titrated with LD4α peptide,[108] the LD4β and LD4γ peptide titrations dramatically perturbed the GIT1 PBD spectrum too. The same set of residues was moved in the same directions with the same intermediate exchange phenomena (Figure 4-1). In fact, the final spectra of GIT1 PBD titrated with 5 equivalences of LD4α, LD4β and LD4γ peptides were overlaid reasonably well (Figure 4-2). The shifted peaks mapped to the same binding interface – the groove between helix H1 and H4. The similar shifted distance and intermediate exchange phenomena indicated similar binding affinities. Thus, from NMR CSP titrations, we were able to conclude that the LD4β and LD4γ peptides interacted with GIT1 PBD in a manner very similar to the LD4α peptide.

Biacore SPR binding assay was used to compare the overall binding affinities between GIT1 PBD and the LD4 peptides. As summarized in the Table 4-1, GIT1 PBD bound to the three splice variants with very similar binding affinity (7.0 μm, 6.5 μm and 6.5 μm, respectively). It confirmed the observation by NMR titration experiments. Thus, the overall binding affinity between GIT1 PBD and the three splice variants were very close.

4.3.2   *Definition of minimal LD4 peptide length for the GIT1 PBD interaction*

With the aim of defining the minimal LD4 peptide length, we tested the binding of GIT1 PBD and LD4 peptides with different lengths. The Biacore SPR showed that the peptide ended at K277 (designated LD4_S1) retained the similar binding affinity while the peptide ended at D275 (designated LD4_S2) didn't have detectable binding signal (Table 4-1). Also, the interaction between GIT1 and S272 phosphorylation of LD4_S2 was also too weak to be detected by SPR.

Similarly, NMR CSP titration experiments indicated that the LD4_S1 had the same perturbation pattern with the three LD4 variants tested (Figure 4-2). The final spectra also overlaid satisfactorily with the final spectra of the three LD4 variants, which indicated the 5 residues extension in the C terminus did not have direct interaction with GIT1 PBD. NMR was able to detect the binding of LD4_S2, although the peak shifting is much weaker. The LD4_S2 S272 phosphorylation interacted even weaker with only slight peak shifts for GIT1 residues Y751, K755 and K758 (Figure 4-3). Therefore, paxillin K277 was the C terminal end of the minimal LD4 sequence for GIT1 binding. Comparing LD4_S1 and LD4_S2, the deletion of F276 and K277 dramatically decreased GIT1 binding. The two peptides had very similar CD spectra (Figure 4-4), indicating similar secondary structure compositions. Thus, the tighter binding of LD4_S1could not be attributed to a more helical structure. We also compared the ^{15}N HSQC spectra titrated with 5 e.q. of LD4_S1 or 5 e.q. of LD4_S2 peptides. As shown in Figure 4-5a, most peaks shifted to the same directions by both peptides, with a much larger distance for LD4_S1 peptide. However, some peaks moved in different directions. This could be
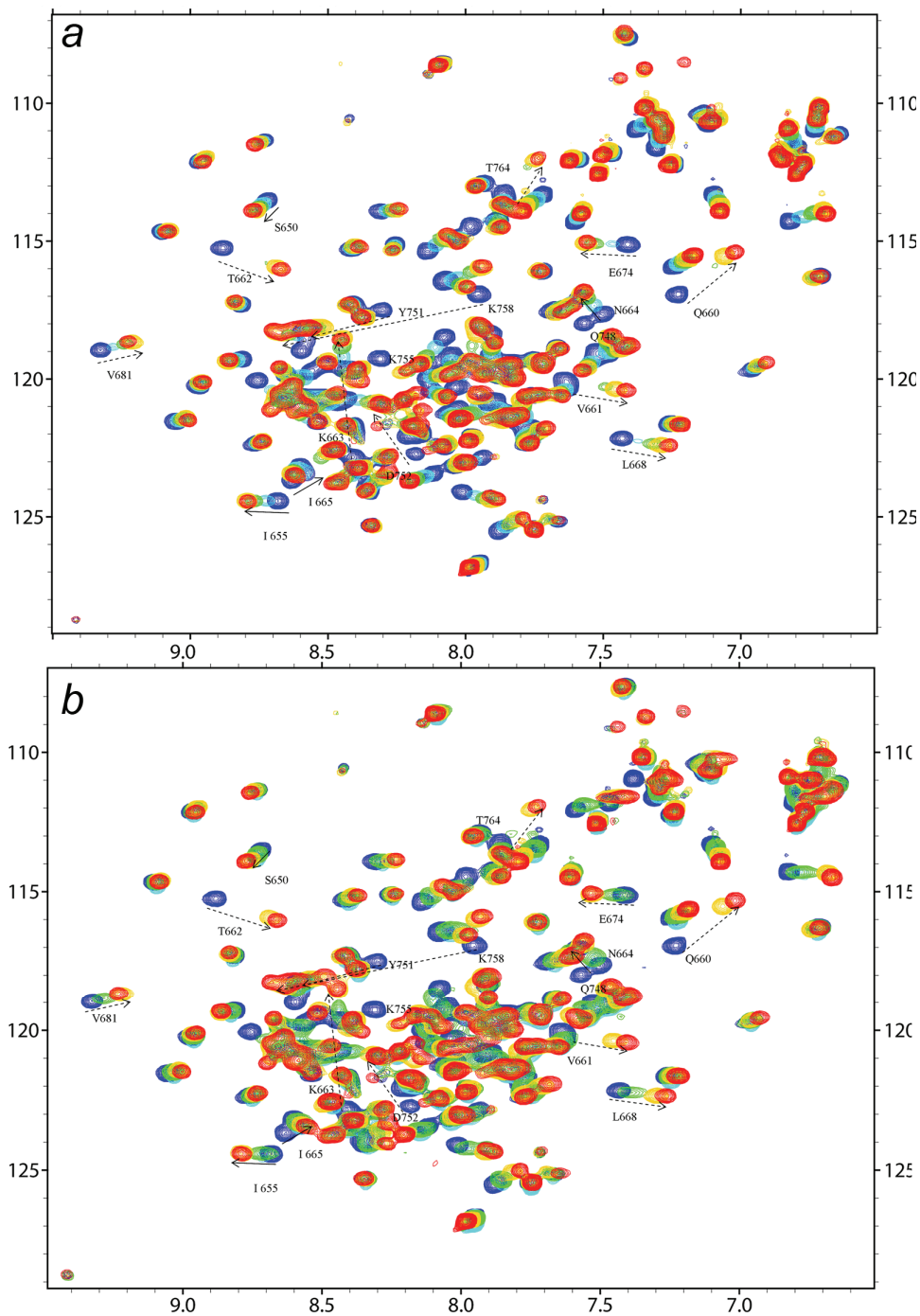
Figure 4-1   NMR CSP titrations with LD4 peptides derived from isoform β and γ.
NMR CSP titrations observing ^{15}N GIT1 PBD. GIT1 PBD: LD4 peptide color coded as Blue – 1:0; Cyan – 1:0.5; Green – 1:1; Gold – 1:2; Red - 1:5 (a) GIT1 + LD4 isoform β. (b) GIT1 + LD4 isoform γ.
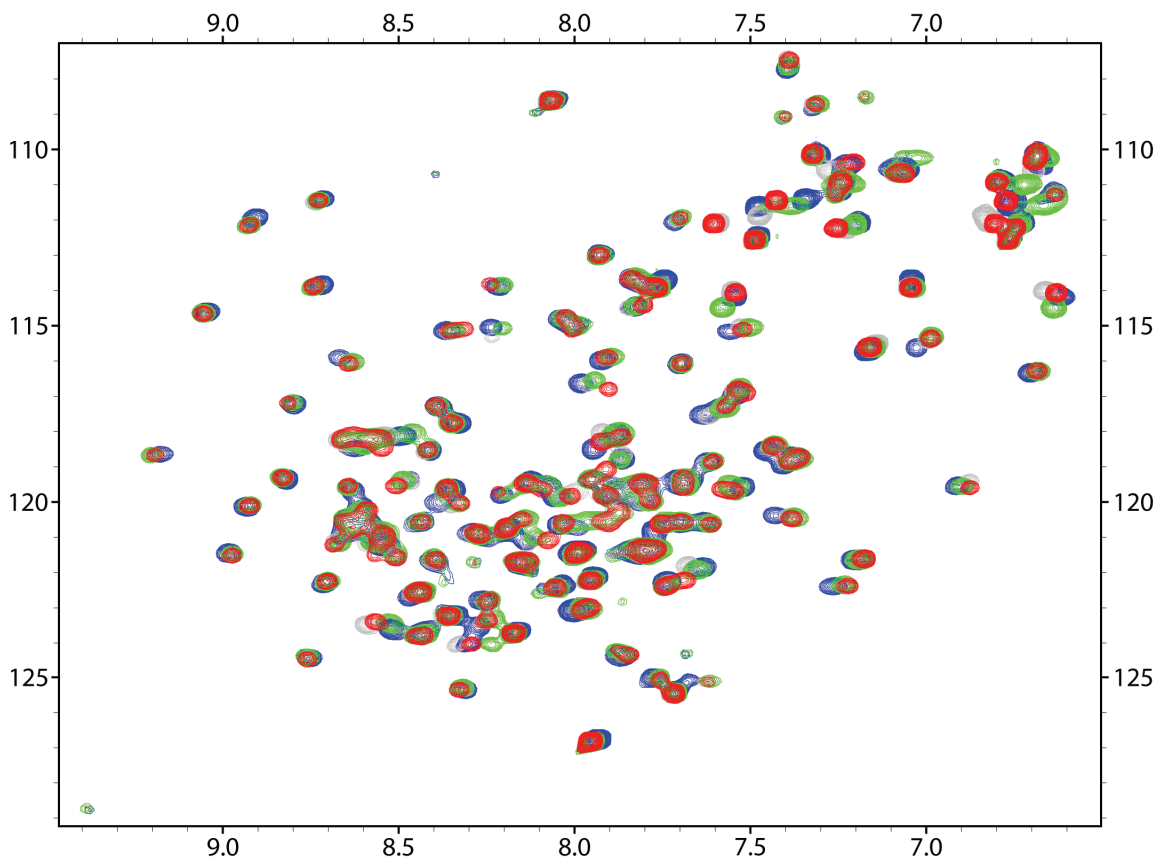
Figure 4-2   GIT1 PBD titrated with 5 equivalences of LD4α, LD4β and LD4γ peptides.

All spectra overlaid well. Blue - LD4α; Gray - LD4β; Green - LD4γ; Red – LD4_S1.
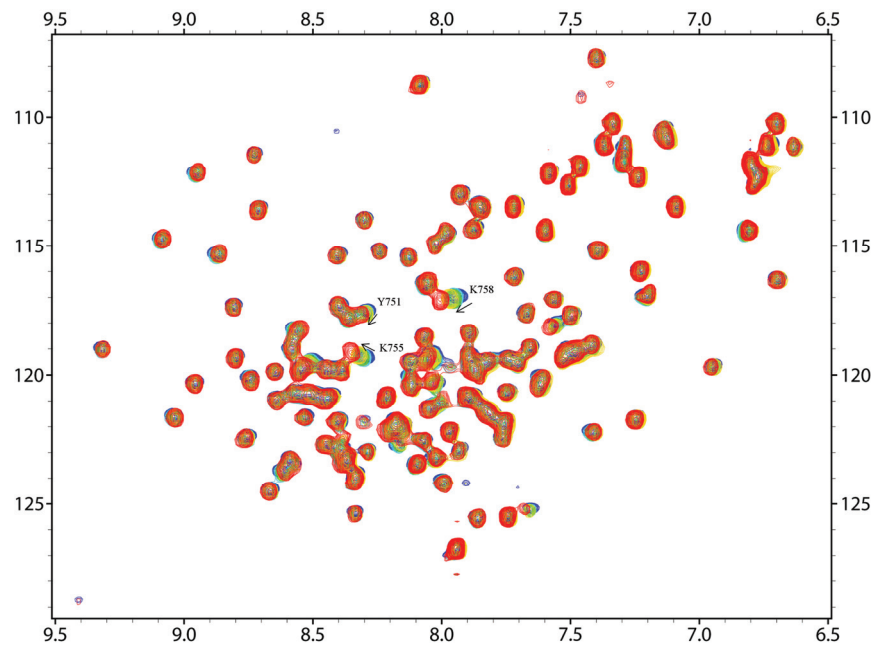
Figure 4-3   The LD4_S2 S272p interacted very weakly to GIT1 PBD.
*Binding only caused slight shifts to GIT1 residues Y751, K755 and K758. Blue-GIT1
alone; Green-GIT1+1:2 LD4_S2p; Red-GIT1+1:10 LD4_S2p.*
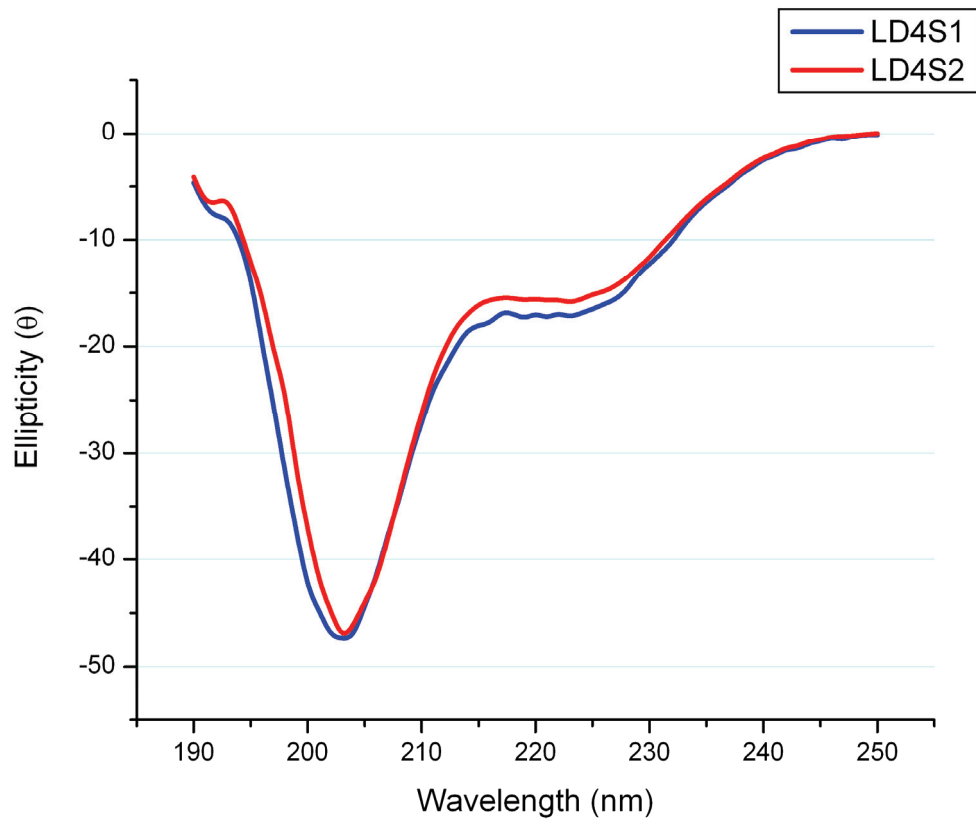
Figure 4-4   The CD spectra of LD4S1 and LD4S2 were very similar.
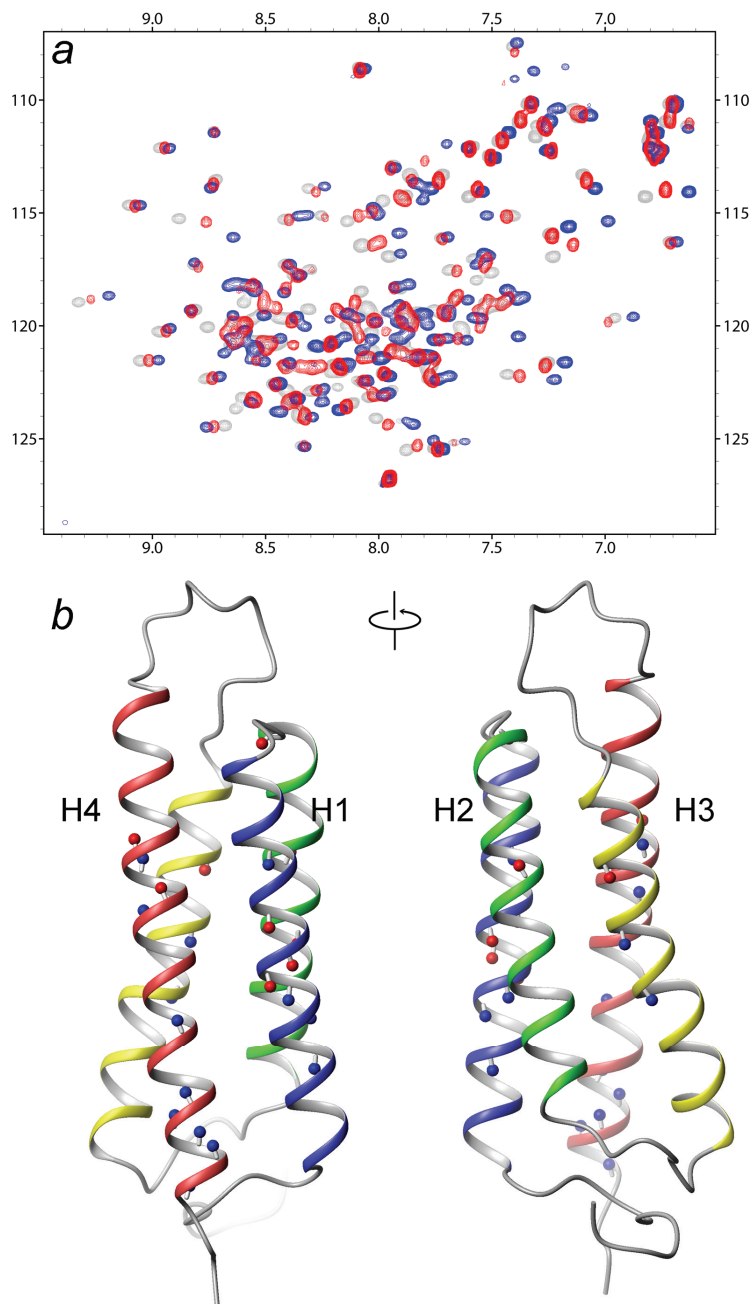
Figure 4-5   The comparison of GIT1 PBD titrated with LD4_S1 and LD4_S2 peptide.
(a) PBD spectrum titrated with either LD4_S1 or LD4_S2 peptides. Gray – PBD alone; Blue – PBD+LD4_S1; Red – PBD+LD4_S2. (b) The difference mapped to the PBD structure. All the little balls represent the amide protons that have composite shift larger than 0.06. The blue balls are peaks shifted in the same direction upon either LD4_S1 titration or LD4_S2 titration. The red balls are peaks shifted in different directions upon the two peptide titrations. Since the CSP direction difference can be attributed to the 2 residues difference at the C terminal end of LD4 peptide, it seems the LD4 C terminus interacts with the upper half of the H1/H4 surface.

accounted as the additional attachment site for F276 and K277, besides the overall tighter interaction. Those peaks were mapped to the upper half of the H1 and H4 when the N and C termini pointed down (Figure 4-5b). Peaks in H2 and H3 were shifted much less than those in H1 and H4, so they might be caused by the proximity or binding induced conformational change. Hence, the F276 and K277 might create an additional attachment at the upper half of H1 and H4. The observation was further confirmed with the complex model driven by independent data set.

4.3.3   *Both LD2 and LD4 motifs interacted with GIT1 PBD in the context of paxillin*

We have demonstrated that the ^{15}N GIT1 PBD titrated with unlabeled paxillin 133-288 was an average of LD2 and LD4 peptide titrations.[108] In order to directly investigate the paxillin-GIT1 interaction, we did titration experiment the opposite way. By NMR, we compared the ^{15}N HSQC spectra of ^{15}N labeled paxillin 133-288 in the absence and presence of unlabeled GIT1 PBD.

In the absence of GIT1 binding, paxillin gave a spectrum with most amide proton peaks clustering in the center within 1ppm range in proton dimension, which indicated lack of a well defined structure. In the presence of 0.8 e.q. of GIT1, the paxillin spectrum is greatly perturbed (Figure 4-6a). While more than half of the residues remained unchanged, about 20 peaks were largely shifted and some new peaks appeared in the more dispersed region. The addition of more GIT1 up to 2.4 e.q. resulted in a spectrum very similar to that with only 0.8 eq. of GIT1 (Figure 4-6b). Thus, it seems GIT1 PBD strongly bound to paxillin and the binding changed the conformation of part of the paxillin. A defined structure in a certain region was stabilized by the GIT1 binding.

In order to map the perturbed region in paxillin, at least the backbone resonance assignment was required. However, as a disordered protein, it was prohibitively difficult to make the unambiguous assignment. Previously, we had experimentally determined the structure of LD2 peptide and obtained its backbone assignment.[31,44] In addition, LD2 and LD4 peptides natural abundance ^{15}N HSQC spectra appeared to superimpose well with the spectrum of paxillin 133-288 (Figure 4-7). Thus, the partial assignment of paxillin 133-288 was obtained by adapting the assignment of the LD2 peptide,[31] based on the assumption that the ordered LD motifs were relatively independent and the synthesized LD peptides can represent the LD motifs in the paxillin. The peaks from the LD4 motif were also identified.

The partial paxillin assignment made it possible to identify the GIT1 binding site by directly mapping the GIT1 titration data. As shown in Figure 4-6, peaks from both LD2 and LD4 motifs experienced dramatic shifts. It directly exhibited the interactions between GIT1 PBD and both LD2 and LD4 motifs in the context of paxillin. Most shifted peaks were from either the LD2 or LD4 motif, excluding the possibility of LD3 motif as a binding site. Also, peaks other than from the LD2 and LD4 motifs kept at the same position, which ruled out the possibility of a wrapped loop between LD2 and LD4 motifs, as predicted for paxillin-FAK interaction.[31,44] Otherwise, we should expect to
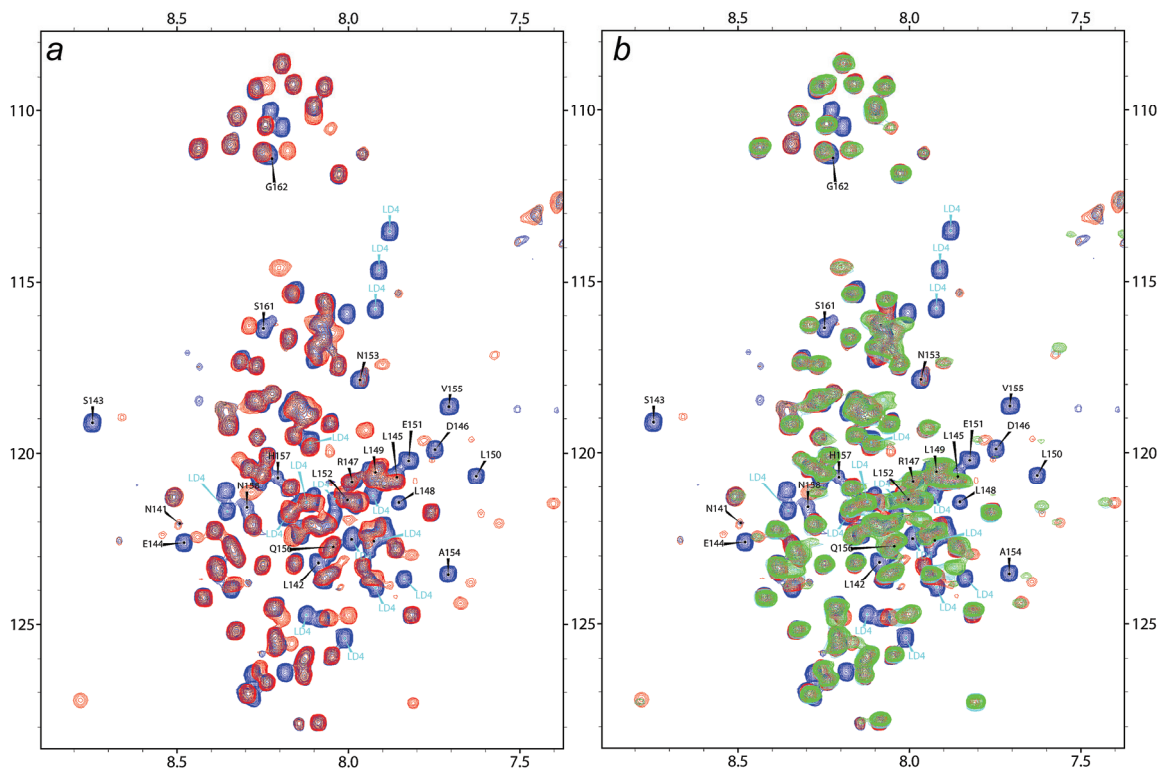
Figure 4-6 ^{15}N paxillin titrated with unlabeled GIT1 PBD.
Blue – 1:0; Red – 1:0.8; Cyan – 1:1.6; Green – 1:2.4.
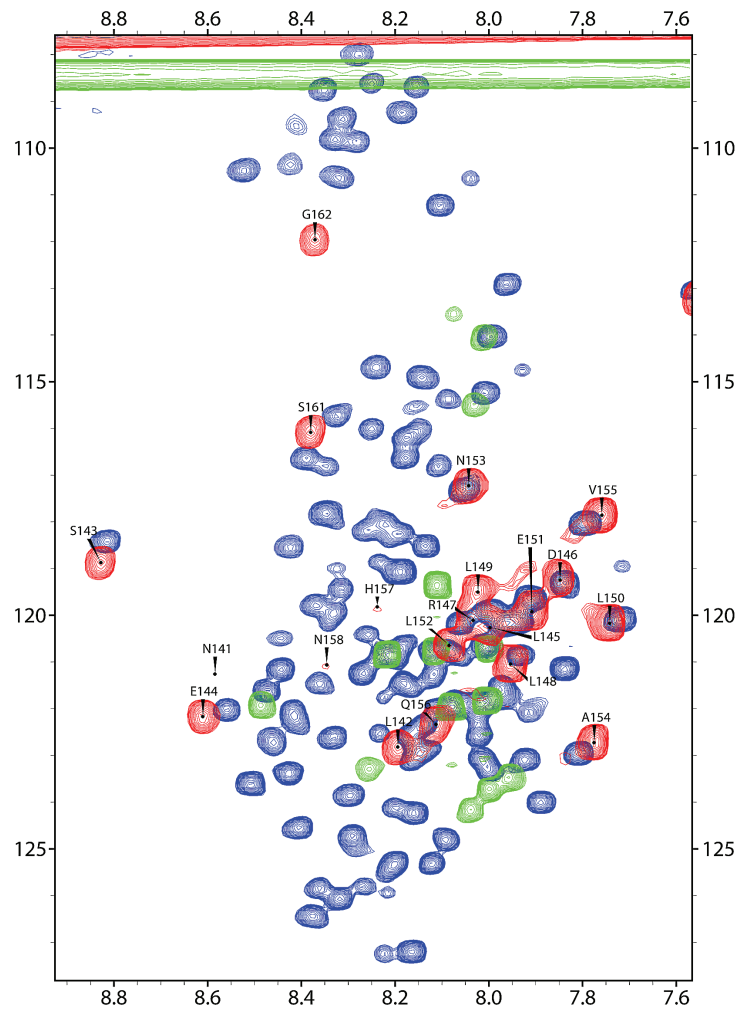
Figure 4-7   Superposition of the spectrum of paxillin 133-288 with the natural abundance HSQC spectra of LD2 and LD4 peptides.
The spectra overlaid well, allowing the partial assignment of paxillin. Blue – paxillin 133-288; Red – LD2 peptide; Green – LD4 peptide.

see remarkable chemical shift perturbations for most of the peaks, because the wrapping around was a huge conformational change for most of the residues in the paxillin 133-288.

4.3.4 NMR driving HADDOCK complex modeling

Although we have determined that both LD2 and LD4 bound to GIT1 PBD on the H1/H4 surface, the interaction details were still unclear without a complex structure. By NMR, we determined the solution structure of GIT1 PBD as a 4-helix bundle.[108] We had also solved the solution structure of LD2 peptide as a single α-helix.[31] CD spectrum suggested LD4 peptide was also an α-helix in solution (Figure 4-8). The binding site on GIT1 PBD had been precisely mapped by NMR CSP titrations experiments.[108] Thus, we try to model the complex structures by HADDOCK.[77] For GIT1, the active and passive residues were defined according to the NMR CSP titration data and solvent exposure surface data.[108] For the peptides, the CSP based binding site mapping was not applicable, because all the peaks experienced shifts. Thus, paxillin sequence conservations were used to define the active and passive residues.

Figure 4-9 showed the complex models with the best intermolecular energy out of 200 poses calculated by HADDOCK with explicit solvent refinement. Both LD2 and LD4 peptides attached to the GIT1 H1/H4 groove as an additional helix to the bundle. The complex models for both peptides were very similar, in both distances and helix axis angles (Figure 4-10). Both interactions were dominated by hydrophobic interactions between the hydrophobic GIT1 groove and the hydrophobic LD side chains, e.g. the residues at Ø-3, Ø+0, Ø+3, Ø+4, Ø+7 and Ø+10 (colored yellow in Figure 4-9). It might explain the reason why both LD2 and LD4 peptides could interact with the GIT1 H1/H4 face in a very similar manner.[108]
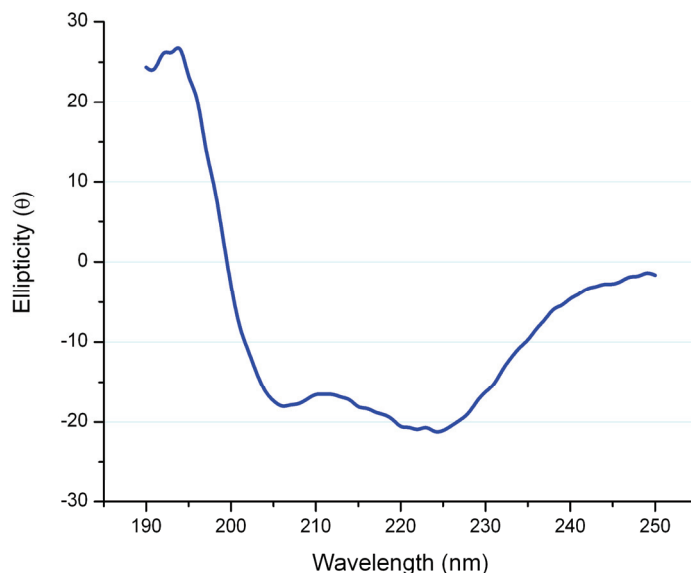
Figure 4-8   CD spectrum suggested LD4 α peptide was an α-helix in solution.

84

Figure 4-9   Complex structures of GIT1 PBD bound with LD2 peptide or LD4 peptide.
Both peptides bound to the hydrophobic groove between H1 and H4. The GIT1 PBD surfaces were colored according to surface charge, with blue for positive and red for negative. The hydrophobic residues from LD peptides were colored in yellow. (a) GIT1 PBD + LD2 peptides; (b) GIT1 PBD + LD4 peptides.
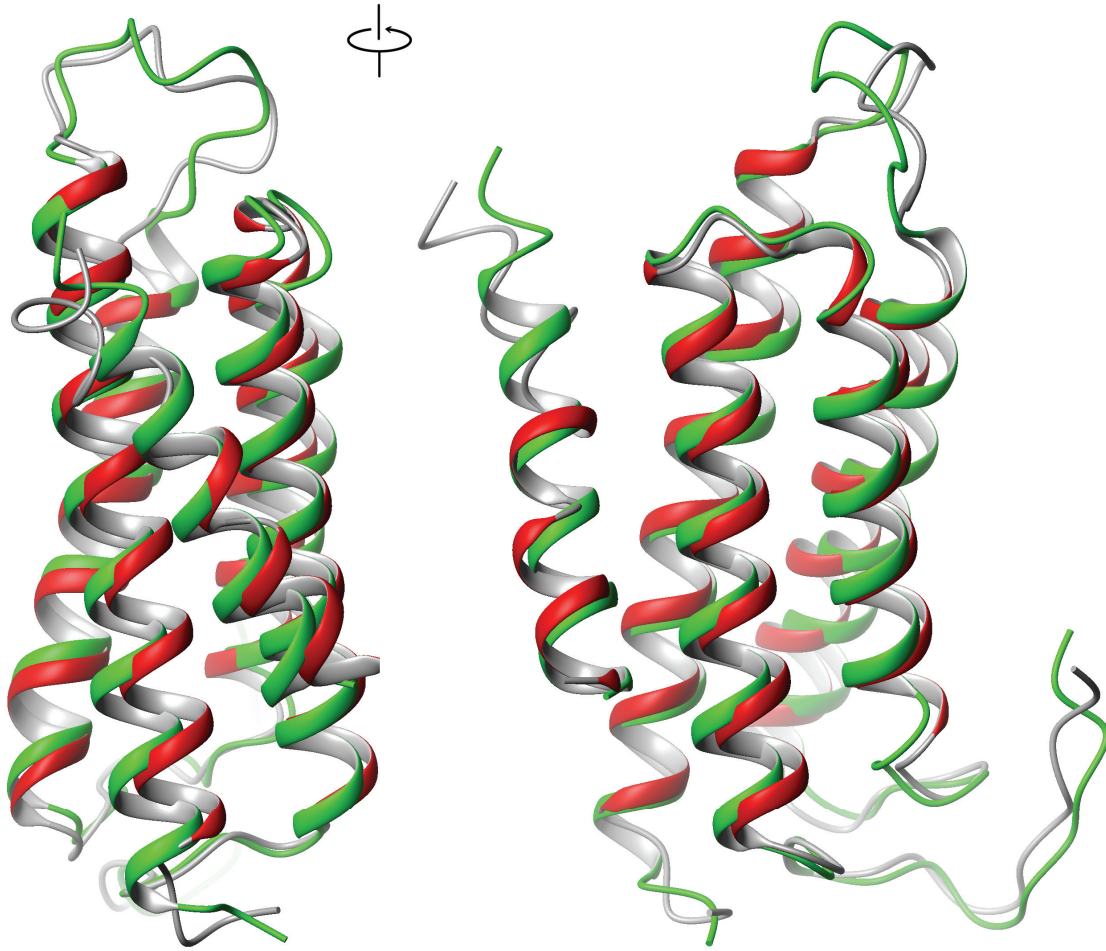
Figure 4-10   The overlay of GIT1 PBD complex structures bound with LD2 or LD4 peptides.
The two complex structures were very similar. Red – GIT1 PBD + LD2 peptide; Green - GIT1 PBD + LD4 peptide.

Besides hydrophobic interactions, there were some well defined salt bridge and H-bond connections between the LD peptides and the two polar and positive charged ridges.[108] The conserved D146 and D267 (Ø+1) formed a salt bridge with GIT1 K758 for both LD peptides. Similarly, this salt bridge is conserved between FAK K1032 and paxillin D146 (Ø+1) in the FAK-LD2 complex structure.[32] There were H-bonds between GIT1 E659 and paxillin N141 (LD2 Ø-4) or S261 (LD4 Ø-5). For GIT1-LD4 complex structure, a network of salt bridges and H-bonds was built by the E265Pax-K663GIT1-E268Pax-Q666GIT1-S272Pax connections (Figure 4-11). Experimentally, these GIT1 residues were among those mostly shifted upon both LD peptide titrations.[108] However, in GIT1-LD2 complex structure, the network was broken by the substitution of E268 (LD4 Ø+2) with R147 (LD2 Ø+2), leaving only E144Pax-K663GIT1 and Q666GIT1-E151Pax (Figure 4-11). The breaking of a well defined H-bond network might contribute to the observation that LD4 peptides bound to GIT1 slightly stronger than the LD2 peptide.[108] The binding affinity difference was also supported by the slightly larger buried surface areas (1474.64 Å^2 for LD4 compared to 1417.26 Å^2 for LD2) and intermolecular energy calculations (-446.162 Kcal/Mol for LD4 compared to -315.141 Kcal/Mol for LD2).

With the complex structures, we would like to answer why the H1/H4 surface of GIT1 can bind both LD2 and LD4 peptides, while the analogous FAK surface binds only LD2 peptides.[31,44,110] As shown in Figure 4-12, overall the two surfaces were similar as both featured a hydrophobic groove sandwiched by two polar and positive charged ridges. However, FAK had an additional positive charged patch at the upper end of helix 4, formed by K1018 and K1019. The analogous position on GIT1 was T743 and Q744, which is in proximity to the positive charged LD4 K277 in our GIT1-LD4 complex structure. Different from LD4, the same position in LD2 peptide was Q156, which should not cause charge repulsion. Therefore, it might be possible that the binding specificity of FAK H1/H4 site is determined by the positive charged patch at the upper half of H4.

There were no contacts between GIT1 PBD and the last 5 residues in LD4 peptide, which explained why LD4α, LD4β, LD4γ and LD4_S1 peptides behaved very alike in both NMR CSP titrations and Biacore SPR binding assays. However, the paxillin F276 and K277 made extensive contacts with the H1 and H4, correspondingly. It explicated the huge difference between LD4_S1 and LD4_S2 peptides and confirmed our hypothesis that the binding affinity change was the result of additional contacts.

The complex structures were validated by the recent mutation experiment, in which GIT1-paxillin binding was weakened to different extent by mutations such as K663E, K758E, I747E, T662E and Q666E.[62] Those residues were right in the binding interface in both complex structures. Our model also explained the reason why Y751F mutation didn't affect binding was that the main interaction here was the hydrophobic stacking of the tyrosine aromatic ring against the side chain of paxillin L152 and V155(for LD2 complex) or L273 and F276 (for LD4 complex).
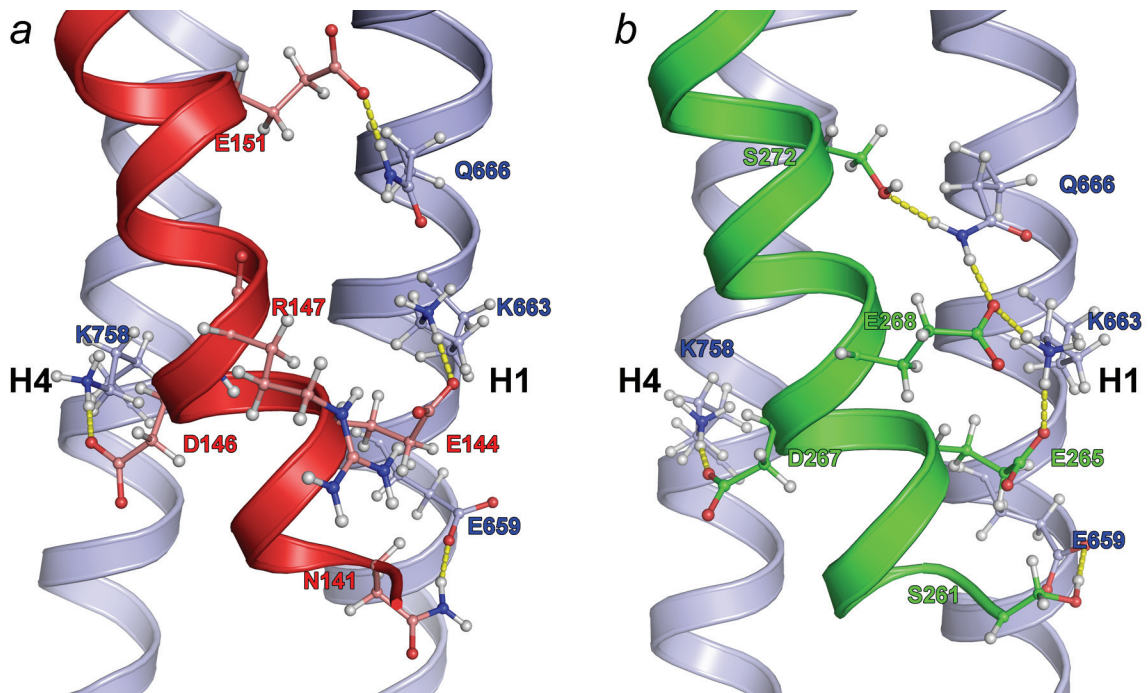
Figure 4-11   The side chain salt bridge and H-bond network.
(a) The complex structure of GIT1 PBD and the LD2 peptide. (b) The complex structure of GIT1 PBD and the LD4 peptide.
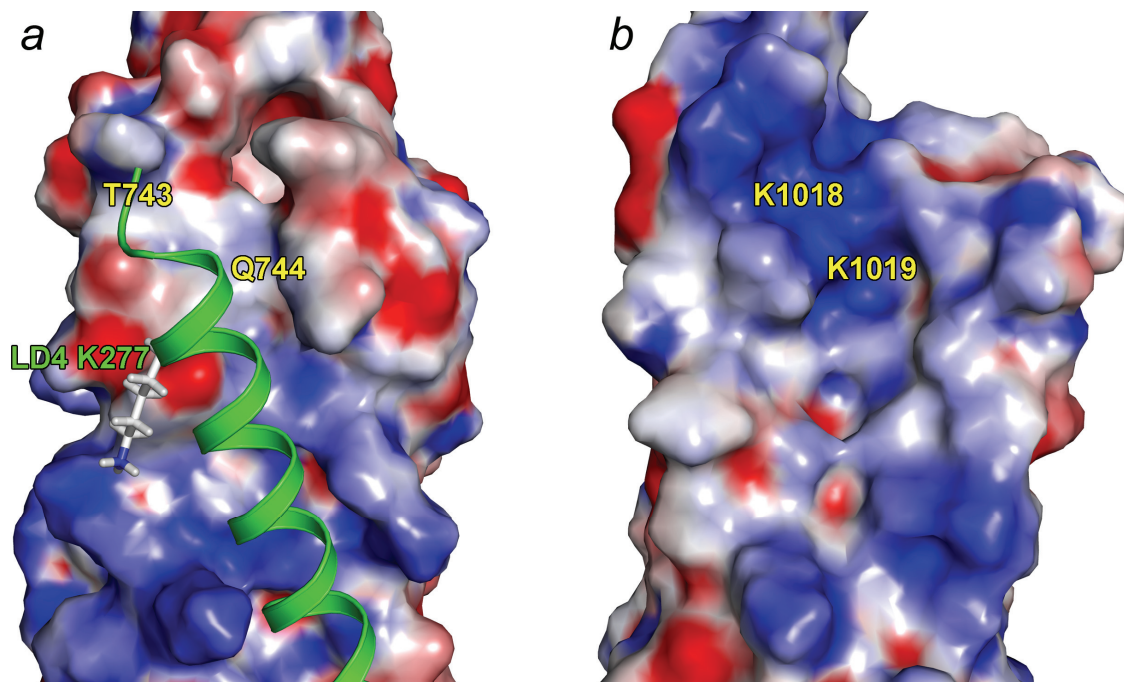
Figure 4-12   H1/H4 surface of GIT1 PBD and FAK FAT.
Surfaces were colored according to the charge, with blue for positive and red for negative. (a) GIT1 PBD bound with LD4 peptide. The LD4 K277 is in proximity to GIT1 T743 and Q744. (b) FAK FAT domain surface map. The two Lys create a large positive charged patch that may prevent LD4 binding by repelling LD4 K277.

4.4 Discussion

Both NMR CSP titrations and Biacore SPR binding assay revealed that there was no significant difference between the three paxillin LD4 peptide variants and the shorter LD4_S1. They interacted with the same interface on GIT1 with similar binding affinity. It disproved the hypothesis that paxillin splice variants might bind GIT1 differently. However, their interaction to other proteins, e.g. FAK and vinculin, should also be tested in order to conclude if the specificity is determined by the LD4 proximal sequences. Since there was no sequence homology in the last 5 residues, the study also confirmed that the last 5 residues were not directly involved in GIT1 interactions.

Comparing the LD4_S1 and the LD4_S2, the deletion of the last two residues dramatically decreased the GIT1 interaction. Why were the two residues were so critical? There were two possible reasons: 1. the two residues capped the peptide C terminus and made the peptide more helical (more ordered); 2. the two residues made additional attachments to GIT1. The more ordered peptides should bind stronger than the less ordered ones because of the entropy loss upon binding. The CD data disfavored the first hypothesis by showing that the two residues did not change the overall secondary structure. So the stronger binding was not the result of increased helicity by the C terminal capping. In contrast, both NMR CSP titration and complex modeling supported the second hypothesis. The GIT1-LD4 complex even showed extensive contact between GIT1 and F276/K277. Thus, it is likely the two residues made additional contacts that dramatically increased the binding affinity.

The CSP titration with ^{15}N labeled paxillin 133-288 substantiated our previous study that both paxillin LD2 and LD4 were able to interact with GIT1 PBD.[108] It also confirmed our observation that the two LD motifs bound PBD with a similar affinity,[108] since peaks from both motifs experienced perturbation when there were more LD motifs than PBD (LD2:LD4:PBD=1:1:0.8 for the first point titration). With only partial assignment, we were able to differentiate paxillin 133-288 resonance from either LD2 motif, LD4 motif or the rest of the protein. GIT1 PBD perturbed both LD2 and LD4 motifs and those were the only region perturbed. So it is likely GIT1 PBD interacts with both LD2 and LD4 motifs but not at the same time. There was no wrap-around for the linker region, supporting the hypothesis that one paxillin interacts with two copies of PBD simultaneously. Since GIT1 can dimerize through the coiled-coil domain,[104] it is highly possible the dimer interacts with paxillin through both LD2 and LD4 motifs concurrently, achieving comparable affinity as the FAK-paxillin interaction.

We deemed the complex models reliable. The reason included: 1. It was a small system and there were no conformational changes upon binding;[108] 2. The top lowest binding energy models were classified to the largest groups for both dockings; 3. The models were validated by mutation data. Our GIT1-LD2 complex structure resembled the FAT-LD2 peptide complex structure (1OW8 chain A,D, [32]), with an overall backbone RMSD of 1.38 Å. A close analysis found that FAT-LD2 binding was mainly hydrophobic interactions too, with side chain salt bridges and H-bonds. Most of the side chain interactions were even conserved between FAT and GIT1. This, along with the

supporting experimental data, demonstrated the reliability of our complex structure models. For the GIT1-LD4 complex structure, the lack of contacts between GIT1 and the last five residues in LD4 explained the similarities within LD4 splice variants and LD4_S1 peptides. The binding surfaces overlapped more than 90% for LD2 peptide and LD4 peptides, excluding the possibility of simultaneous LD2 and LD4 interactions.

Most importantly, our complex structures suggested that FAK K1018 and K1019 might be the determinant for its binding specificity in favor of LD2 to LD4 on the H1/H4 surface. Without this positive charged patch, GIT1 is able to interact with both LD2 and LD4 on the H1/H4 surface. Further mutation experiments are needed to confirm this conclusion. Thus, the complex models provide structural insight for the GIT1-paxillin interaction and are presented as a starting point for applications such as structure based ligand screening.

CHAPTER 5. CONCLUSION AND DISCUSSION

5.1 The NMR structure of GIT1 PBD

5.1.1 The accuracy and quality of the structure

The accuracy of the NMR derived PBD structure was supported by the H/D exchange experiment. As seen in Figure 3-1d, all the solvent accessible amide protons (blue) were mapped to the 3 inter-helix loops and the 2 ends. So the flexible regions identified by the H/D exchange experiment correspond well to the loops in the structure. The protected peaks were mapped exclusively to the center of all 4 helices, which approved the accuracy of the structure. Also, the low target function value (0.70) also indicated the good correlation between the structure and NOE data. In addition, the PBD structure is very close to the previously reported GIT1 PBD homology model.[62] Thus, the accuracy of the GIT1 PBD presented in this dissertation is supported from more than three independent data sources.

The quality of the GIT1 PBD structure is elucidated by the final structure statistics. As seen in Table 3-1, the structure was derived from a total of 2440 NOE distance restraints and 189 dihedral angle restraints. Among them, there are 398 long range NOE restraints, most of which were inter-helix NOE correlations. On average, there are 18.07 NOE restraints per residue. So the structure was supported by a good number of restraints.

The NMR structure bundle is of good precision too. The overall backbone RMSD value is 0.39 Å and the heavy atom RMSD value is 0.77 Å. In the Ramachandran plot, 88% of residues are in the most favored region and only 1 residue is in the disallowed region. It was the D643 which located at the very N terminal end in an unstructured region.

Thus, all the statistics support that the structure presented in this dissertation is a high quality NMR structure.

5.1.2 Structure analysis

Overall, GIT1 PBD is a rod shaped molecule, about 58 Å in height and 26 Å in diameter. The molecule shape was validated by an analytical ultra-centrifugation (AUC) experiment, which derived an f/f_0 value of 1.24. At the same time, the SAXS model also matched well to our structure in overall shape.[62] The low resolution SAXS model showed that H3 and H4 appeared to be longer than H1 and H2, which can be explained by the fact that the H3/H4 loop was longer than the H1/H2 loop in our structure.

The solution structure of GIT1 PBD is a 4 helix bundle with right handed up-and-down antiparallel topology. The individual helices form an average axis angle of 19.1º, which is very close to the ideal antiparallel helix packing angle of 20º.[102] This suggests a well organized inter-helix side chain network. Actually, a well conserved hydrophobic side chain network has been identified from GIT1 PBD, FAK FAT and vinculin Vt domain.

The GIT1 PBD is a well folded monomer in solution, which is different from the dimeric FAK FAT domain.[91,101] The GIT1 PBD spectra taken at different temperatures revealed that all the peaks have similar intensity upon temperature change. It is an indication of the absence of local regions with higher mobility. However, the FAK FAT domain had local line broadening in the loop region between H1 and H2. The comparison explained the reason why FAT can dimerize by swapping the H1. The H1/H2 loop of PBD has the flexible sequence of *EFKH* comparing to the *PAPP* found in FAT. We proposed that the conformational strain introduced by the proline rich loop is the cause of the higher dynamics of FAT H1 and the related dimerization behavior.

5.1.3   *The comparison to FAK FAT domain*

GIT1 PBD, the FAT domain of FAK and the vinculin Vt domain share a well conserved hydrophobic core, while the surface residues of the three domains exhibit more diversity. This ensures the three domains to have a very similar overall fold while keeping distinct binding specificity. There are more common points within the three domains: 1. they are all independent helix bundle proteins; 2. they share a conserved inter-helical side chain network; 3. they can all bind to the paxillin LD motifs and are responsible for focal adhesion targeting. Thus, the three protein domains are very similar in both structure and function. This suggests a common ancestor for the three domains.

However, there are many differences between GIT1 PBD and FAK FAT. Although they can both bind to paxillin at LD2 and LD4 motifs, PBD binds them at the same site – the H1/H4 face, while FAT binds them simultaneously at H1/H4 face and H2/H3 face. The binding specificity can be elucidated by comparing the surface residues of PBD and FAT.

5.1.3.1   The H1/H4 face

Both GIT1 PBD and FAK FAT can bind paxillin LD2 motif at the groove surface between H1 and H4. There are many similarities at this surface. They are hydrophobic grooves sandwiched by positive charged and polar ridges. Both of the hydrophobic grooves are elongated with similar size, accommodating the same set of hydrophobic side chains from the LD2 motif. The two positive charged Lys residues (one at each ridge, see Figure 3-6) locate exactly at the same positions on both PBD surface and FAT surface. In the complex structures, the two Lys residues function as attachment points to anchor the same two negative charged residues in the LD2 motif. More specifically, the

K663GIT1 or K934FAK forms salt bridge with D146 in the LD2 motif. The K758GIT1 or K1033FAK form another salt bridge with E144 in the LD2 motif. Besides the two conserved salt bridges, there are more conserved hydrogen bond pairs between LD2 and the polar residues from the ridges. Thus, we can conclude that the H1/H4 surface is highly similar between GIT1 PBD and FAK FAT. This perfectly explained the reason why the paxillin LD2 motif can bind to both GIT1 PBD and FAK FAT.

5.1.3.2   The H2/H3 face

The FAK FAT domain can bind to LD4 at its H2/H3 face, but the GIT1 PBD can not bind to any LD motif at its H2/H3 face. The different binding specificity is rooted in the distinct surface residues. Although both surfaces have a hydrophobic groove, the residues at the ridges are different. On the H2/H3 surface of GIT1 PBD, the hydrophobic groove is closed by two bulky residues, the H688 and Y719, at the center of the groove. Differently, the two positions are occupied by two small residues (G959 and A996) on the FAT surface. So FAT does not have the problem of a closed groove. In addition, the positive charged residues (K956 and R963) on FAT ridge are also substituted by negativly charged ones (E685 and T692). Since the positive charged surface patch may be responsible for the anchoring of the two negative charged residues in the LD4 motif, the lack of those positive charged residues might be another reason why GIT1 PBD cannot bind to the LD4 peptide. Thus, there is major difference comparing the H2/H3 surface of PBD and the FAT domain.

5.1.4   *Other 4 helix bundle proteins in FA*

GIT1 PBS shares the 4-helix bundle structure with both FAK FAT and vinculin Vt. Besides, based on the high sequence homology, we also predict the same folding for FAK2 (PYK2) and GIT2 (PKL) C terminal paxillin binding domain.   It appears that this is an evolutionarily preferred structure for the binding of paxillin LD motifs. In addition, there are more 4-helix bundle protein domains existing in focal adhesion components, e.g. p130^{CAS} (1Z23 [111]) , Talin VBS2 (1U89 [112]) , VBS3 (2B0H [113]) and α-Catenin (1H6G [114]). They are all bundles composed of 4 amphipathic helices in a right handed up-and-down topology. All the helices were bundled up with 4 anti-parallel helices at an average inter-helix angle of 20º.

The reason for the concentration of so many 4-helix bundle proteins in focal adhesions remains unclear. Some features of the helical bundle may contribute to the answer:

1. 4-helix bundle proteins tend to be small and stable. The up-and-down topology enables the forming of a bundle with minimal loop residues. Besides, this also helps to increase the stability by aligning helix dipole moment so that opposite charged ends are against each other.

2. Although small in size, the stability allows them to expose a relative large binding surface (between helices), which results in a large binding energy.

3. The bundle offers 4 faces, which makes it possible for the existence of multiple binding sites.

4. The bundle often binds α helix parallel to its elongated groove, which may help fix the angle between the two binding partners. This is of more importance for scaffold proteins to be able to bring its binders into proximity with the right distance and orientation.

Thus, four-helix bundle protein domains have the features suitable for functioning at the focal adhesion. It may be the reason why so many four-helix bundle proteins exist in the FA.

5.2   The GIT1-paxillin interaction studies

The GIT1-paxillin interaction has been studied extensively in this dissertation. With ^{15}N labeled GIT1 PBD, we studied the interaction between PBD and different LD peptides, including LD2, LD4, LD4 S272p, LD4 splice variants, and LD4 shorts. The studies mapped the binding site and compared the binding affinities. More quantitatively comparison of binding affinity has also been made by Biacore surface plasma resonance binding assays. With ^{15}N labeled paxillin, we studied the interaction the other way, in which the GIT1 binding site on the paxillin was also mapped.

5.2.1   Binding site on GIT1

The binding of LD peptides perturbed the ^{15}N HSQC spectrum of GIT1 PBD and allowed mapping the binding site on GIT1. Both LD2 and LD4 peptides perturbed the same set of GIT1 spectra, which mapped to the groove between H1 and H4. The binding site for both LD2 and LD4 is on the H1/H4 surface. This observation corresponds well with the SAXS data that the LD4 peptide bound to GIT1 PBD at the H1/H4 surface.[62]

The perturbations focused at the center of the H1/H4 groove, in which many residues experienced long range shifts and intermediate exchange phenomena. Residues in the H2 and H3 experienced small shifts too, but the small shifts were the results of proximity to the H1/H4 site. The reason includes: 1. the shifts were very small compared to residues at the H1/H4 site; 2. the buried H2/H3 residues had longer shifts than the exposed residues. This exactly showed that it is the proximity to the H1/H4 that caused those shifts. The GIT1-LD interactions were not at the surface of H2/H3, since the H2/H3 solvent exposed residues had less perturbation.

Recently, in order to pinpoint the paxillin binding site on GIT1, a series of PBD point mutations were made and their ability to bind the full length paxillin were

examined.[62] The mutation data perfectly matches our results, providing more biochemical support. The residues that are critical to paxillin/GIT1 binding are mapped to the H1 and H4 site, while most mutations on H2 and H3 have minimal effect. In the study described before, perturbed residues also cluster in H1 and H4, with minimal changes in H2 and H3. Furthermore, while both K663E and K758E mutations abolished paxillin binding, both the Ks were among the most dramatically perturbed residues in the titration experiments.

Also, among different species, H1/H4 has higher sequence identity than H2/H3. So from the evolutionary point of view, H1/H4 seems to be more important than H2/H3. Since the function of GIT1 PBD is paxillin binding, it is likely that the paxillin binding site is located to the more conserved H1/H4 surface.

Thus, many evidences converge to identify that the GIT1 paxillin binding site is in the hydrophobic groove between H1 and H4. The conclusion is supported by data from NMR CSP titration experiments, biochemical mutagenesis experiments, SAXS and sequence analysis.

The binding of LD2 and LD4 peptides resulted in a very similar chemical shift perturbation pattern. This is the direct evidence that the LD2 peptide and LD4 peptide bind GIT1 in a very similar manner. The observation was confirmed by the complex structure modeling. In the complex structures, the LD2 peptide and the LD4 peptide interact with GIT1 at the same angle and distance in the same pocket. Most of the side chain interactions are even conserved, for example, the salt bridges between the two anchoring Lys on GIT1 to the Glu and Asp residues on the LD2 and LD4 peptides. More details about the complex structure are discussed in section 5.3.

5.2.2   Binding site on paxillin

In order to discover the GIT1 binding site on the paxillin, NMR CSP titration experiments were performed from two directions. Firstly, the experiments with ^{15}N labeled GIT1 PBD identified LD2 and LD4 motifs as the binding site. Secondly, the experiment with ^{15}N labeled paxillin confirmed that both LD2 and LD4 motifs participate in GIT1 binding. It also excluded the possibility of LD3 as a binding site.

5.2.2.1   Studies from ^{15}N labeled GIT1

Using ^{15}N labeled GIT1 PBD, we tested the interactions to peptides derived from paxillin LD2 and LD4 motifs. The NMR CSP titration experiments discovered a PBD-LD2 interaction, in addition to confirming the PBD-LD4 interaction. Both paxillin LD2 and LD4 peptides were able to dramatically perturb the NMR spectra of GIT1 PBD. Thus either LD2 or LD4 or both motifs could be the GIT1 binding site.

To determine which LD motif is the real GIT1 binding site, their binding affinity needed to be compared. The competitive NMR titration experiments were used to directly compare the binding affinity. The experiments showed that LD2 peptide could further perturb the spectrum in the presence of equal molar amounts of LD4 peptide. Similarly, the LD4 peptide could further perturb the spectrum in the presence of equal molar amount of LD2 peptide. And the peaks from the two competitive experiments moved to a central position between the positions of single peptide titration (Figure 3-12). So it is clear that the two peptides bind to GIT1 PBD at very similar binding affinities; otherwise if one peptide binds significantly stronger than the other, only the stronger one can further perturb the PBD spectrum. The observation was confirmed by the Biacore surface plasma resonance binding assay. It showed the Kd values for PBD-LD2 and PBD-LD4 interaction were different by only three folds, which is close enough for them to compete if equal molar concentrations co-exist.

The CSP titration with ^{15}N labeled GIT1 and unlabeled paxillin 133-288 further confirmed that the LD2 and LD4 motifs both interact with GIT1. The PBD spectrum titrated with paxillin 133-288 was the average of PBD spectrum titrated with equal molar of both LD2 and LD4 peptides (Figure 3-12). It means the GIT1 responds to paxillin the same as to the LD2 and LD4 peptides together.

Thus, all the evidences above support that the paxillin GIT1 binding site is both the LD2 and the LD4 motifs. Why did previous studies find that only LD4 is required for GIT binding? One noticeable difference is that we are studying GIT1 protein, while previous studies focus on GIT2 (p95-PKL).[38,61] Although GIT1 and GIT2 have 65% sequence identity and 85% of similarity,[49] they are different in many ways, e.g. GIT1 exists mainly in full-length, but GIT2 is largely alternatively spliced;[53] although GIT2 is uniformly distributed, GIT1 only expresses in certain cells;[115] GIT1 binds paxillin with a much higher affinity than does GIT2 (p95-PKL);[53] Src, FAK activity is required for GIT2, not GIT1, to localize to FA;[116] GIT1, not GIT2, can localized to the FA with constitutively active PAK.[60,107] The other difference is that we are studying protein interactions in the absence of other context proteins. Since paxillin LD motifs are the binding sites shared with other proteins, e.g. FAK/PYK2, GIT1/GIT2, vinculin, Actopaxin or Clathrin, it is possible the LD2 motif is not always available for GIT1 binding, probably due to masking or allosteric hindering by other proteins; and this could be an interesting way of regulation. With this in mind, the current study provides a more direct observation that LD2 can actually interact to GIT1. Since it is still unknown how paxillin dynamically interacts with multiple partners, this new finding may shed more light on the signal integration function of paxillin.

5.2.2.2   Studies from ^{15}N labeled paxillin

The NMR CSP titrations observing the ^{15}N labeled paxillin provided more direct information about the GIT1 binding site. With the innovative approach, we were able to partially assign paxillin. Using the partial assignment, the CSP titrations unambiguously identified the GIT1 binding site is both the LD2 and LD4 motifs.

Firstly, we want to discuss the correctness of the partial paxillin assignment. The assignment was based on the comparison of the spectra of paxillin and its component LD motifs. The paxillin 133-288 is a flexible protein, with folded LD motifs connected by unfolded linker sequences. It can be described as "beads-on-the-string" model, in which the individual motifs are relatively independent and the HSQC peak positions are at the same positions as the HSQC peaks of individual motifs. We found that the HSQC spectra of individual LD motifs matched well with the paxillin HSQC spectrum, which meant there were no interactions between the LD motifs and the rest of the protein. The LD motifs in the paxillin construct folded and behaved the same as the isolated LD peptides. This validated our approach to assign the paxillin by the assignment of the individual LD motifs.

The partial assignment enabled us to identify which residues were from the LD2 motif, which were from the LD4 motif and which were from the rest of the protein. In the subsequent titration experiments, it directly showed which part of the paxillin was involved in the GIT1 binding.

The titration of unlabeled GIT1 PBD into the ^{15}N labeled paxillin 133-288 perturbed all the peaks from both the LD2 motif and the LD4 motif. Thus, this is the direct evidence that both LD2 and LD4 motifs are involved in GIT1 binding. At the same time, all the peaks other than from the LD2 and LD4 motifs remains exactly at the same position. So they are not involve in the GIT1 binding for the rest of the paxillin construct, including the LD3 motif and the linker sequences. It excluded the possibility of wrapping around like the situation of the FAT-paxillin interaction, where the paxillin LD2 and LD4 motifs bind to two opposite face of the FAT and the sequences in between wrap around the FAT. The data supports the model that the LD2 motif binds to a molecule of GIT1 and the LD4 motif binds to a second molecule of GIT1, because the residues in between do not have any perturbation.

Thus, data from this part of study directly confirmed that the paxillin GIT1 binding site is both LD2 and LD4 motifs.

5.2.2.3   Model of GIT1-paxillin interactions

The novel finding that the paxillin LD2 motif can also interact with GIT1 PBD opens a new window toward the understanding of FA signaling. Current data supports the idea that FAK binds paxillin in the way that paxillin LD2 and LD4 motifs bind simultaneously to FAT in both the H1/H4 face and the H2/H3 face.[44] With two binding sites, it seems the FAT should have much stronger affinity than GIT1 PBD if the LD4 motif is the only binding site for PBD. With the LD2 motif as another binding site, GIT1 might achieve the comparable binding affinity by interacting with the two LD motifs simultaneously. However, both LD motifs share the same binding site; it seems not possible for them to bind PBD at the same time. This is true until we consider the fact that GIT1 can dimerize through its coiled-coil domain.[104] Thus, we propose that GIT1 interacts with paxillin as a dimer, with one protomer binds the LD2 motif while the other

binds the LD4 motif at the same time. The two LD motifs could belong to one paxillin or two copies of paxillin. The former possibility could lead to similar binding affinity comparable to the FAK-paxillin interaction; the later possibility could result in the cross-linking of paxillin molecules in the FA, which is even more exciting. Indeed, GIT1 is found primarily as an oligomer together with PIX,[59,104] and the oligomerization is essential for GIT1 FA localization.[105,106] Our hypothesis exactly explained the observations in that only the oligomerized GIT1 can provide multi LD binding PBDs that interact with a network of paxillins, locating themselves into the FA and clusterize the FA at the same time. This may also explain why previous studies only found the LD4 motif as the GIT1 binding site – the LD2 motif alone may not be strong enough for GIT1 binding.

5.2.3   *The effect of paxillin S272 phosphorylation*

Since the paxillin S272 phosphorylation promotes GIT1 FA localization,[89] the effect of S272 phosphorylation was also tested. We found that the LD4 S272p peptide bound GIT1 PBD in a manner very similar to the same length of LD4 peptide. The same set of residues had perturbation shifts in the same direction and distance. The final spectra of the PBD titrated with either LD4 or LD4p can be overlaid. Most peaks were in the same position, while a small number of peaks were in different positions. However, neither peptide produced a consistently larger chemical shift perturbation throughout the PBD binding site than the other, indicating the two peptides had very similar affinity to PBD. This notion was confirmed quantitively by the Biacore surface plasma resonance experiment. The Kd value for PBD binding with LD4 or LD4p was very close (7.2 μM and 10.2 μM, respectively).

There is phosphorylation induced LD4 helix unfolding in the interaction between LD4 and the FAT domain of FAK.[45] The same phosphorylation might induce helix unfolding too in the GIT1 interaction. Since GIT1 binds the same to LD4 and LD4p, there might be some other interaction to compensate for the entropy lost for LD4p. The most plausible theory is that the phosphate group of LD4p created more contacts which compensated for the entropy lost.

5.2.4   *Paxillin splice variants and short LD4 peptides*

The three paxillin splice variant differ right after the LD4 motif, so we tested GIT1 binding with the three LD4 peptides, of which the sequence is the same in the N terminus and different in the C terminus. Both NMR CSP titration and Biacore SPR showed that the three splice variants bound GIT1 PBD the same: same binding affinity, same binding site and same intermediate exchange phenomena. This is different from a previous report that the splice variants bind differently to FA proteins.[35] There are three reasons that might contribute to the discrepancy: 1. while the 3 spice variants bind the same to GIT1, they might interact differently to other paxillin LD4 binding FA proteins, e.g. FAK, actopaxin and PAK3; 2. the LD4 peptides might not be long enough to show

the difference; 3. the difference might be caused by conformational changes far into the insertions.

By serial deletion of the LD4 peptide at the C terminus, we identified the minimal LD4 length for GIT1 binding. The LD4 peptide ending at K277 is enough and sufficient to interact with GIT1 PBD. The deletion of two residues at the C terminal end dramatically decreased GIT1 binding. In the GIT1-LD4 complex structure, the two residues, F276 and K277, have extensive interactions with PBD. However, there is no interaction between GIT1 and the residues beyond K277. Thus the complex structure explained the reason why F276 and K277 are required for effective GIT1 binding and the reason why the LD4 splice variants did not exhibit any difference.

5.3   The GIT1-LD complex structures

5.3.1   The reliability of the complex structure models

Although accurate, HADDOCK can still make mistakes in generating the complex structure models. So the correctness of the model has to be validated before further analysis. We deem the complex structures correct and reliable. The reasons include:

1. The complex structures were validated by mutation data.[62] All the GIT1 mutations that proved to be important for paxillin binding involved residues that had extensive side chain contacts in the complex structures. All the GIT1 mutations that proved to be non-important for paxillin binding involved residues that had no contacts in the complex structures.

2. The binding site that was used to guide HADDOCK docking was obtained experimentally from the CSP titration experiments. Extensive binding site information was obtained for GIT1 PBD.

3. The structure of the binding partners, the GIT1 PBD and LD peptides, were experimentally determined by NMR, which is more accurate than homology models.

4. There was no conformational change upon interacting with both LD2 peptide and LD4 peptide. Since HADDOCK treats residues with very limited flexibility, this is an important prerequisite.

5. Both complex structures were small systems that had a smaller number of possible conformations to sample, so it was less challenging for HADDOCK to find the best solution.

6. During post docking analysis, the top lowest binding energy models were classified to the largest groups. In other words, the models were solutions that had the lowest binding energy and were similar to the majority of all the possible solutions.

5.3.2 *The reason why GIT1 PBD binds both LD2 and LD4 motifs*

The complex structures found that the sequence similarity of LD2 and LD4 motifs determined the ability for the two LD peptides to bind to GIT1 PBD. The interactions were primarily mediated by hydrophobic interactions with specificity mediated by side chain salt bridges and H-bonds.

The two LD peptides had well conserved hydrophobic side chains. As aligned in Figure 3-8c, the side chains of Φ-3, Φ+0, Φ +3, Φ +4, Φ +7 and Φ +10 positions were highly conserved in both properties and positions in the helices. In the complex structures, all of those residues were inserted into the hydrophobic groove between H1 and H4. So the conservation of the hydrophobic residues provided the basis for both LD peptides to bind to the PBD H1/H4 site.

Besides the hydrophobic residues, the conservation of the charged and polar residues further contributed to the fact that both LD peptides can bind to GIT1. For example, the Glu at Φ-1 and the Asp at Φ+1 had the conserved salt bridges to GIT1 K663 and K758, respectively. Also, the H-bond between GIT1 Q666 and paxillin LD Φ+6 residue were conserved too.

With the conservation of the large hydrophobic contacts, the well defined salt bridges and H-bonds, the LD2 and LD4 peptides were very alike. Thus both LD2 and LD4 motifs can interact with GIT1 PBD at the H1/H4 site. It also explained why the two peptides perturbed the same set of residues in a very similar manner in the CSP titration experiments.

5.3.3 *The specificity compared to FAT domain of FAK*

Since there are so many similarities between the two LD motifs, why does the H1/H4 site of FAK FAT domain only interact with the LD2 motif? What determines the specificity of GIT1 PBD versus the FAK FAT domain?

The answer might be the subtle difference of the H1/H4 surface. Overall the two surfaces are very similar as both featured a hydrophobic groove sandwiched by two polar and positive charged ridges. However, the FAK H1/H4 has an additional positive charged patch at the upper end of helix 4 (K1018 and K1019). The same positions on the GIT1 surface are T743 and Q744 which accommodate well the LD4 K277 residue. So this may explain why the LD4 motif can bind to the H1/H4 site of GIT1, but not the H1/H4 site of FAK FAT.

5.4 Implications for the understanding of FA disassembly

Focal adhesion disassembly, or FA turn over, is important to ensure cells switching from stationary to motile. It is as important as FA assembly if not more important. However current research has focused more on the FA assembly. The mechanism of FA disassembly is just starting to be unveiled. More and more evidence indicates that PAK is one of the key regulators for FA disassembly,[12,45,89] Combining previous knowledge and discoveries in this dissertation, we proposed a model of FA disassembly initiation (Figure 5-1).

This model involves the switch of assembly activator FAK to the disassembly activator PAK, under the control of paxillin S272 phosphorylation. Detailed discussion follows the Figure 5-1.

1. FAK and GIT depend on paxillin for FA localization. The interactions between paxillin LD motifs and the FAK FAT domain or the GIT1 PBD are critical for their FA targeting. At the beginning of FA assembly, FAK is recruited to FA through paxillin. The FAT domain of FAK binds paxillin LD2 and LD4 motifs simultaneously, with LD2 binding to the H1/H4 site and LD4 binding to the H2/H3 site. The interaction is cooperative at two helix-helix interfaces, bearing a large binding energy.[31,44] FAK is one of the master activators of FA assembly. The recruitment of FAK activates multiple downstream signaling, resulting in the FA assembly at cell leading edge as well as the cell survival and proliferation.

2. The paxillin S272 (in the LD4 motif) phosphorylation is the major switch event to change FA assembly to FA disassembly. The kinase that phosphorylates the S272 is still unknown. The phosphorylation of S272 leads to the dissociation of paxillin LD4 from the binding site at the FAT H2/H3 face. The mechanism involves the unfolding of the LD4 motif by the S272 phosphorylation.[45] Without the interaction at the LD4 motif, the cooperative binding at the LD2 motif may not be strong enough to compete with other paxillin binding partners, in particular, the GIT1.

3. GIT1 PBD is capable of binding to either the LD2 or the LD4 motif with roughly the same affinity. And the LD4 S272 phosphorylation does not influence the PBD interaction. So, without FAK occupying the LD motifs, GIT1 is able to bind to paxillin, through either LD2 or LD4 or LD4p motifs. Considering the fact the GIT1 is primarily found to be dimerized in the FA, we also propose that the interaction involves the one molecule of paxillin and a dimer of GIT1 (it could be a GIT1 homodimer or heterodimer with GIT2). Thus, one copy of GIT1 binds to the LD2 motif and the other GIT1 binds to the LD4 motif, achieving a similar binding affinity as the FAK-paxillin interaction. The recruitment of GIT1 into FA also brings in PAK and PIX via the interaction of GIT1-PIX-PAK. PAK is one of the master activators to promote focal adhesion disassembly. The PAK activity can activate downstream signaling to promote FA disassembly.
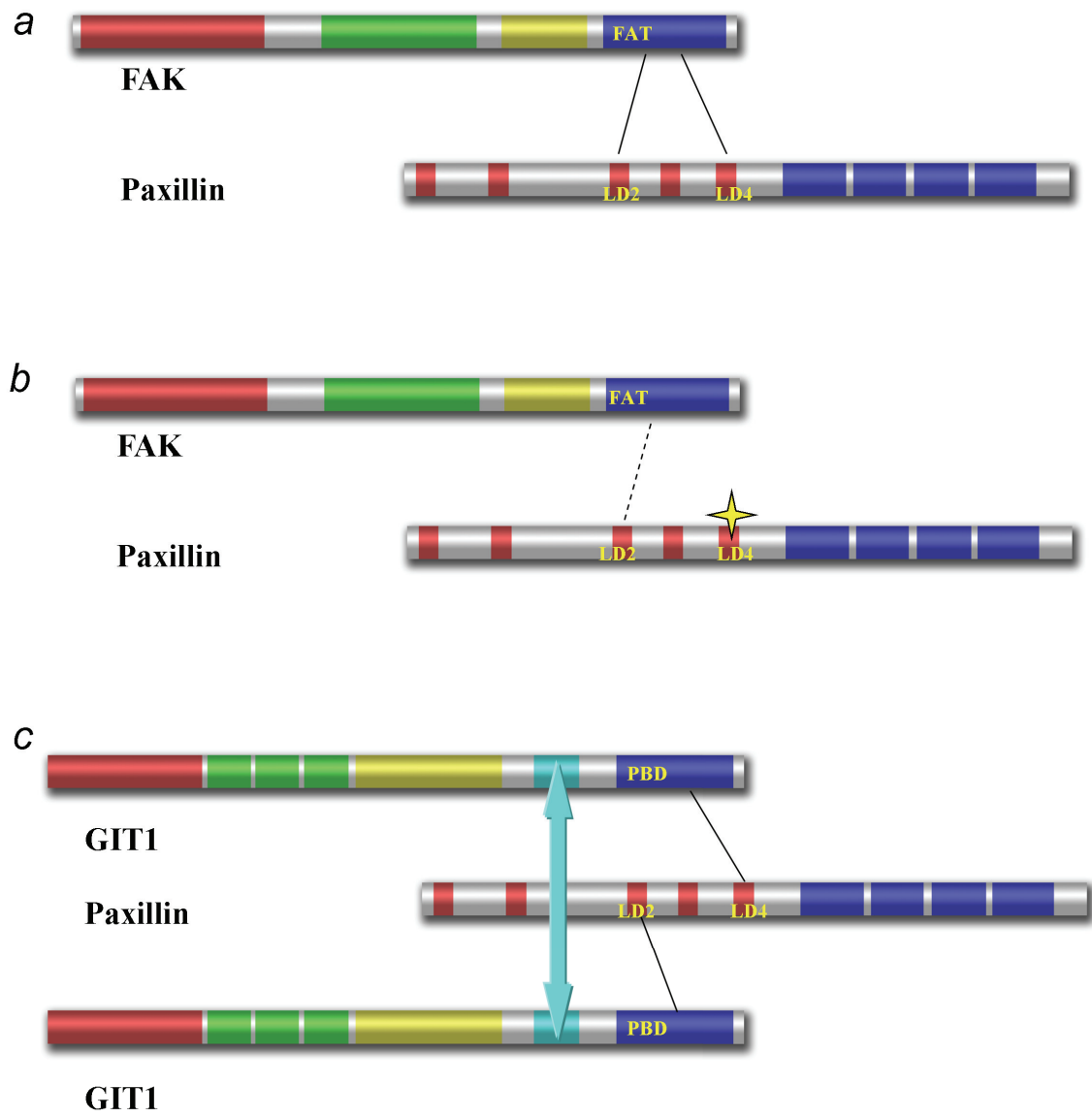
Figure 5-1   Model of focal adhesion disassembly initiation.

Thus, cell achieves FA assembly to the disassembly switch through the paxillin phosphorylation and the recruitment of different activation complexes. The paxillin S272 phosphorylation shifts the balance of paxillin interaction with FAK versus GIT1 in favor of GIT1. This would decrease FAK while increasing PAK in the FA; the balance shift may promote regulated FA assembly and disassembly cycle.

This model integrates existing data to provide a reasonable working theory for FA disassembly, a critical cell function. However, there are still a lot of questions to be answered, for example, which kinase phosphorylates the S272 and which phosphatase reverses the S272p and at what time point? Nevertheless, the structural and functional studies of GIT1 PBD presented in this dissertation shed more light on the function of GIT proteins in the integration of focal adhesion signaling.

LIST OF REFERENCES

1.   Lauffenburger, D.A. & Horwitz, A.F. Cell migration: a physically integrated molecular process. *Cell* **84**, 359-69 (1996).

2.   Zamir, E. & Geiger, B. Components of cell-matrix adhesions. *J Cell Sci* **114**, 3577-9 (2001).

3.   Lo, S.H. Focal adhesions: what's new inside. *Dev Biol* **294**, 280-91 (2006).

4.   Cohen, M., Joester, D., Geiger, B. & Addadi, L. Spatial and temporal sequence of events in cell adhesion: from molecular recognition to focal adhesion assembly. *Chembiochem* **5**, 1393-9 (2004).

5.   Ponta, H., Sherman, L. & Herrlich, P.A. CD44: from adhesion molecules to signalling regulators. *Nat Rev Mol Cell Biol* **4**, 33-45 (2003).

6.   Vestweber, D. & Blanks, J.E. Mechanisms that regulate the function of the selectins and their ligands. *Physiol Rev* **79**, 181-213 (1999).

7.   Small, J.V., Stradal, T., Vignal, E. & Rottner, K. The lamellipodium: where motility begins. *Trends Cell Biol* **12**, 112-20 (2002).

8.   Koleske, A.J. Do filopodia enable the growth cone to find its way? *Sci STKE* **2003**, pe20 (2003).

9.   Oliver, T., Dembo, M. & Jacobson, K. Separation of propulsive and adhesive traction stresses in locomoting keratocytes. *J Cell Biol* **145**, 589-604 (1999).

10.   Beglova, N., Blacklow, S.C., Takagi, J. & Springer, T.A. Cysteine-rich module structure reveals a fulcrum for integrin rearrangement upon activation. *Nat Struct Biol* **9**, 282-7 (2002).

11.   Shimaoka, M. et al. Stabilizing the integrin alpha M inserted domain in alternative conformations with a range of engineered disulfide bonds. *Proc Natl Acad Sci U S A* **99**, 16737-41 (2002).

12.   Zhao, Z.S., Manser, E., Loo, T.H. & Lim, L. Coupling of PAK-interacting exchange factor PIX to GIT1 promotes focal complex disassembly. *Mol Cell Biol* **20**, 6354-63 (2000).

13.   Manser, E. et al. Expression of constitutively active alpha-PAK reveals effects of the kinase on actin and focal complexes. *Mol Cell Biol* **17**, 1129-43 (1997).

14.   Frost, J.A., Khokhlatchev, A., Stippec, S., White, M.A. & Cobb, M.H. Differential effects of PAK1-activating mutations reveal activity-dependent and - independent effects on cytoskeletal regulation. *J Biol Chem* **273**, 28191-8 (1998).

15.   Kiosses, W.B., Daniels, R.H., Otey, C., Bokoch, G.M. & Schwartz, M.A. A role for p21-activated kinase in endothelial cell migration. *J Cell Biol* **147**, 831-44 (1999).

16.   Choquet, D., Felsenfeld, D.P. & Sheetz, M.P. Extracellular matrix rigidity causes strengthening of integrin-cytoskeleton linkages. *Cell* **88**, 39-48 (1997).

17.   Harris, A.K., Wild, P. & Stopak, D. Silicone rubber substrata: a new wrinkle in the study of cell locomotion. *Science* **208**, 177-9 (1980).

18.   Balaban, N.Q. et al. Force and focal adhesion assembly: a close relationship studied using elastic micropatterned substrates. *Nat Cell Biol* **3**, 466-72 (2001).

19.   Bershadsky, A.D. et al. Assembly and mechanosensory function of focal adhesions: experiments and models. *Eur J Cell Biol* **85**, 165-73 (2006).

20.   Moro, L. et al. Integrin-induced epidermal growth factor (EGF) receptor activation requires c-Src and p130Cas and leads to phosphorylation of specific EGF receptor tyrosines. *J Biol Chem* **277**, 9405-14 (2002).

21.   Borges, E., Jan, Y. & Ruoslahti, E. Platelet-derived growth factor receptor beta and vascular endothelial growth factor receptor 2 bind to the beta 3 integrin through its extracellular domain. *J Biol Chem* **275**, 39867-73 (2000).

22.   Schneller, M., Vuori, K. & Ruoslahti, E. Alphavbeta3 integrin associates with activated insulin and PDGFbeta receptors and potentiates the biological activity of PDGF. *EMBO J* **16**, 5600-7 (1997).

23.   Hood, J.D., Frausto, R., Kiosses, W.B., Schwartz, M.A. & Cheresh, D.A. Differential alphav integrin-mediated Ras-ERK signaling during two pathways of angiogenesis. *J Cell Biol* **162**, 933-43 (2003).

24.   Breitschopf, K., Haendeler, J., Malchow, P., Zeiher, A.M. & Dimmeler, S. Posttranslational modification of Bcl-2 facilitates its proteasome-dependent degradation: molecular characterization of the involved signaling pathway. *Mol Cell Biol* **20**, 1886-96 (2000).

25.   Stromblad, S., Becker, J.C., Yebra, M., Brooks, P.C. & Cheresh, D.A. Suppression of p53 activity and p21WAF1/CIP1 expression by vascular cell integrin alphaVbeta3 during angiogenesis. *J Clin Invest* **98**, 426-33 (1996).

26. Hynes, R.O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673-87 (2002).

27. Xiong, J.P. et al. Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science* **294**, 339-45 (2001).

28. Turner, C.E., Glenney, J.R., Jr. & Burridge, K. Paxillin: a new vinculin-binding protein present in focal adhesions. *J Cell Biol* **111**, 1059-68 (1990).

29. Brown, M.C., Curtis, M.S. & Turner, C.E. Paxillin LD motifs may define a new family of protein recognition domains. *Nat Struct Biol* **5**, 677-8 (1998).

30. Dawid, I.B., Breen, J.J. & Toyama, R. LIM domains: multiple roles as adapters and functional modifiers in protein interactions. *Trends Genet* **14**, 156-62 (1998).

31. Liu, G., Guibao, C.D. & Zheng, J. Structural insight into the mechanisms of targeting and signaling of focal adhesion kinase. *Mol Cell Biol* **22**, 2751-60 (2002).

32. Hoellerer, M.K. et al. Molecular recognition of paxillin LD motifs by the focal adhesion targeting domain. *Structure* **11**, 1207-17 (2003).

33. Brown, M.C., Perrotta, J.A. & Turner, C.E. Identification of LIM3 as the principal determinant of paxillin focal adhesion localization and characterization of a novel motif on paxillin directing vinculin and focal adhesion kinase binding. *J Cell Biol* **135**, 1109-23 (1996).

34. Brown, M.C., Perrotta, J.A. & Turner, C.E. Serine and threonine phosphorylation of the paxillin LIM domains regulates paxillin focal adhesion localization and cell adhesion to fibronectin. *Mol Biol Cell* **9**, 1803-16 (1998).

35. Mazaki, Y., Hashimoto, S. & Sabe, H. Monocyte cells and cancer cells express novel paxillin isoforms with different binding properties to focal adhesion proteins. *J Biol Chem* **272**, 7437-44 (1997).

36. Nikolopoulos, S.N. & Turner, C.E. Actopaxin, a new focal adhesion protein that binds paxillin LD motifs and actin and regulates cell adhesion. *J Cell Biol* **151**, 1435-48 (2000).

37. Nikolopoulos, S.N. & Turner, C.E. Integrin-linked kinase (ILK) binding to paxillin LD1 motif regulates ILK localization to focal adhesions. *J Biol Chem* **276**, 23499-505 (2001).

38. Turner, C.E. et al. Paxillin LD4 motif binds PAK and PIX through a novel 95-kD ankyrin repeat, ARF-GAP protein: A role in cytoskeletal remodeling. *J Cell Biol* **145**, 851-63 (1999).

39. Tong, X. & Howley, P.M. The bovine papillomavirus E6 oncoprotein interacts with paxillin and disrupts the actin cytoskeleton. *Proc Natl Acad Sci U S A* **94**, 4412-7 (1997).

40. Hashimoto, S., Tsubouchi, A., Mazaki, Y. & Sabe, H. Interaction of paxillin with p21-activated Kinase (PAK). Association of paxillin alpha with the kinase-inactive and the Cdc42-activated forms of PAK3. *J Biol Chem* **276**, 6037-45 (2001).

41. Mitra, S.K., Hanson, D.A. & Schlaepfer, D.D. Focal adhesion kinase: in command and control of cell motility. *Nat Rev Mol Cell Biol* **6**, 56-68 (2005).

42. Hauck, C.R., Hsia, D.A., Puente, X.S., Cheresh, D.A. & Schlaepfer, D.D. FRNK blocks v-Src-stimulated invasion and experimental metastases without effects on cell motility or growth. *EMBO J* **21**, 6289-302 (2002).

43. Sieg, D.J. et al. FAK integrates growth-factor and integrin signals to promote cell migration. *Nat Cell Biol* **2**, 249-56 (2000).

44. Bertolucci, C.M., Guibao, C.D. & Zheng, J. Structural features of the focal adhesion kinase-paxillin complex give insight into the dynamics of focal adhesion assembly. *Protein Sci* **14**, 644-52 (2005).

45. Bertolucci, C.M., Guibao, C.D. & Zheng, J.J. Phosphorylation of Paxillin LD4 Destabilizes Helix Formation and Inhibits Binding to Focal Adhesion Kinase. *Biochemistry* (2007).

46. Goehler, H. et al. A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol Cell* **15**, 853-65 (2004).

47. Blagoveshchenskaya, A.D., Thomas, L., Feliciangeli, S.F., Hung, C.H. & Thomas, G. HIV-1 Nef downregulates MHC-I by a PACS-1- and PI3K-regulated ARF6 endocytic pathway. *Cell* **111**, 853-66 (2002).

48. Brown, A., Wang, X., Sawai, E. & Cheng-Mayer, C. Activation of the PAK-related kinase by human immunodeficiency virus type 1 Nef in primary human peripheral blood lymphocytes and macrophages leads to phosphorylation of a PIX-p95 complex. *J Virol* **73**, 9899-907 (1999).

49. Hoefen, R.J. & Berk, B.C. The multifunctional GIT family of proteins. *J Cell Sci* **119**, 1469-75 (2006).

50. Mazaki, Y. et al. An ADP-ribosylation factor GTPase-activating protein Git2-short/KIAA0148 is involved in subcellular localization of paxillin and actin cytoskeletal organization. *Mol Biol Cell* **12**, 645-62 (2001).

51.   Vitale, N. et al. GIT proteins, A novel family of phosphatidylinositol 3,4, 5-trisphosphate-stimulated GTPase-activating proteins for ARF6. *J Biol Chem* **275**, 13901-6 (2000).

52.   Premont, R.T. et al. beta2-Adrenergic receptor regulation by GIT1, a G protein-coupled receptor kinase-associated ADP ribosylation factor GTPase-activating protein. *Proc Natl Acad Sci U S A* **95**, 14082-7 (1998).

53.   Premont, R.T., Claing, A., Vitale, N., Perry, S.J. & Lefkowitz, R.J. The GIT family of ADP-ribosylation factor GTPase-activating proteins. Functional diversity of GIT2 through alternative splicing. *J Biol Chem* **275**, 22373-80 (2000).

54.   Di Cesare, A. et al. p95-APP1 links membrane transport to Rac-mediated reorganization of actin. *Nat Cell Biol* **2**, 521-30 (2000).

55.   Paris, S., Za, L., Sporchia, B. & de Curtis, I. Analysis of the subcellular distribution of avian p95-APP2, an ARF-GAP orthologous to mammalian paxillin kinase linker. *Int J Biochem Cell Biol* **34**, 826-37 (2002).

56.   Bagrodia, S. et al. A tyrosine-phosphorylated protein that binds to an important regulatory region on the cool family of p21-activated kinase-binding proteins. *J Biol Chem* **274**, 22393-400 (1999).

57.   Yin, G., Haendeler, J., Yan, C. & Berk, B.C. GIT1 functions as a scaffold for MEK1-extracellular signal-regulated kinase 1 and 2 activation by angiotensin II and epidermal growth factor. *Mol Cell Biol* **24**, 875-85 (2004).

58.   Schmitz, U., Ishida, M. & Berk, B.C. Angiotensin II stimulates tyrosine phosphorylation of phospholipase C-gamma-associated proteins. Characterization of a c-Src-dependent 97-kD protein in vascular smooth muscle cells. *Circ Res* **81**, 550-7 (1997).

59.   Premont, R.T. et al. The GIT/PIX complex: an oligomeric assembly of GIT family ARF GTPase-activating proteins and PIX family Rac1/Cdc42 guanine nucleotide exchange factors. *Cell Signal* **16**, 1001-11 (2004).

60.   Brown, M.C., West, K.A. & Turner, C.E. Paxillin-dependent paxillin kinase linker and p21-activated kinase localization to focal adhesions involves a multistep activation pathway. *Mol Biol Cell* **13**, 1550-65 (2002).

61.   West, K.A. et al. The LD4 motif of paxillin regulates cell spreading and motility through an interaction with paxillin kinase linker (PKL). *J Cell Biol* **154**, 161-76 (2001).

62.   Schmalzigaug, R. et al. GIT1 utilizes a focal adhesion targeting-homology domain to bind paxillin. *Cell Signal* **19**, 1733-44 (2007).

109

63.   RCSB PDB statistics. (www.RCSB.org, 2008). Accessed on March 4, 2008.

64.   Jennison, R.C. *Fourier Transforms and Convolutions*, (Pergamon Press, NY, 1961).

65.   Masse, J.E. & Keller, R. AutoLink: automated sequential resonance assignment of biopolymers from NMR data by relative-hypothesis-prioritization-based simulated logic. *J Magn Reson* **174**, 133-51 (2005).

66.   Keller, R. *The Computer Aided Resonance Assignment Tutorial* (CANTINA Verlag, 2004).

67.   Guntert, P. Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-78 (2004).

68.   Guntert, P. Structure calculation of biological macromolecules from NMR data. *Q Rev Biophys* **31**, 145-237 (1998).

69.   Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**, 209-27 (2002).

70.   Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* **24**, 171-89 (2002).

71.   Braun, W. Distance geometry and related methods for protein structure determination from NMR data. *Q Rev Biophys* **19**, 115-57 (1987).

72.   Kaptein, R., Zuiderweg, E.R., Scheek, R.M., Boelens, R. & van Gunsteren, W.F. A protein structure from nuclear magnetic resonance data. lac repressor headpiece. *J Mol Biol* **182**, 179-82 (1985).

73.   Havel, T.F. & Wuthrich, K. An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *J Mol Biol* **182**, 281-94 (1985).

74.   Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289-302 (1999).

75.   Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. & Shimada, I. A novel NMR method for determining the interfaces of large protein-protein complexes. *Nat Struct Biol* **7**, 220-3 (2000).

76. Worrall, J.A., Reinle, W., Bernhardt, R. & Ubbink, M. Transient protein interactions studied by NMR spectroscopy: the case of cytochrome C and adrenodoxin. *Biochemistry* **42**, 7068-76 (2003).

77. Dominguez, C., Boelens, R. & Bonvin, A.M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* **125**, 1731-7 (2003).

78. Brunger, A.T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**, 905-21 (1998).

79. Rieping, W. et al. ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23**, 381-2 (2007).

80. Jockusch, B.M. et al. The molecular architecture of focal adhesions. *Annu Rev Cell Dev Biol* **11**, 379-416 (1995).

81. Sastry, S.K. & Burridge, K. Focal adhesions: a nexus for intracellular signaling and cytoskeletal dynamics. *Exp Cell Res* **261**, 25-36 (2000).

82. Sastry, S.K. et al. Quantitative changes in integrin and focal adhesion signaling regulate myoblast cell cycle withdrawal. *J Cell Biol* **144**, 1295-309 (1999).

83. Ilic, D. et al. Extracellular matrix survival signals transduced by focal adhesion kinase suppress p53-mediated apoptosis. *J Cell Biol* **143**, 547-60 (1998).

84. Cary, L.A. & Guan, J.L. Focal adhesion kinase in integrin-mediated signaling. *Front Biosci* **4**, D102-13 (1999).

85. Manser, E., Leung, T., Salihuddin, H., Zhao, Z.S. & Lim, L. A brain serine/threonine protein kinase activated by Cdc42 and Rac1. *Nature* **367**, 40-6 (1994).

86. Ridley, A.J. & Hall, A. The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70**, 389-99 (1992).

87. Ridley, A.J., Paterson, H.F., Johnston, C.L., Diekmann, D. & Hall, A. The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. *Cell* **70**, 401-10 (1992).

88. Nobes, C.D. & Hall, A. Rho, rac and cdc42 GTPases: regulators of actin structures, cell adhesion and motility. *Biochem Soc Trans* **23**, 456-9 (1995).

89.   Nayal, A. et al. Paxillin phosphorylation at Ser273 localizes a GIT1-PIX-PAK complex and regulates adhesion and protrusion dynamics. *J Cell Biol* **173**, 587-9 (2006).

90.   Brown, M.C. & Turner, C.E. Paxillin: adapting to change. *Physiol Rev* **84**, 1315-39 (2004).

91.   Arold, S.T., Hoellerer, M.K. & Noble, M.E. The structural basis of localization and signaling by the focal adhesion targeting domain. *Structure* **10**, 319-27 (2002).

92.   Thomas, J.W. et al. The role of focal adhesion kinase binding in the regulation of tyrosine phosphorylation of paxillin. *J Biol Chem* **274**, 36684-92 (1999).

93.   Neidhardt, F.C., Bloch, P.L. & Smith, D.F. Culture medium for enterobacteria. *J Bacteriol* **119**, 736-47 (1974).

94.   Goddard, T.D. & Kneller, D.G. SPARKY 3. (University of California, San Francisco).

95.   Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93 (1995).

96.   Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291 (1993).

97.   Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**, 51-5, 29-32 (1996).

98.   Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-96 (1991).

99.   Guex, N. & Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-23 (1997).

100.   Sharma, D. & Rajarathnam, K. 13C NMR chemical shifts can predict disulfide bond formation. *J Biomol NMR* **18**, 165-71 (2000).

101.   Prutzman, K.C. et al. The focal adhesion targeting domain of focal adhesion kinase contains a hinge region that modulates tyrosine 926 phosphorylation. *Structure* **12**, 881-91 (2004).

102.   Chothia, C., Levitt, M. & Richardson, D. Helix to helix packing in proteins. *J Mol Biol* **145**, 215-50 (1981).

103.   Hayashi, I., Vuori, K. & Liddington, R.C. The focal adhesion targeting (FAT) region of focal adhesion kinase is a four-helix bundle that binds paxillin. *Nat Struct Biol* **9**, 101-6 (2002).

104.   Paris, S., Longhi, R., Santambrogio, P. & de Curtis, I. Leucine-zipper-mediated homo- and hetero-dimerization of GIT family p95-ARF GTPase-activating protein, PIX-, paxillin-interacting proteins 1 and 2. *Biochem J* **372**, 391-8 (2003).

105.   Kim, S., Lee, S.H. & Park, D. Leucine zipper-mediated homodimerization of the p21-activated kinase-interacting factor, beta Pix. Implication for a role in cytoskeletal reorganization. *J Biol Chem* **276**, 10581-4 (2001).

106.   Loo, T.H., Ng, Y.W., Lim, L. & Manser, E. GIT1 activates p21-activated kinase through a mechanism independent of p21 binding. *Mol Cell Biol* **24**, 3849-59 (2004).

107.   Manabe, R., Kovalenko, M., Webb, D.J. & Horwitz, A.R. GIT1 functions in a motile, multi-molecular signaling complex that regulates protrusive activity and cell migration. *J Cell Sci* **115**, 1497-510 (2002).

108.   Zhang, Z.M., Simmerman, J.A., Guibao, C.D. & Zheng, J.J. GIT1 paxillin-binding domain is a four-helix bundle and it binds to both paxillin LD2 and LD4 motifs. *J Biol Chem* (2008).

109.   Schaller, M.D. Paxillin: a focal adhesion-associated adaptor protein. *Oncogene* **20**, 6459-72 (2001).

110.   Bertolucci, C.M., Guibao, C.D. & Zheng, J.J. Phosphorylation of paxillin LD4 destabilizes helix formation and inhibits binding to focal adhesion kinase. *Biochemistry* **47**, 548-54 (2008).

111.   Briknarova, K. et al. The serine-rich domain from Crk-associated substrate (p130cas) is a four-helix bundle. *J Biol Chem* **280**, 21908-14 (2005).

112.   Fillingham, I. et al. A vinculin binding domain from the talin rod unfolds to form a complex with the vinculin head. *Structure* **13**, 65-74 (2005).

113.   Gingras, A.R. et al. Structural and dynamic characterization of a vinculin binding site in the talin rod. *Biochemistry* **45**, 1805-17 (2006).

114.   Yang, J., Dokurno, P., Tonks, N.K. & Barford, D. Crystal structure of the M-fragment of alpha-catenin: implications for modulation of cell adhesion. *Embo J* **20**, 3645-56 (2001).

115. Schmalzigaug, R., Phee, H., Davidson, C.E., Weiss, A. & Premont, R.T. Differential Expression of the ARF GAP Genes GIT1 and GIT2 in Mouse Tissues. *J Histochem Cytochem* **55**, 1039-48 (2007).

116. Brown, M.C., Cary, L.A., Jamieson, J.S., Cooper, J.A. & Turner, C.E. Src and FAK kinases cooperate to phosphorylate paxillin kinase linker, stimulate its focal adhesion localization, and regulate cell spreading and protrusiveness. *Mol Biol Cell* **16**, 4316-28 (2005).

117. Artavanis-Tsakonas, S., Rand, M.D. & Lake, R.J. Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770-6 (1999).

118. Greenwald, I. LIN-12/Notch signaling: lessons from worms and flies. *Genes Dev* **12**, 1751-62 (1998).

119. Egan, S.E., St-Pierre, B. & Leow, C.C. Notch receptors, partners and regulators: from conserved domains to powerful functions. *Curr Top Microbiol Immunol* **228**, 273-324 (1998).

120. Maillard, I. & Pear, W.S. Notch and cancer: best to avoid the ups and downs. *Cancer Cell* **3**, 203-5 (2003).

121. Gridley, T. Notch signaling and inherited disease syndromes. *Hum Mol Genet* **12 Spec No 1**, R9-13 (2003).

122. Mumm, J.S. & Kopan, R. Notch signaling: from the outside in. *Dev Biol* **228**, 151-65 (2000).

123. Yancopoulos, G.D. et al. Vascular-specific growth factors and blood vessel formation. *Nature* **407**, 242-8 (2000).

124. Nakagawa, O. et al. Members of the HRT family of basic helix-loop-helix proteins act as transcriptional repressors downstream of Notch signaling. *Proc Natl Acad Sci U S A* **97**, 13655-60 (2000).

125. Matsuno, K., Go, M.J., Sun, X., Eastman, D.S. & Artavanis-Tsakonas, S. Suppressor of Hairless-independent events in Notch signaling imply novel pathway elements. *Development* **124**, 4265-73 (1997).

126. Ordentlich, P. et al. Notch inhibition of E47 supports the existence of a novel signaling pathway. *Mol Cell Biol* **18**, 2230-9 (1998).

127. Jehn, B.M., Bielke, W., Pear, W.S. & Osborne, B.A. Cutting edge: protective effects of notch-1 on TCR-induced apoptosis. *J Immunol* **162**, 635-8 (1999).

128. Shawber, C.J. & Kitajewski, J. Notch function in the vasculature: insights from zebrafish, mouse and man. *Bioessays* **26**, 225-34 (2004).

129.   Villa, N. et al. Vascular expression of Notch pathway receptors and ligands is restricted to arterial vessels. *Mech Dev* **108**, 161-4 (2001).

130.   Krebs, L.T. et al. Notch signaling is essential for vascular morphogenesis in mice. *Genes Dev* **14**, 1343-52 (2000).

131.   Del Amo, F.F. et al. Expression pattern of Motch, a mouse homolog of Drosophila Notch, suggests an important role in early postimplantation mouse development. *Development* **115**, 737-44 (1992).

132.   Reaume, A.G., Conlon, R.A., Zirngibl, R., Yamaguchi, T.P. & Rossant, J. Expression analysis of a Notch homologue in the mouse embryo. *Dev Biol* **154**, 377-87 (1992).

133.   Loomes, K.M. et al. Characterization of Notch receptor expression in the developing mammalian heart and liver. *Am J Med Genet* **112**, 181-9 (2002).

134.   McCright, B. et al. Defects in development of the kidney, heart and eye vasculature in mice homozygous for a hypomorphic Notch2 mutation. *Development* **128**, 491-502 (2001).

135.   Xue, Y. et al. Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. *Hum Mol Genet* **8**, 723-30 (1999).

136.   Joutel, A. et al. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature* **383**, 707-10 (1996).

137.   Oda, T. et al. Mutations in the human Jagged1 gene are responsible for Alagille syndrome. *Nat Genet* **16**, 235-42 (1997).

138.   Young, L.S. & Rickinson, A.B. Epstein-Barr virus: 40 years on. *Nat Rev Cancer* **4**, 757-68 (2004).

139.   Henkel, T., Ling, P.D., Hayward, S.D. & Peterson, M.G. Mediation of Epstein-Barr virus EBNA2 transactivation by recombination signal-binding protein J kappa. *Science* **265**, 92-5 (1994).

140.   Grossman, S.R., Johannsen, E., Tong, X., Yalamanchili, R. & Kieff, E. The Epstein-Barr virus nuclear antigen 2 transactivator is directed to response elements by the J kappa recombination signal binding protein. *Proc Natl Acad Sci U S A* **91**, 7568-72 (1994).

141.   Waltzer, L. et al. The human J kappa recombination signal sequence binding protein (RBP-J kappa) targets the Epstein-Barr virus EBNA2 protein to its DNA responsive elements. *Embo J* **13**, 5633-8 (1994).

142.   Zimber-Strobl, U. et al. Epstein-Barr virus nuclear antigen 2 exerts its transactivating function through interaction with recombination signal binding protein RBP-J kappa, the homologue of Drosophila Suppressor of Hairless. *Embo J* **13**, 4973-82 (1994).

143.   Rabson, M., Gradoville, L., Heston, L. & Miller, G. Non-immortalizing P3J-HR-1 Epstein-Barr virus: a deletion mutant of its transforming parent, Jijoye. *J Virol* **44**, 834-44 (1982).

144.   Cohen, J.I., Wang, F., Mannick, J. & Kieff, E. Epstein-Barr virus nuclear protein 2 is a key determinant of lymphocyte transformation. *Proc Natl Acad Sci U S A* **86**, 9558-62 (1989).

145.   Farrell, C.J. et al. Inhibition of Epstein-Barr virus-induced growth proliferation by a nuclear antigen EBNA2-TAT peptide. *Proc Natl Acad Sci U S A* **101**, 4625-30 (2004).

146.   Kovall, R.A. & Hendrickson, W.A. Crystal structure of the nuclear effector of Notch signaling, CSL, bound to DNA. *Embo J* **23**, 3441-51 (2004).

147.   Wilson, J.J. & Kovall, R.A. Crystal structure of the CSL-Notch-Mastermind ternary complex bound to DNA. *Cell* **124**, 985-96 (2006).

148.   Hsieh, J.J., Nofziger, D.E., Weinmaster, G. & Hayward, S.D. Epstein-Barr virus immortalization: Notch2 interacts with CBF1 and blocks differentiation. *J Virol* **71**, 1938-45 (1997).

149.   Fuchs, K.P. et al. Mutational analysis of the J recombination signal sequence binding protein (RBP-J)/Epstein-Barr virus nuclear antigen 2 (EBNA2) and RBP-J/Notch interaction. *Eur J Biochem* **268**, 4639-46 (2001).

APPENDIX A. A STRUCTURAL AND FUNCTIONAL STUDY ON THE INTERACTION BETWEEN EBNA2 AND NOTCH EFFECTOR CSL

A.1   Introduction

A.1.1   Notch signaling pathway

Notch signaling is evolutionarily conserved and involves cell fate decision in metazoans through intercellular communications.[117] This signaling pathway is widely used in organisms from *C. elegans* to human. The interaction of trans-membrane receptor Notch with its ligand DSL (Delta, Serrate, Lag-2) on the neighboring cell surface activates the intracellular signaling pathway and controls binary cell-fate determination.[117-119] Various cancers and developmental disorders have been found to relate to abnormal Notch signaling.[120,121]

CSL is the collective name of CBF1/RBP-Jκ, Su(H), Lag-1 for mammalian, *Drosophila melanogaster* and *Caenorhabditis elegans* orthologs, respectively. It is the central nuclear effector in the canonical Notch signaling pathway. In the nucleus, CSL specifically binds to target DNA and acts as an expression switch in working with different co-regulators. Without an activation signal, CSL binds co-repressor complexes and keeps its target gene silent. Upon activation by the ligand DSL from the adjacent cell, receptor Notch is cleaved by presenilin-dependent proteolysis and the Notch intracellular domain (NICD) is released.[122] In the nucleus, migrated NICD binds CSL and results in the displacement of co-repressors by co-activators, which then turns on transcription by chromatin remodeling.[122] The products of CSL controlled genes belong to the basic Helix-Loop-Helix (bHLH) family of transcription repressors, including Hairy/Enhancer of Split (HES) and HES related repressor protein (HERP).[123,124] HES and HERP are transcription factors themselves and further control downstream gene expression in different contexts. Although there is evidence for CSL independent Notch signaling,[125-128] CSL is still believed to be the major effecter of the Notch pathway.[117]

Notch signaling plays an essential role in development. Take the vascular system for example, Notch components are found widely expressed in various vessels, including system artery, pulmonary artery and veins.[129-133] In mice, mutation or knockout of different Notch components results in multiple system development disorders. Most of the investigations found embryonic lethality with vascular development failure.[130,134,135] In humans, two diseases caused by Notch component deficiency were recorded to have cardiovascular system abnormality: Alagille syndrome (AGS) and Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL) with Notch ligand Jagged1 and receptor Notch3 mutation, respectively.[13,136,137] The current opinion on Notch function in vascular development is that it is dispensable in vasculogenesis, a process of forming the primary homogeneously sized vascular plexus, but essential for angiogenesis, a remodeling step following vasculogenesis.[128]

A.1.2   Epstein-Barr virus transforms B lymphocytes by hijacking the Notch pathway

Epstein-Barr Virus (EBV) is a common human herpesvirus that is the major cause of infectious mononucleosis and is tightly associated with several cancers, especially B cell lymphomas in immunocompromised individuals.[138] EBV achieves strong transforming potential by hijacking the Notch signaling pathway. EBV Nuclear Antigen 2 (EBNA2) is a latent viral gene product that interacts with the host Notch nuclear effector CSL.[139-142] Three layers of evidence proved that the EBNA2-CSL interaction is essential for EBV associated B cell transformation: while an EBNA2 gene deleted EBV strain can not transform B cells;[143] restorations of the EBNA2 gene restores transformation function.[144] Furthermore, disruption of EBNA2-CSL interaction inhibits EBV induced growth proliferation.[145]

Thus, both NICD and EBNA2 can bind the Notch nuclear effector CSL and both interactions have critical functions. The crystal structure of *C. elegans* Lag-1 bound with DNA was determined by Kovall *et al* in 2004.[146] Core Lag-1 comprises three subdomains among which the central beta-trefoil domain (BTD) was found to bind with NICD by native gel electrophoresis.[146] Recent studies resolved crystal structures of CSL complexed with NICD and co-activator Mastermind (Figure A-1).[100,147] The new complex structure revealed that Notch binds CSL BTD via the same stretch of residues as we used in the preliminary binding test. Thus far, there is no CSL complex structure available to show how EBNA2 binds CSL. Although some researchers believed the binding site for EBNA2 might be similar to that of NICD,[146,147] some mutation data indicates that NICD and EBNA2 have distinct binding sites on CSL.[148,149] In this project, we studied the interaction between CSL and EBNA2 and compared it with CSL-NICD interaction.

A.2   Experimental procedures

A.2.1   CSL construct and peptide synthesis

The BTD of the human CSL179-333 was subcloned into a pET15 vector. The protein was overexpressed in E. coli strain BL21(DE3) and ^{15}N labeled by growing the cells in MOPS-buffered medium containing 15NH4Cl (1 g/L).[93] Proteins were purified by the Ni^{2+}-charged His-Bind Resin (Novagen, San Diego, CA) according to the manufacturer's protocol. The His tag was removed by thrombin cleavage (EMD Biosciences, San Diego, CA) at room temperature for 4 hours. The digested protein was dialyzed in 50 mM potassium phosphate (pH 7.2), 50 mM NaCl, 5 mM 1,4-dithiothreitol (DTT), and 5 mM EDTA and further purified by gel filtration on a HiLoad 26/60 Superdex 75 HR column (GE Healthcare, UK). The EBNA2 peptide and NICD peptide were chemically synthesized and purified by HPLC at the Hartwell Center of Bioinformatics and Biotechnology, St. Jude Children's Research Hospital, Memphis, TN. The CSL target DNA was synthesized as single strand DNA and annealed at 95 °C and cooled down to 25°C.
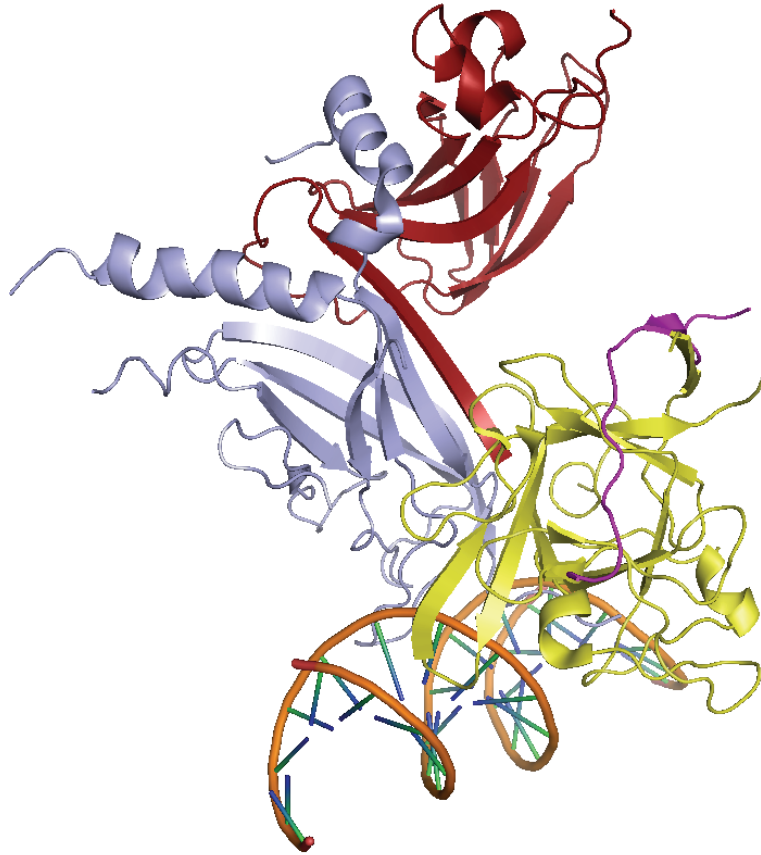
Figure A-1   Crystal structure of CSL bound with NICD.
The Crystal structure of CSL bound with NICD (2FO1, [147]) CSL is composed of 3 subdomains: N terminal domain(NTD) in blue, central beta-trefoil domain(BTD) in yellow and C terminal domain (CTD) in red. The NICD binds to BTD through a stretch of residues from the RAM domain. (Magenta).

A.2.2   NMR chemical shift perturbation titration studies

The ^{15}N labeled CSL BTD sample was approximately 400 μl at a concentration of 0.4 mM in 50 mM potassium phosphate buffer (pH 7.2), 50 mM NaCl, 5 mM EDTA, 5 mM DTT, and 5% (v/v) D$_2$O. All peptide stocks were prepared with the same buffer and the pH was readjusted to 7.2 before titration. All NMR spectra were acquired at 25°C with a Bruker Avance 600 MHz spectrometer equipped with cryoprobe. NMR data was processed by NMRpipe and analyzed with the software Sparky.[94]

A.3   Results and discussion

Based on the Lag-1 structure and sequence alignment from different organisms,[146] we designed several constructs for the central beta-trefoil domain(BTD) of human CSL homologue CBF1/RBP-Jκ. An optimal construct (hRBP-Jκ 179-333) and solution condition was obtained by extensive solubility screening and NMR analysis. An ^{15}N HSQC experiment was carried out for initial protein fold quality assessment. As shown in Figure A-2, the protein is very sensitive to pH changes and our screening found that pH 7.2 is the best pH to balance stability and deprotonation at low pH.

Using our CSL BTD construct and synthetic peptides from both hNotch1 and EBNA2, we were able to show that there is strong binding between CSL BTD and both peptides. The data indicated that the adding of both peptides can dramatically stabilize the CSL BTD structure. As shown in the ^{15}N HSQC spectrum, there are only a few dispersed peaks when observing ^{15}N-labeled CSL BTD alone (Figure A-4). A large number of peaks are clustering in the center, which is an indication of poor folding. However, upon adding the unlabeled peptides, almost all the peaks show up and nicely dispersed (Figure A-5). The experimental conditions were carefully controlled to insure the same solution environment, including pH, salt concentration, protein concentration and temperature. The spectra ware collected and processed with the same set of parameters too. The dramatic spectral change is a strong indication of protein- peptide interaction. Our explanation is that by binding to CSL BTD, both peptides dramatically stabilized the protein structure.

Interestingly, we noticed that there is significant peak shifting comparing hNotch1 peptide bound form and EBNA2 peptide bound form (Figure A-6). This abnormally large difference, as shown in the overlaid spectrum, led us to speculate that there might be different binding sites for the two peptides, which is consistent with the mutation data.[148,149] The other possibility is that the binding of different peptides to the same site might induce different conformational changes, which could open a new door to the study of Notch signaling and the EBNA2 hijacking mechanism.

At the same time, we also found that there was an interaction between CSL BTD and its target DNA sequence when it co-existed with, as we observed that the synthesized DNA fragment can further perturb the peaks from the EBNA2 peptide bound CSL BTD (Figure A-7). This indicated that the CSL binding site for EBNA2 peptide and DNA
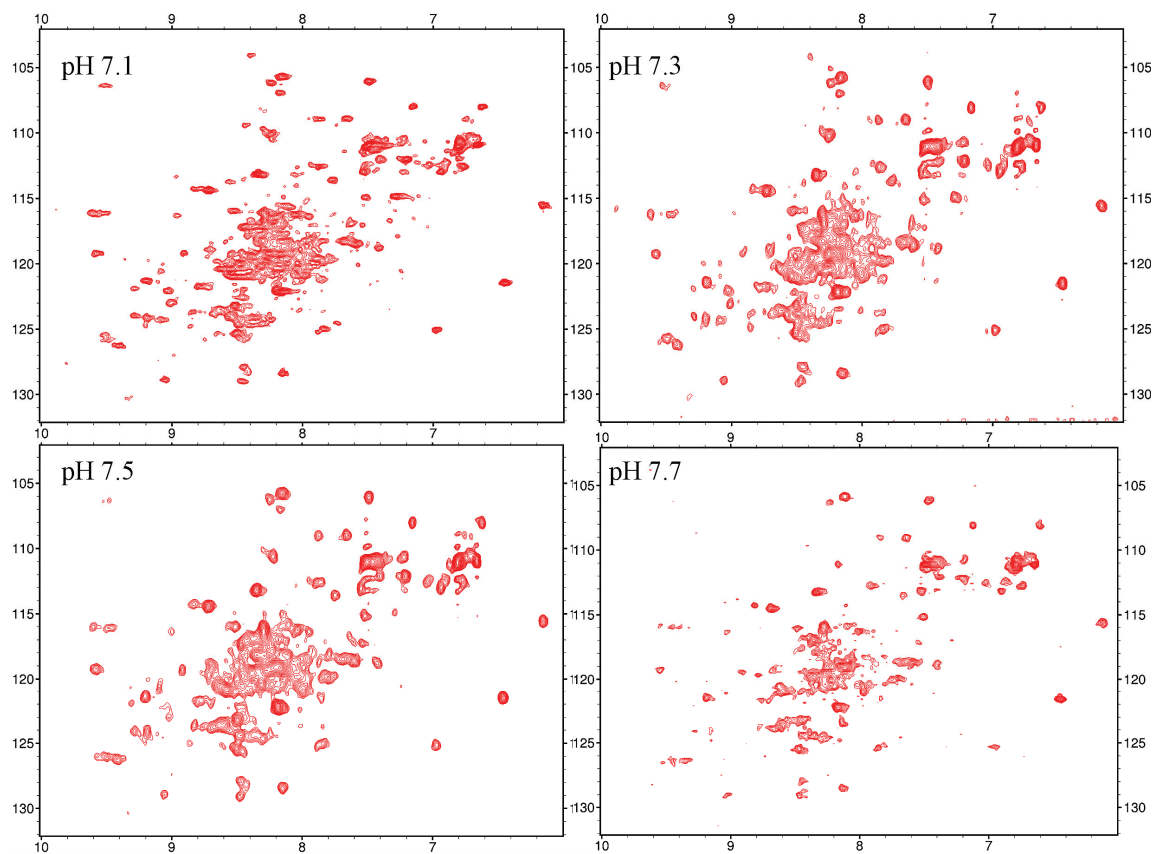
Figure A-2   CSL BTD fold is very sensitive to pH change.
The pH screening found that pH 7.2 is the best balance point for protein stability and deprotonation effect.

```
EBNA2      H L P S G P P W W P P I C
hNotch1    R R Q H G Q L W F P E G F
```

Figure A-3   The alignment of EBNA2 and Notch sequence used in the study.
The highlighted area is proposed as the CSL binding site.[146]
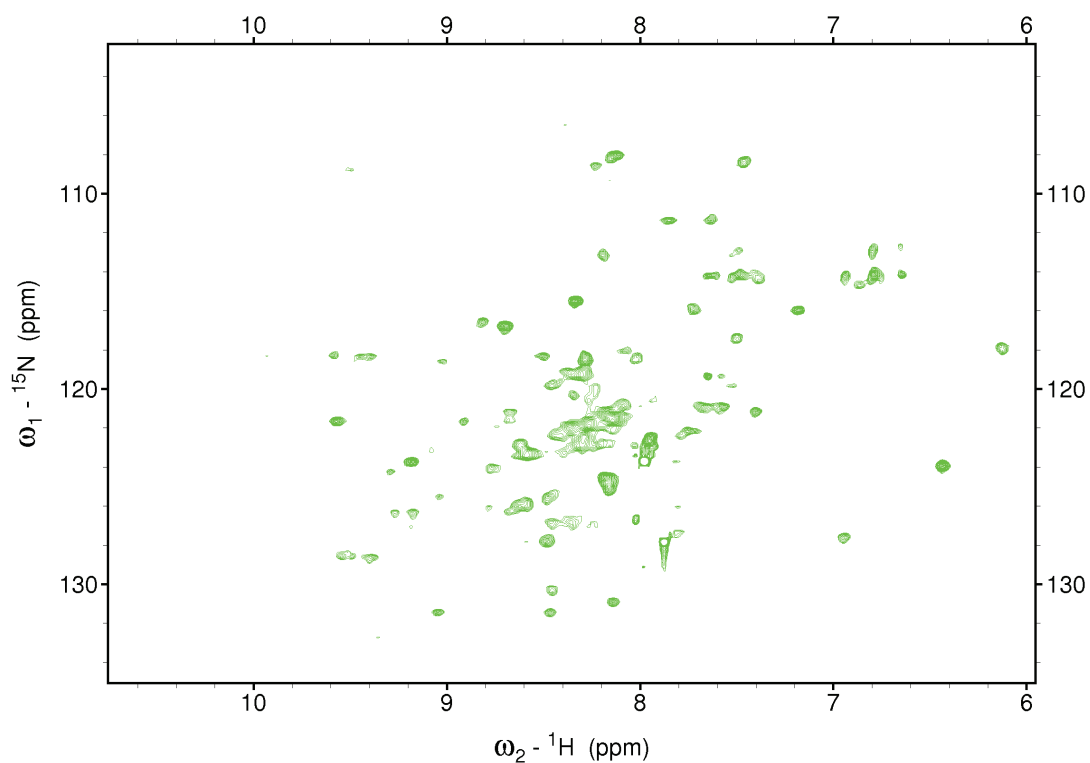
Figure A-4   15N HSQC spectrum of CSL BTD.
Peaks are poorly dispersed, which is an indication of poor folding.
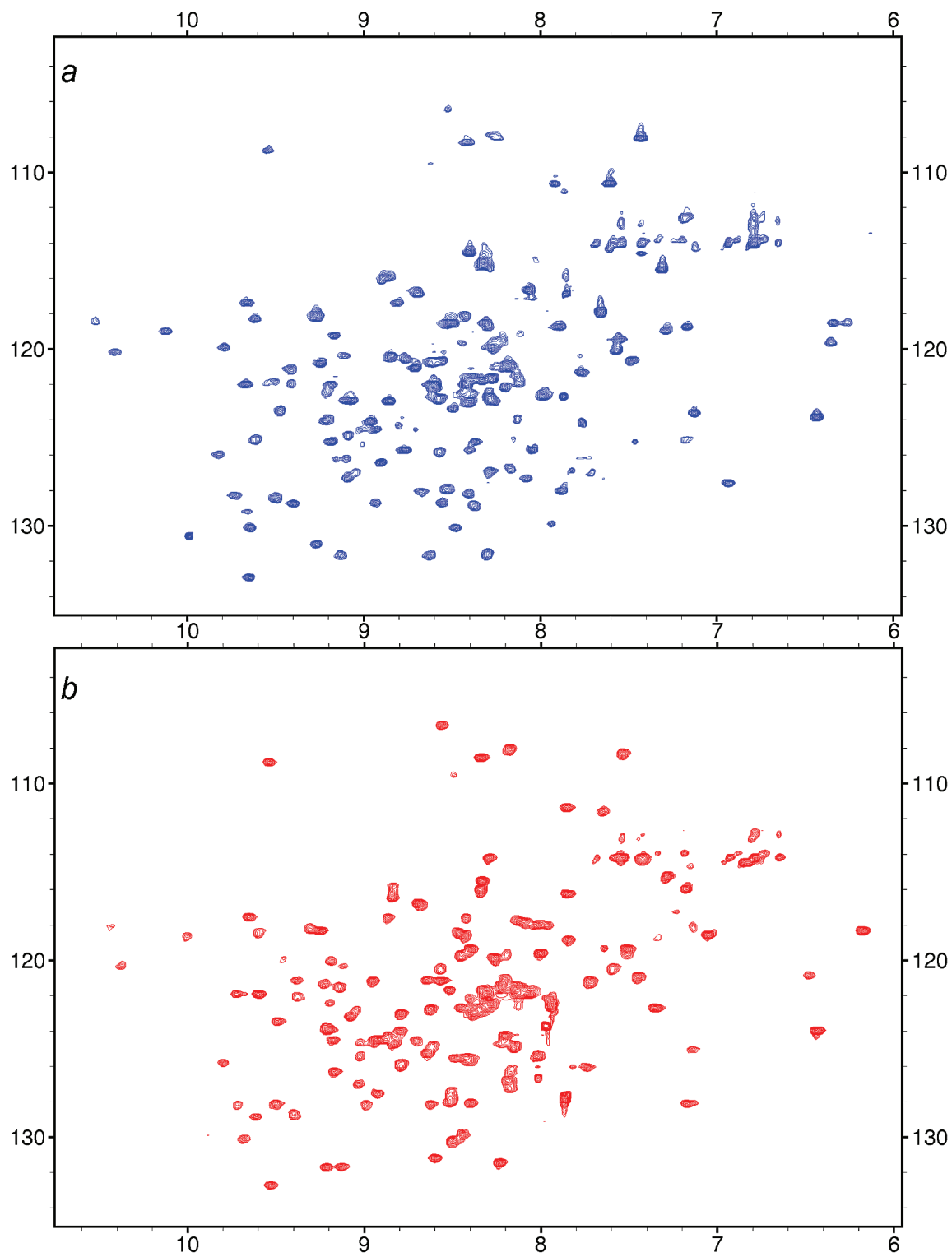
Figure A-5 ^{15}N HSQC spectra of CSL BTD bound with different peptides.
(a) CSL BTD + hNotch1 peptide (1:5); (b) CSL BTD + EBNA2 peptide (1:5).
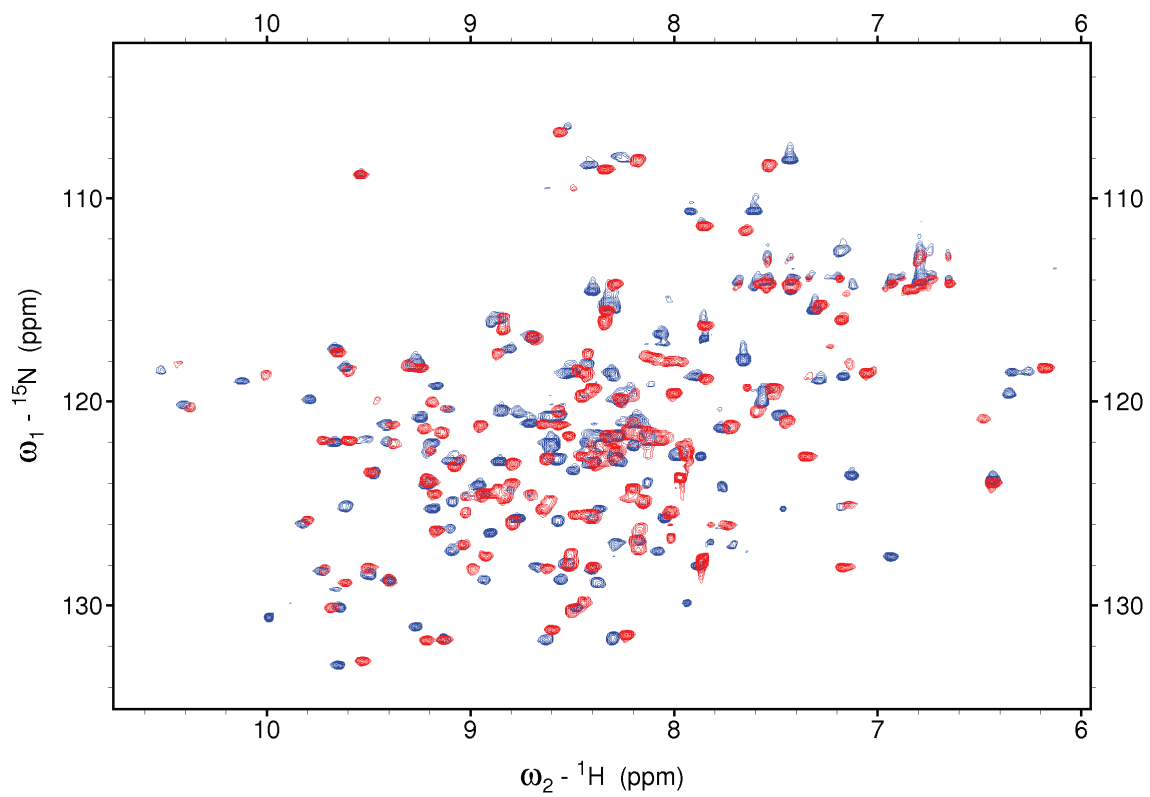
Figure A-6   Overlayed ¹⁵N HSQC spectra of CSLBTD bound with both peptides.
Blue -- CSL BTD + hNotch1 peptide (1:5); Red -- CSL BTD + EBNA2 peptide (1:5).
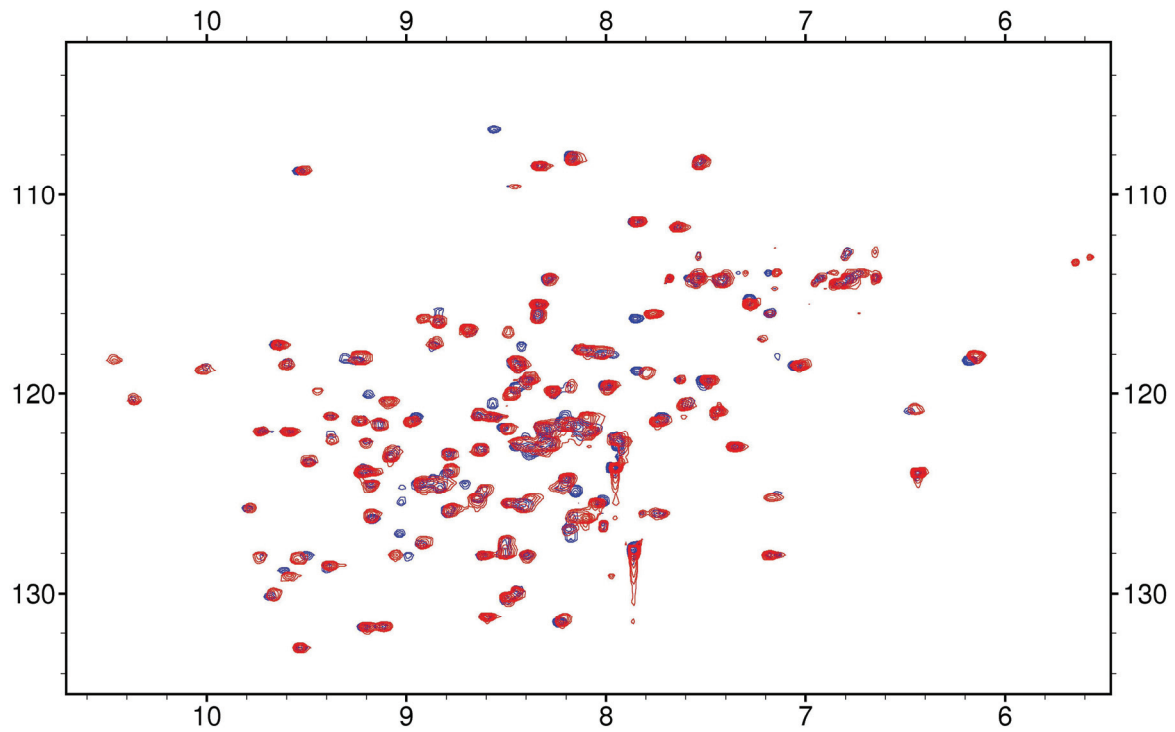
Figure A-7   The DNA target interacts with CSL BTD in the presence of EBNA2.
*Blue – CSL BTD + EBNA2 peptide (1:5) ; Red – CSL BTD + EBNA2 peptide (1:5) +
DNA target (1:5). It showed that the DNA can further perturb the CSL spectrum,
indicating the binding sites for EBNA2 and DNA are not overlapping.*

were not overlapping, and EBNA2 and DNA can bind CSL BTD simultaneously. It supported the hypothesis that EBNA2 functions as a transcription regulator because EBNA2 binding did not exclude DNA binding.

Studies had showed that EBV transformed B cells by hijacking Notch signaling with its latent protein EBNA2. However, there was no detailed model on how EBNA2 achieves its goal. Although how NICD interacts with CSL has been elucidated,[100,147] conflict still remained regarding whether EBNA2 shared the same binding site.[147-149] On basis of our preliminary data, we can conclude that EBNA2 might bind dramatically differently from NICD, either binding to a different site or inducing distinct conformational changes on CSL. However the question can not be answered without the CSL-EBNA2 complex structure determination. The complex structure can also help elucidate the mechanism by which EBNA2 hijacks Notch signaling. Hence, the future direction of this project is to solve the complex structure of CSL/EBNA2.

APPENDIX B.   NMR DATA PROCESSING SCRIPTS

NMRPipe script for processing 2D ^{15}N-HSQC spectrum:

```
FIN=ser
OUT=hsqc.ft5
bruk2pipe -in ./ser -bad 0.0 -aswap -DMX -decim 1920 -dspfvs 20 -grpdly
68  \
  -xN               2048  -yN               256  \
  -xT               1024  -yT               128  \
  -xMODE             DQD  -yMODE  Echo-AntiEcho  \
  -xSW          9615.385  -ySW         1519.988  \
  -xOBS          599.913  -yOBS          60.796  \
  -xCAR            4.618  -yCAR         118.913  \
  -xLAB               1H  -yLAB             15N  \
  -ndim                2  -aq2D          States  \
|nmrPipe -fn CBF \
|nmrPipe -fn SOL -mode 1 -fl 7 -fs 3 -poly              \
|nmrPipe -fn SP -size 1024 -off 0.5 -end 0.98 -c 0.5  -pow 2   \
|nmrPipe -fn GM -g1 16 -g2 10 \
|nmrPipe -fn ZF -size 2048                                     \
|nmrPipe -fn FT                                               \
|nmrPipe -fn EXT -left -sw                                    \
|nmrPipe -fn PS -p0 102.6 -p1 -8.4 -di                        \
|nmrPipe -fn TP \
|nmrPipe -fn CBF \
|nmrPipe -fn LP -ord  8 -fb -after -x1 1 -xn 128 -pred 384     \
|nmrPipe -fn SP -size 512 -off 0.5 -end 0.95 -c 0.5  -pow 2     \
|nmrPipe -fn GM -g1 16 -g2 18 \
|nmrPipe -fn ZF -size 512                                      \
|nmrPipe -fn FT                                              \
|nmrPipe -fn PS -p0 87 -p1 0 -di                             \
|nmrPipe -fn TP -verb                                        \
 -ov -out $OUT
```

NMRPipe script for processing 3D ^{15}N-NOESY HSQC spectrum:

```
FIN=ser
OUT=ph.ft

bruk2pipe -in $FIN    \
  -DMX -decim 2080 -bad 0.0 -dspfvs 20 -aswap  -grpdly 68       \
 -xN    1024             -yN     98              -zN     128        \
 -xT    512              -yT     49              -zT     64         \
 -xMODE DQD              -yMODE  Echo-Antiecho   -zMODE  States-TPPI  \
 -xSW   5699.088         -ySW    9956.353        -zSW    5699.088    \
 -xOBS  599.9128226      -yOBS   150.8532966     -zOBS   599.9128226  \
 -xCAR  4.705            -yCAR   38.500          -zCAR   4.705       \
 -xLAB  H1               -yLAB   C13             -zLAB   H1          \
 -ndim  3                -aq2D        States-TPPI                    \
|nmrPipe -fn SP -size 512 -off 0.44 -end 1 -c 0.5 -pow 2            \
|nmrPipe -fn ZF -size 1024                                         \
|nmrPipe -fn FT                                                   \
|nmrPipe -fn EXT -x1 10.0ppm -xn 0.0ppm -sw                \
|nmrPipe -fn PS -p0 14 -p1 0 -di                          \
```

```
|nmrPipe -fn TP                                                    \
|nmrPipe -fn SP -size 49 -off 0.44 -end 1 -c 0.5                   \
|nmrPipe -fn ZF -size 98                                           \
|nmrPipe -fn FT                                                    \
|nmrPipe -fn PS -p0 -90 -p1 0 -di                     \
|nmrPipe -fn TP                                                    \
|nmrPipe -fn ZTP -verb                                            \
|nmrPipe -fn SP -size 64 -off 0.44 -end 1 -c 0.5                  \
|nmrPipe -fn ZF -size 128                                         \
|nmrPipe -fn FT -alt                                      \
|nmrPipe -fn PS -p0 0 -p1 0 -di                       \
|nmrPipe -fn ZTP -verb                                        \
 -ov -out $OUT

proj3D -y $OUT y.ft
proj3D -z $OUT z.ft
```

NMRPipe script for lineal predication of the same 3D ^{15}N-NOESY HSQC spectrum:

```
FIN=ph.ft
OUT=lp.ft

nmrPipe -in $FIN -fn TP                                            \
|nmrPipe -fn HT                                                   \
|nmrPipe -fn FT    -inv                                           \
|nmrPipe -fn LP -ord  8 -fb -after -x1 1 -xn 49 -pred 49    \
|nmrPipe -fn SP -size 98 -off 0.44 -end 1.0 -c 0.5     \
|nmrPipe -fn ZF -size 196                                \
|nmrPipe -fn FT    -di                                   \
|nmrPipe -fn TP -verb                                    \
|nmrPipe -fn ZTP                                             \
|nmrPipe -fn HT                                          \
|nmrPipe -fn FT    -inv                                      \
|nmrPipe -fn LP -ord  8 -fb -after -x1 1 -xn 64 -pred 64    \
|nmrPipe -fn SP -size 128 -off 0.44 -end 1.0 -c 0.5    \
|nmrPipe -fn ZF -size 256                                \
|nmrPipe -fn FT    -di                                   \
|nmrPipe -fn ZTP -verb                                   \
 -out $OUT


proj3D -y $OUT y_lp.ft
proj3D -z $OUT z_lp.ft
```

APPENDIX C. GIT1 PBD ATOM ASSIGNMENTS

	PPM			ASSIGNMENT						
1	7.897	0.000	H	135		90	120.682	0.000	N	41
2	126.468	0.000	N	135		91	8.516	0.000	H	53
3	7.883	0.000	H	83		92	120.580	0.000	N	53
4	125.278	0.000	N	83		93	8.175	0.000	H	6
7	8.291	0.000	H	103		94	120.567	0.000	N	6
8	125.145	0.000	N	103		95	8.630	0.000	H	23
11	8.672	0.000	H	20		96	120.513	0.000	N	23
12	124.211	0.000	N	20		97	7.726	0.000	H	78
13	8.007	0.000	H	50		98	120.454	0.000	N	78
14	124.059	0.000	N	50		99	8.561	0.000	H	89
15	8.317	0.000	H	104		100	120.428	0.000	N	89
16	123.920	0.000	N	104		101	8.470	0.000	H	52
17	8.346	0.000	H	24		102	120.430	0.000	N	52
18	123.856	0.000	N	24		105	8.111	0.000	H	124
23	8.581	0.000	H	30		106	120.377	0.000	N	124
24	123.388	0.000	N	30		109	8.013	0.000	H	35
31	8.033	0.000	H	121		110	120.035	0.000	N	35
32	122.993	0.000	N	121		111	7.616	0.000	H	51
33	8.587	0.000	H	105		112	120.073	0.000	N	51
34	122.967	0.000	N	105		113	8.927	0.000	H	60
35	8.306	0.000	H	5		114	120.063	0.000	N	60
36	122.988	0.000	N	5		115	8.727	0.000	H	37
39	8.405	0.000	H	28		116	119.920	0.000	N	37
40	122.794	0.000	N	28		117	8.609	0.000	H	16
41	8.267	0.000	H	76		118	119.823	0.000	N	16
42	122.591	0.000	N	76		119	7.661	0.000	H	26
43	7.936	0.000	H	45		120	119.721	0.000	N	26
44	122.669	0.000	N	45		121	8.131	0.000	H	84
47	8.418	0.000	H	10		122	119.692	0.000	N	84
48	122.292	0.000	N	10		125	8.541	0.000	H	122
53	8.728	0.000	H	81		126	119.565	0.000	N	122
54	122.083	0.000	N	81		127	6.977	0.000	H	42
55	8.011	0.000	H	13		128	119.540	0.000	N	42
56	122.081	0.000	N	13		129	8.338	0.000	H	65
57	7.429	0.000	H	33		130	119.466	0.000	N	65
58	122.009	0.000	N	33		131	8.137	0.000	H	86
61	8.147	0.000	H	8		132	119.413	0.000	N	86
62	121.678	0.000	N	8		133	8.474	0.000	H	31
63	7.924	0.000	H	102		134	119.378	0.000	N	31
64	121.689	0.000	N	102		139	8.787	0.000	H	17
67	8.372	0.000	H	85		140	119.201	0.000	N	17
68	121.571	0.000	N	85		141	7.739	0.000	H	73
69	7.236	0.000	H	94		142	119.184	0.000	N	73
70	121.359	0.000	N	94		143	8.356	0.000	H	127
75	8.489	0.000	H	111		144	119.183	0.000	N	127
76	121.215	0.000	N	111		145	7.684	0.000	H	75
77	9.016	0.000	H	56		146	119.215	0.000	N	75
78	121.204	0.000	N	56		147	7.886	0.000	H	107
81	8.052	0.000	H	54		148	118.991	0.000	N	107
82	121.063	0.000	N	54		153	8.046	0.000	H	112
87	8.015	0.000	H	118		154	118.971	0.000	N	112
88	120.681	0.000	N	118		155	7.544	0.000	H	109
89	8.437	0.000	H	41		156	118.956	0.000	N	109

159	8.300	0.000	H	120
160	118.879	0.000	N	120
161	8.561	0.000	H	34
162	118.829	0.000	N	34
163	7.426	0.000	H	92
164	118.651	0.000	N	92
165	7.643	0.000	H	91
166	118.674	0.000	N	91
169	8.046	0.000	H	132
170	118.473	0.000	N	132
175	7.902	0.000	H	110
176	118.092	0.000	N	110
179	8.463	0.000	H	69
180	117.889	0.000	N	69
181	8.307	0.000	H	66
182	117.625	0.000	N	66
185	7.485	0.000	H	29
186	117.483	0.000	N	29
187	7.674	0.000	H	87
188	117.356	0.000	N	87
189	8.377	0.000	H	57
190	117.109	0.000	N	57
191	8.769	0.000	H	49
192	117.127	0.000	N	49
193	7.589	0.000	H	40
194	116.950	0.000	N	40
195	7.241	0.000	H	25
196	116.846	0.000	N	25
197	7.961	0.000	H	123
198	116.721	0.000	N	123
199	8.055	0.000	H	22
200	116.329	0.000	N	22
201	6.722	0.000	H	72
202	116.163	0.000	N	72
203	7.224	0.000	H	97
204	115.883	0.000	N	97
205	7.683	0.000	H	63
206	115.860	0.000	N	63
209	8.104	0.000	H	80
210	115.198	0.000	N	80
211	8.398	0.000	H	82
212	115.020	0.000	N	82
215	7.408	0.000	H	39
216	115.011	0.000	N	39
217	8.215	0.000	H	108
218	114.766	0.000	N	108
219	8.000	0.000	H	48
220	114.588	0.000	N	48
221	9.056	0.000	H	74
222	114.360	0.000	N	74
223	7.942	0.000	H	130
224	114.070	0.000	N	130
225	7.857	0.000	H	88
226	114.030	0.000	N	88
227	8.278	0.000	H	90
228	113.793	0.000	N	90
229	8.678	0.000	H	15
230	113.474	0.000	N	15
231	7.827	0.000	H	44
232	113.311	0.000	N	44
235	7.913	0.000	H	61
236	112.714	0.000	N	61
237	7.827	0.000	H	129
238	112.785	0.000	N	129
239	8.906	0.000	H	12
240	111.652	0.000	N	12
241	8.038	0.000	H	7
242	108.338	0.000	N	7
243	7.397	0.000	H	93
244	107.623	0.000	N	93
245	9.256	0.000	H	46
246	118.748	0.000	N	46
247	7.583	0.000	H	113
248	117.871	0.000	N	113
251	8.378	0.000	H	128
252	120.424	0.000	N	128
253	8.839	0.000	H	27
254	114.868	0.000	N	27
255	8.668	0.000	H	99
256	111.031	0.000	N	99
257	9.363	0.000	H	70
258	128.644	0.000	N	70
269	8.543	0.000	H	43
270	118.477	0.000	N	43
272	57.437	0.000	CA	135
273	55.507	0.000	CA	83
276	52.440	0.000	CA	103
277	59.540	0.000	CA	50
279	7.715	0.000	H	100
280	125.137	0.000	N	100
282	7.717	0.000	H	68
283	124.945	0.000	N	68
285	50.341	0.000	CA	100
286	54.266	0.000	CA	68
288	65.977	0.000	CA	20
290	61.473	0.000	CA	104
291	59.607	0.000	CA	24
295	59.913	0.000	CA	105
296	65.296	0.000	CA	30
297	8.108	0.000	H	134
298	123.302	0.000	N	134
301	8.362	0.000	H	79
302	123.042	0.000	N	79
304	58.930	0.000	CA	79
306	59.919	0.000	CA	28
308	55.232	0.000	CA	5
310	58.161	0.000	CA	76
312	50.924	0.000	CA	10
314	55.128	0.000	CA	121
315	59.574	0.000	CA	45
317	7.756	0.000	H	106
318	122.150	0.000	N	106
319	57.679	0.000	CA	106
322	55.404	0.000	CA	81

323	51.544	0.000	CA	13
326	57.954	0.000	CA	33
327	8.202	0.000	H	115
328	121.882	0.000	N	115
330	55.576	0.000	CA	115
332	52.888	0.000	CA	8
333	8.151	0.000	H	58
334	121.817	0.000	N	58
335	59.746	0.000	CA	58
338	61.676	0.000	CA	102
339	7.768	0.000	H	36
340	121.547	0.000	N	36
342	55.369	0.000	CA	36
346	7.769	0.000	H	131
347	121.795	0.000	N	131
349	57.782	0.000	CA	131
352	7.812	0.000	H	55
353	121.350	0.000	N	55
358	7.831	0.000	H	21
359	121.049	0.000	N	21
361	58.506	0.000	CA	21
362	55.025	0.000	CA	55
363	7.885	0.000	H	119
364	120.684	0.000	N	119
366	55.404	0.000	CA	119
367	7.895	0.000	H	133
368	120.783	0.000	N	133
370	56.059	0.000	CA	133
372	59.402	0.000	CA	94
374	59.471	0.000	CA	85
376	67.363	0.000	CA	111
378	67.398	0.000	CA	56
380	58.368	0.000	CA	54
382	65.881	0.000	CA	118
384	64.571	0.000	CA	128
386	58.644	0.000	CA	89
388	61.091	0.000	CA	53
390	66.433	0.000	CA	52
392	53.922	0.000	CA	41
394	54.473	0.000	CA	6
396	68.121	0.000	CA	23
398	57.920	0.000	CA	78
400	59.195	0.000	CA	124
402	59.371	0.000	CA	35
404	59.960	0.000	CA	51
406	55.304	0.000	CA	60
407	55.375	0.000	CA	37
410	66.443	0.000	CA	16
412	67.768	0.000	CA	26
414	61.643	0.000	CA	84
416	58.849	0.000	CA	86
426	58.133	0.000	CA	107
428	55.447	0.000	CA	122
430	59.422	0.000	CA	31
432	57.990	0.000	CA	42
435	58.635	0.000	CA	65
436	67.159	0.000	CA	127
438	59.602	0.000	CA	120
440	63.541	0.000	CA	75
442	60.676	0.000	CA	73
444	58.348	0.000	CA	91
447	7.479	0.000	H	32
448	118.878	0.000	N	32
451	7.486	0.000	H	62
452	119.030	0.000	N	62
456	59.315	0.000	CA	109
458	58.061	0.000	CA	92
459	65.690	0.000	CA	112
461	8.039	0.000	H	38
462	118.962	0.000	N	38
463	58.312	0.000	CA	38
465	57.990	0.000	CA	34
468	56.772	0.000	CA	132
470	58.506	0.000	CA	110
472	54.129	0.000	CA	69
474	52.785	0.000	CA	66
475	8.289	0.000	H	116
476	117.246	0.000	N	116
478	60.539	0.000	CA	116
480	66.915	0.000	CA	57
482	56.576	0.000	CA	29
484	59.850	0.000	CA	87
486	62.779	0.000	CA	49
488	58.713	0.000	CA	40
490	58.506	0.000	CA	25
492	59.367	0.000	CA	123
493	8.046	0.000	H	114
494	116.145	0.000	N	114
495	63.606	0.000	CA	114
498	57.541	0.000	CA	22
499	65.950	0.000	CA	72
503	53.336	0.000	CA	97
505	56.128	0.000	CA	63
507	56.645	0.000	CA	80
509	62.607	0.000	CA	82
511	55.542	0.000	CA	39
513	68.397	0.000	CA	108
515	64.330	0.000	CA	48
517	61.642	0.000	CA	74
519	64.916	0.000	CA	130
521	61.608	0.000	CA	88
523	64.571	0.000	CA	90
525	57.748	0.000	CA	15
527	59.540	0.000	CA	44
529	66.777	0.000	CA	129
531	46.167	0.000	CA	12
533	44.892	0.000	CA	7
535	63.434	0.000	CA	93
537	69.190	0.000	CA	46
539	59.126	0.000	CA	113
548	60.784	0.000	CA	61
550	67.076	0.000	CA	27
554	61.255	0.000	CA	70
565	55.948	0.000	CA	43

568	30.457	0.000	CB	135	671	17.466	0.000	CB	122
570	39.993	0.000	CB	103	673	29.817	0.000	CB	42
571	18.203	0.000	CB	83	675	32.569	0.000	CB	65
573	37.703	0.000	CB	20	677	68.600	0.000	CB	127
574	29.255	0.000	CB	50	679	32.505	0.000	CB	120
576	29.065	0.000	CB	24	681	41.401	0.000	CB	86
577	38.910	0.000	CB	104	683	27.705	0.000	CB	31
579	37.004	0.000	CB	30	685	28.921	0.000	CB	17
581	18.203	0.000	CB	121	687	30.777	0.000	CB	73
583	27.985	0.000	CB	105	690	42.297	0.000	CB	107
585	42.149	0.000	CB	5	693	30.137	0.000	CB	132
587	32.177	0.000	CB	28	695	37.625	0.000	CB	112
589	36.115	0.000	CB	45	696	28.473	0.000	CB	38
590	43.293	0.000	CB	10	697	28.217	0.000	CB	109
593	18.394	0.000	CB	81	699	40.825	0.000	CB	34
594	44.245	0.000	CB	13	701	32.249	0.000	CB	92
596	41.768	0.000	CB	33	703	29.817	0.000	CB	91
597	40.498	0.000	CB	8	705	28.537	0.000	CB	110
599	29.763	0.000	CB	58	707	43.129	0.000	CB	69
601	17.759	0.000	CB	115	709	30.201	0.000	CB	66
603	34.083	0.000	CB	102	711	39.289	0.000	CB	29
605	30.017	0.000	CB	85	713	29.433	0.000	CB	87
607	32.304	0.000	CB	94	715	68.472	0.000	CB	57
609	31.478	0.000	CB	111	719	36.089	0.000	CB	40
611	30.907	0.000	CB	56	721	28.473	0.000	CB	25
613	41.514	0.000	CB	54	723	31.801	0.000	CB	123
615	38.148	0.000	CB	118	725	30.841	0.000	CB	22
617	30.081	0.000	CB	41	727	27.897	0.000	CB	114
619	30.145	0.000	CB	53	730	30.521	0.000	CB	97
622	29.128	0.000	CB	89	732	40.313	0.000	CB	63
624	38.211	0.000	CB	52	734	38.969	0.000	CB	80
626	38.656	0.000	CB	128	737	30.073	0.000	CB	39
628	41.196	0.000	CB	6	740	28.665	0.000	CB	48
630	41.514	0.000	CB	78	742	62.520	0.000	CB	74
632	27.858	0.000	CB	124	745	62.648	0.000	CB	88
634	29.509	0.000	CB	35	747	26.938	0.000	CB	90
636	32.050	0.000	CB	51	749	64.056	0.000	CB	15
638	18.013	0.000	CB	60	753	64.056	0.000	CB	44
640	17.695	0.000	CB	37	755	63.160	0.000	CB	61
642	68.382	0.000	CB	16	757	68.856	0.000	CB	129
644	31.859	0.000	CB	26	762	69.462	0.000	CB	93
646	37.703	0.000	CB	84	764	28.874	0.000	CB	46
647	7.884	0.000	H	59	765	28.112	0.000	CB	113
648	119.783	0.000	N	59	768	68.216	0.000	CB	27
649	7.825	0.000	H	125	778	39.417	0.000	CB	43
650	119.193	0.000	N	125	780	19.386	0.000	CB	68
651	7.881	0.000	H	77	781	18.554	0.000	CB	100
652	119.581	0.000	N	77	783	56.378	0.000	CA	134
653	7.841	0.000	H	19	785	32.889	0.000	CB	134
654	119.337	0.000	N	19	787	42.489	0.000	CB	79
660	58.312	0.000	CA	125	789	41.849	0.000	CB	106
662	41.145	0.000	CB	125	794	41.785	0.000	CB	21
663	66.552	0.000	CA	19	795	18.042	0.000	CB	55
665	31.673	0.000	CB	19	796	17.338	0.000	CB	36
667	59.745	0.000	CA	77	797	30.521	0.000	CB	131
668	29.689	0.000	CB	77	800	32.569	0.000	CB	133
669	32.505	0.000	CB	59	801	18.042	0.000	CB	119

806	55.805	0.000	CA	62
808	29.369	0.000	CB	32
810	42.553	0.000	CB	62
812	36.985	0.000	CB	116
813	8.138	0.000	H	117
814	122.451	0.000	N	117
816	57.990	0.000	CA	117
818	39.929	0.000	CB	117
824	69.208	0.000	CB	130
825	22.283	0.000	CB	108
826	59.746	0.000	CA	59
827	59.693	0.000	CA	32
828	60.379	0.000	CA	17
833	3.975	0.000	HA	49
834	3.667	0.000	HA	27
835	4.090	0.000	HA	74
836	4.281	0.000	HA	15
837	4.435	0.000	HA	94
838	4.229	0.000	HA	33
839	3.382	0.000	HA	72
840	4.073	0.000	HA	25
841	2.473	0.000	HA	42
842	4.134	0.000	HA	135
844	4.582	0.000	HA	100
846	4.175	0.000	HA	50
847	4.665	0.000	HA	103
848	3.592	0.000	HA	20
849	4.061	0.000	HA	24
850	4.134	0.000	HA	104
851	4.300	0.000	HA	134
853	4.040	0.000	HA	28
854	4.311	0.000	HA	5
855	3.831	0.000	HA	76
856	4.029	0.000	HA	121
857	4.290	0.000	HA	45
858	4.811	0.000	HA	13
859	4.519	0.000	HA	117
860	3.790	0.000	HA	102
861	3.977	0.000	HA	58
862	4.842	0.000	HA	8
863	3.988	0.000	HA	85
864	3.287	0.000	HA	111
865	3.824	0.000	HA	115
866	3.985	0.000	HA	54
867	3.578	0.000	HA	118
868	4.038	0.000	HA	124
869	4.517	0.000	HA	6
870	4.001	0.000	HA	35
871	3.790	0.000	HA	128
872	4.401	0.000	HA	41
873	3.485	0.000	HA	52
874	3.972	0.000	HA	53
875	4.183	0.000	HA	89
876	3.826	0.000	HA	23
877	4.103	0.000	HA	37
878	3.679	0.000	HA	16
879	3.943	0.000	HA	122

880	3.877	0.000	HA	31
881	3.957	0.000	HA	65
882	3.805	0.000	HA	127
883	3.892	0.000	HA	120
884	4.343	0.000	HA	43
885	3.783	0.000	HA	34
886	4.255	0.000	HA	84
888	3.528	0.000	HA	19
889	4.052	0.000	HA	125
890	3.950	0.000	HA	59
891	3.965	0.000	HA	77
892	4.665	0.000	HA	97
893	4.207	0.000	HA	39
895	4.540	0.000	HA	29
896	3.789	0.000	HA	40
897	3.862	0.000	HA	87
898	3.942	0.000	HA	113
899	3.978	0.000	HA	91
900	4.055	0.000	HA	110
901	4.191	0.000	HA	132
902	4.191	0.000	HA	107
903	3.236	0.000	HA	112
904	3.876	0.000	HA	38
905	4.280	0.000	HA	82
906	4.138	0.000	HA	90
907	3.745	0.000	HA	108
908	4.238	0.000	HA	48
909	4.280	0.000	HA	80
910	4.065	0.000	HA	130
912	4.590	0.000	HA	44
913	3.923	0.000	HA	129
914	4.380	0.000	HA	61
915	4.438	0.000	HA	63
917	4.196	0.000	HA	22
918	3.949	0.000	HA	123
919	3.922	0.000	HA	57
920	4.488	0.000	HA	69
921	4.238	0.000	HA	81
922	3.390	0.000	HA	56
923	4.144	0.000	HA	83
924	3.760	0.000	HA	46
925	3.982	0.000	HA	60
926	3.834	0.000	HA	17
927	4.236	0.000	HA	105
928	3.622	0.000	HA	30
929	5.008	0.000	HA	10
930	4.172	0.000	HA	78
931	4.288	0.000	HA	133
932	4.119	0.000	HA	55
933	4.119	0.000	HA	21
934	4.151	0.000	HA	131
935	4.098	0.000	HA	106
936	4.109	0.000	HA	36
937	4.056	0.000	HA	51
938	3.389	0.000	HA	26
939	4.214	0.000	HA	75
940	3.749	0.000	HA	73

941	3.961	0.000	HA	109	1425	1.359	0.000	HB	102
942	4.013	0.000	HA	62	1426	3.761	0.000	HA3	99
944	4.130	0.000	HA	92	1427	3.999	0.000	HA2	99
946	3.887	0.000	HA	7	1430	4.365	0.000	HB	93
949	3.738	0.000	HA3	12	1431	1.761	0.000	HB	91
951	4.574	0.000	HA	66	1432	2.000	0.000	HB	92
952	4.320	0.000	HA	116	1434	3.120	0.000	HB2	90
953	3.950	0.000	HA	93	1435	2.473	0.000	HB3	90
976	1.760	0.000	HB	128	1437	1.872	0.000	HB3	89
977	1.887	0.000	HB	131	1438	2.005	0.000	HB2	89
1374	29.671	0.000	CD	120	1439	3.976	0.000	HA	88
1375	1.507	0.000	HB	83	1440	3.814	0.000	HB	88
1376	2.877	0.000	HB2	103	1441	1.964	0.000	HB	87
1377	2.423	0.000	HB3	103	1443	3.124	0.000	HB3	84
1378	1.874	0.000	HB	20	1444	3.939	0.000	HB	82
1379	1.689	0.000	HG12	20	1445	1.518	0.000	HB	81
1380	1.084	0.000	HG13	20	1446	2.735	0.000	HB2	80
1382	0.746	0.000	HD1	20	1447	2.630	0.000	HB3	80
1383	0.860	0.000	HG2	20	1453	1.526	0.000	HB3	79
1384	2.083	0.000	HB3	50	1454	3.721	0.000	HA	79
1385	1.891	0.000	HB3	135	1455	1.995	0.000	HB2	79
1386	2.065	0.000	HB2	135	1456	1.799	0.000	HB	78
1387	4.305	0.000	HB	127	1457	1.918	0.000	HB	77
1388	4.248	0.000	HB	129	1460	1.381	0.000	HB3	76
1389	4.279	0.000	HB	130	1461	1.627	0.000	HB2	76
1390	1.836	0.000	HB3	132	1462	3.764	0.000	HB3	75
1391	2.081	0.000	HB2	132	1463	3.837	0.000	HB2	75
1392	1.848	0.000	HB3	133	1464	3.867	0.000	HB	74
1393	1.766	0.000	HB2	133	1465	1.876	0.000	HB	73
1394	1.766	0.000	HB3	134	1466	2.215	0.000	HB	72
1395	1.848	0.000	HB2	134	1467	1.253	0.000	HB3	69
1397	2.055	0.000	HB3	124	1468	1.903	0.000	HB2	69
1398	2.162	0.000	HB2	124	1469	3.952	0.000	HA	68
1399	1.861	0.000	HB	123	1470	1.281	0.000	HB	68
1400	1.376	0.000	HB	122	1473	1.804	0.000	HB2	62
1401	1.483	0.000	HB	121	1474	0.977	0.000	HB3	62
1402	1.873	0.000	HB	120	1476	4.005	0.000	HB	61
1403	4.310	0.000	HA	119	1477	1.489	0.000	HB	60
1404	1.536	0.000	HB	119	1480	1.904	0.000	HB	59
1405	1.876	0.000	HB	118	1484	1.843	0.000	HB3	58
1406	2.674	0.000	HB3	117	1485	2.074	0.000	HB2	58
1407	2.976	0.000	HB2	117	1487	2.162	0.000	HB	56
1408	3.117	0.000	HB	116	1488	1.443	0.000	HB	55
1409	1.325	0.000	HB	115	1489	1.757	0.000	HB	54
1412	4.198	0.000	HA	114	1492	3.243	0.000	HB2	53
1413	3.077	0.000	HB2	114	1493	3.155	0.000	HB3	53
1414	2.712	0.000	HB3	114	1494	1.823	0.000	HB	52
1415	2.152	0.000	HB	113	1497	1.820	0.000	HB3	51
1416	1.700	0.000	HB	112	1498	2.035	0.000	HB2	51
1417	2.014	0.000	HB	111	1499	2.194	0.000	HB2	50
1418	1.987	0.000	HB3	110	1502	3.588	0.000	HB2	45
1419	2.235	0.000	HB2	110	1503	3.296	0.000	HB3	45
1420	1.316	0.000	HB	108	1506	4.087	0.000	HB2	44
1421	1.651	0.000	HB	107	1507	3.972	0.000	HB3	44
1422	1.563	0.000	HB	106	1508	2.575	0.000	HB	43
1423	2.247	0.000	HB	105	1509	2.356	0.000	HB3	42
1424	2.805	0.000	HB	104	1512	1.523	0.000	HB3	41

1513	1.829	0.000	HB2	41	1606	1.693	0.000	HG2	131
1516	3.239	0.000	HB2	40	1607	27.185	0.000	CG	131
1517	3.150	0.000	HB3	40	1608	3.155	0.000	HD	131
1520	1.787	0.000	HB3	39	1609	43.478	0.000	CD	131
1521	2.196	0.000	HB2	39	1610	1.227	0.000	HG2	130
1522	2.128	0.000	HB3	38	1611	21.382	0.000	CG2	130
1523	2.205	0.000	HB2	38	1612	1.288	0.000	HG2	129
1526	1.324	0.000	HB	37	1613	22.431	0.000	CG2	129
1527	1.637	0.000	HB	36	1614	0.864	0.000	HG2	128
1528	1.850	0.000	HB	35	1615	18.025	0.000	CG2	128
1531	1.367	0.000	HB3	34	1616	1.109	0.000	HG13	128
1532	1.791	0.000	HB2	34	1617	1.672	0.000	HG12	128
1535	1.996	0.000	HB2	33	1618	29.099	0.000	CG1	128
1536	1.333	0.000	HB3	33	1619	0.700	0.000	HD1	128
1539	2.229	0.000	HB	31	1620	14.109	0.000	CD1	128
1541	2.095	0.000	HB	30	1622	1.185	0.000	HG2	127
1542	2.613	0.000	HB	29	1623	21.522	0.000	CG2	127
1543	1.844	0.000	HB	28	1624	2.379	0.000	HG3	124
1544	3.674	0.000	HB	27	1625	2.486	0.000	HG2	124
1545	2.414	0.000	HB	26	1626	33.977	0.000	CG	124
1546	2.177	0.000	HB3	25	1627	1.370	0.000	HG	123
1549	2.169	0.000	HB2	24	1628	24.948	0.000	CG	123
1550	2.065	0.000	HB3	24	1629	1.610	0.000	HD	123
1553	1.658	0.000	HB3	21	1630	29.213	0.000	CD	123
1554	1.868	0.000	HB2	21	1631	2.877	0.000	HE	123
1556	2.122	0.000	HB	19	1632	42.080	0.000	CE	123
1559	3.983	0.000	HB	16	1633	1.445	0.000	HG	120
1560	3.973	0.000	HB	15	1634	24.715	0.000	CG	120
1562	2.577	0.000	HB	6	1635	0.667	0.000	HD1	118
1563	1.563	0.000	HB	5	1636	17.676	0.000	CD1	118
1570	4.221	0.000	HB	23	1639	0.862	0.000	HG2	118
1571	1.809	0.000	HB3	22	1640	1.088	0.000	HG13	118
1572	2.000	0.000	HB2	22	1642	1.681	0.000	HG12	118
1573	3.761	0.000	HB	49	1643	29.213	0.000	CG1	118
1576	3.039	0.000	HB2	48	1644	18.095	0.000	CG2	118
1577	2.727	0.000	HB3	48	1645	2.407	0.000	HG	113
1578	62.831	0.000	CB	49	1646	33.968	0.000	CG	113
1579	1.759	0.000	HB	65	1648	0.704	0.000	HD1	112
1585	45.269	0.000	CA	99	1649	0.655	0.000	HG2	112
1587	2.261	0.000	HG	135	1650	1.004	0.000	HG13	112
1588	34.300	0.000	CG	135	1651	17.466	0.000	CG2	112
1589	1.428	0.000	HG	134	1652	1.560	0.000	HG12	112
1590	24.676	0.000	CG	134	1653	27.046	0.000	CG1	112
1591	1.641	0.000	HD	134	1654	14.040	0.000	CD1	112
1592	28.937	0.000	CD	134	1655	0.815	0.000	HG2	111
1593	2.964	0.000	HE	134	1656	0.354	0.000	HG1	111
1594	42.260	0.000	CE	134	1657	20.892	0.000	CG1	111
1595	1.447	0.000	HG	133	1658	23.479	0.000	CG2	111
1596	24.667	0.000	CG	133	1659	2.383	0.000	HG	110
1597	1.642	0.000	HD	133	1660	34.038	0.000	CG	110
1599	29.073	0.000	CD	133	1662	2.219	0.000	HB	109
1600	2.915	0.000	HE	133	1663	2.456	0.000	HG	109
1601	42.150	0.000	CE	133	1664	33.968	0.000	CG	109
1602	2.155	0.000	HG3	132	1665	0.858	0.000	HD2	106
1603	2.320	0.000	HG2	132	1666	0.804	0.000	HD1	106
1604	36.114	0.000	CG	132	1667	1.531	0.000	HG	106
1605	1.652	0.000	HG3	131	1668	27.046	0.000	CG	106

1669	23.410	0.000	CD1	106
1670	24.668	0.000	CD2	106
1671	2.370	0.000	HG3	105
1672	2.451	0.000	HG2	105
1673	34.108	0.000	CG	105
1674	0.162	0.000	HG1	102
1675	0.631	0.000	HG2	102
1676	20.752	0.000	CG1	102
1677	20.822	0.000	CG2	102
1679	2.064	0.000	HB	94
1680	0.892	0.000	HG1	94
1681	0.898	0.000	HG2	94
1682	20.892	0.000	CG1	94
1683	20.719	0.000	CG2	94
1684	1.364	0.000	HG2	93
1685	22.640	0.000	CG2	93
1686	1.688	0.000	HG	92
1687	25.787	0.000	CG	92
1688	1.582	0.000	HD	92
1689	29.213	0.000	CD	92
1690	2.914	0.000	HE	92
1691	42.150	0.000	CE	92
1692	1.676	0.000	HG	91
1693	27.116	0.000	CG	91
1694	3.152	0.000	HD	91
1695	43.478	0.000	CD	91
1696	2.326	0.000	HG	89
1697	36.276	0.000	CG	89
1698	2.319	0.000	HG	87
1699	34.178	0.000	CG	87
1700	4.005	0.000	HA	86
1701	2.057	0.000	HB	86
1702	1.551	0.000	HG	86
1703	27.046	0.000	CG	86
1704	0.769	0.000	HD1	86
1705	0.801	0.000	HD2	86
1706	25.158	0.000	CD1	86
1707	25.437	0.000	CD2	86
1708	2.000	0.000	HB	85
1709	1.677	0.000	HG	85
1710	28.025	0.000	CG	85
1711	3.220	0.000	HD	85
1712	43.478	0.000	CD	85
1713	3.190	0.000	HB2	84
1714	63.337	0.000	CB	82
1715	0.804	0.000	HG	79
1716	26.766	0.000	CG	79
1717	0.901	0.000	HD	79
1718	25.158	0.000	CD1	79
1719	1.699	0.000	HG	78
1720	27.093	0.000	CG	78
1721	0.867	0.000	HD1	78
1722	0.893	0.000	HD2	78
1723	24.528	0.000	CD2	78
1724	24.109	0.000	CD1	78
1726	1.563	0.000	HG3	77
1727	1.753	0.000	HG2	77

1728	27.046	0.000	CG	77
1729	3.158	0.000	HD	77
1730	43.408	0.000	CD	77
1731	41.829	0.000	CB	76
1732	1.354	0.000	HG	76
1733	27.255	0.000	CG	76
1734	0.548	0.000	HD1	76
1735	0.595	0.000	HD2	76
1736	22.780	0.000	CD2	76
1737	23.200	0.000	CD1	76
1738	60.400	0.000	CB	75
1739	3.192	0.000	HD3	73
1740	3.312	0.000	HD2	73
1741	43.528	0.000	CD	73
1742	1.649	0.000	HG3	73
1743	1.702	0.000	HG2	73
1744	28.164	0.000	CG	73
1745	31.731	0.000	CB	72
1746	0.707	0.000	HG1	72
1747	1.007	0.000	HG2	72
1748	21.871	0.000	CG1	72
1749	24.109	0.000	CG2	72
1750	1.662	0.000	HG	69
1752	0.905	0.000	HD1	69
1753	0.926	0.000	HD2	69
1754	27.232	0.000	CG	69
1755	23.252	0.000	CD1	69
1756	25.130	0.000	CD2	69
1757	1.503	0.000	HB3	66
1758	1.729	0.000	HB2	66
1759	1.323	0.000	HG3	66
1760	1.390	0.000	HG2	66
1761	27.331	0.000	CG	66
1762	3.002	0.000	HD	66
1763	43.519	0.000	CD	66
1764	2.965	0.000	HE	65
1765	1.433	0.000	HG	65
1766	24.676	0.000	CG	65
1767	1.640	0.000	HD	65
1768	28.950	0.000	CD	65
1769	42.224	0.000	CE	65
1770	2.964	0.000	HB3	63
1771	0.736	0.000	HG	62
1772	26.425	0.000	CG	62
1773	0.698	0.000	HD1	62
1774	21.957	0.000	CD1	62
1776	2.385	0.000	HG3	59
1777	2.446	0.000	HG2	59
1778	33.936	0.000	CG	59
1779	2.490	0.000	HG	58
1780	36.720	0.000	CG	58
1781	1.106	0.000	HG2	57
1782	21.374	0.000	CG2	57
1783	0.805	0.000	HG1	56
1784	0.963	0.000	HG2	56
1785	21.050	0.000	CG1	56
1786	24.417	0.000	CG2	56

136

1787	1.700	0.000	HG	54	1854	2.148	0.000	HB2	32
1788	27.108	0.000	CG	54	1855	2.389	0.000	HG2	32
1789	0.866	0.000	HD1	54	1856	36.276	0.000	CG	32
1790	0.894	0.000	HD2	54	1857	2.076	0.000	HB3	32
1791	24.047	0.000	CD1	54	1858	2.209	0.000	HG3	32
1792	24.522	0.000	CD2	54	1859	2.375	0.000	HG	31
1794	0.823	0.000	HG2	52	1860	34.066	0.000	CG	31
1797	17.885	0.000	CG2	52	1861	1.702	0.000	HG12	30
1798	0.701	0.000	HD1	52	1862	1.007	0.000	HG13	30
1799	16.697	0.000	CD1	52	1863	0.751	0.000	HG2	30
1801	1.841	0.000	HG1	52	1864	17.256	0.000	CG2	30
1802	32.368	0.000	CG1	52	1865	0.791	0.000	HD1	30
1804	1.442	0.000	HG	51	1866	13.061	0.000	CD1	30
1805	24.808	0.000	CG	51	1867	28.968	0.000	CG1	30
1806	1.643	0.000	HD	51	1868	2.907	0.000	HE3	28
1807	29.025	0.000	CD	51	1869	42.083	0.000	CE	28
1808	2.918	0.000	HE3	51	1870	1.607	0.000	HD3	28
1809	2.963	0.000	HE2	51	1871	29.300	0.000	CD	28
1810	42.202	0.000	CE	51	1873	24.980	0.000	CG	28
1811	2.390	0.000	HG	50	1874	1.635	0.000	HD2	28
1812	34.036	0.000	CG	50	1875	1.447	0.000	HG2	28
1814	2.228	0.000	HB	46	1876	1.505	0.000	HG3	28
1815	0.865	0.000	HG1	46	1877	2.960	0.000	HE2	28
1816	0.939	0.000	HG2	46	1878	1.157	0.000	HG2	27
1817	20.962	0.000	CG1	46	1879	21.664	0.000	CG2	27
1818	24.407	0.000	CG2	46	1880	0.761	0.000	HG1	26
1819	1.268	0.000	HG3	41	1881	0.965	0.000	HG2	26
1820	1.351	0.000	HG2	41	1882	21.746	0.000	CG1	26
1821	24.689	0.000	CG	41	1883	23.338	0.000	CG2	26
1822	1.603	0.000	HD	41	1884	2.454	0.000	HG2	25
1823	29.258	0.000	CD	41	1885	33.964	0.000	CG	25
1824	2.912	0.000	HE3	41	1886	2.407	0.000	HG3	25
1825	2.961	0.000	HE2	41	1887	2.219	0.000	HB2	25
1826	42.124	0.000	CE	41	1888	2.379	0.000	HG	24
1827	2.365	0.000	HG	39	1889	36.250	0.000	CG	24
1828	36.359	0.000	CG	39	1890	67.532	0.000	CB	23
1829	2.406	0.000	HG3	38	1891	1.110	0.000	HG2	23
1830	2.453	0.000	HG2	38	1892	21.382	0.000	CG2	23
1831	33.970	0.000	CG	38	1893	3.149	0.000	HE	22
1832	1.522	0.000	HG3	35	1894	43.408	0.000	CE	22
1833	1.677	0.000	HG2	35	1895	1.670	0.000	HD	22
1834	27.220	0.000	CG	35	1896	27.185	0.000	CD	22
1835	3.154	0.000	HD	35	1897	1.693	0.000	HG	22
1836	43.417	0.000	CD	35	1898	25.927	0.000	CG	22
1837	0.601	0.000	HD1	34	1899	0.869	0.000	HD1	21
1838	0.704	0.000	HD2	34	1900	24.039	0.000	CD1	21
1839	22.790	0.000	CD1	34	1901	0.893	0.000	HD2	21
1840	1.631	0.000	HG	34	1902	24.528	0.000	CD2	21
1843	26.798	0.000	CG	34	1904	27.084	0.000	CG	21
1845	25.462	0.000	CD2	34	1905	1.701	0.000	HG	21
1846	1.140	0.000	HG	33	1908	29.913	0.000	CG1	20
1847	0.555	0.000	HD1	33	1909	13.200	0.000	CD1	20
1848	27.149	0.000	CG	33	1910	17.046	0.000	CG2	20
1850	23.181	0.000	CD1	33	1911	1.000	0.000	HG1	19
1851	0.597	0.000	HD2	33	1912	1.006	0.000	HG2	19
1852	22.790	0.000	CD2	33	1913	22.361	0.000	CG1	19
1853	4.016	0.000	HA	32	1914	24.109	0.000	CG2	19

1915	1.878	0.000	HB3	17		2011	1.993	0.000	HB3	47
1916	2.329	0.000	HG	17		2012	30.472	0.000	CB	47
1917	34.248	0.000	CG	17		2013	3.562	0.000	HD3	47
1919	1.949	0.000	HB2	17		2014	3.640	0.000	HD2	47
1920	1.190	0.000	HG2	16		2015	49.142	0.000	CD	47
1921	23.002	0.000	CG2	16		2016	2.093	0.000	HG3	47
1922	2.458	0.000	HB3	8		2017	2.269	0.000	HG2	47
1923	2.633	0.000	HB2	8		2018	27.955	0.000	CG	47
1924	1.413	0.000	HB3	13		2019	137.322	0.000	N	47
1925	1.638	0.000	HB2	13		2020	4.413	0.000	HA	64
1926	0.983	0.000	HD1	13		2021	2.373	0.000	HB	64
1928	1.179	0.000	HD2	13		2022	62.467	0.000	CA	64
1929	1.826	0.000	HG	13		2023	32.710	0.000	CB	64
1930	32.710	0.000	CG	13		2024	1.989	0.000	HG2	64
1931	26.276	0.000	CD1	13		2025	28.164	0.000	CG	64
1932	23.689	0.000	CD2	13		2026	3.571	0.000	HD3	64
1934	0.802	0.000	HD1	5		2027	3.615	0.000	HD2	64
1935	0.856	0.000	HD2	5		2028	50.471	0.000	CD	64
1936	23.410	0.000	CD1	5		2029	1.951	0.000	HG3	64
1937	24.598	0.000	CD2	5		2030	134.870	0.000	N	64
1938	1.542	0.000	HG	5		2031	4.541	0.000	HA	67
1939	27.046	0.000	CG	5		2032	2.474	0.000	HB2	67
1940	2.533	0.000	HB3	10		2033	1.700	0.000	HB3	67
1941	2.764	0.000	HB2	10		2034	62.888	0.000	CA	67
1942	1.814	0.000	HB3	97		2036	32.080	0.000	CB	67
1943	1.987	0.000	HB2	97		2037	136.983	0.000	N	67
1944	2.089	0.000	HG3	97		2038	4.295	0.000	HA	71
1945	2.128	0.000	HG2	97		2039	1.627	0.000	HB3	71
1946	35.577	0.000	CG	97		2040	2.355	0.000	HB2	71
1947	1.342	0.000	HB	100		2041	65.950	0.000	CA	71
1966	2.027	0.000	HB3	9		2042	31.448	0.000	CB	71
1968	62.814	0.000	CA	9		2044	28.164	0.000	CG	71
1969	31.987	0.000	CB	9		2045	3.671	0.000	HD3	71
1970	1.689	0.000	HB2	9		2046	3.788	0.000	HD2	71
1971	4.486	0.000	HA	9		2047	50.471	0.000	CD	71
1972	1.895	0.000	HG3	9		2048	1.947	0.000	HG3	71
1974	27.325	0.000	CG	9		2049	1.999	0.000	HG2	71
1975	3.626	0.000	HD3	9		2050	133.890	0.000	N	71
1976	3.764	0.000	HD2	9		2051	4.286	0.000	HA	96
1977	50.541	0.000	CD	9		2052	1.936	0.000	HB3	96
1978	1.995	0.000	HG2	9		2053	2.237	0.000	HB2	96
1979	135.248	0.000	N	9		2054	63.986	0.000	CA	96
1995	4.431	0.000	HA	14		2055	31.737	0.000	CB	96
1996	61.366	0.000	CA	14		2056	3.627	0.000	HD3	96
1997	2.036	0.000	HB3	14		2057	3.759	0.000	HD2	96
1998	1.751	0.000	HB2	14		2058	50.471	0.000	CD	96
1999	32.500	0.000	CB	14		2059	1.951	0.000	HG3	96
2000	2.883	0.000	HD3	14		2060	1.995	0.000	HG2	96
2001	3.140	0.000	HD2	14		2063	27.395	0.000	CG	96
2002	50.191	0.000	CD	14		2064	133.852	0.000	N	96
2003	0.444	0.000	HG3	14		2066	140.076	0.000	N	95
2004	1.013	0.000	HG2	14		2067	4.395	0.000	HA	101
2005	26.137	0.000	CG	14		2068	1.818	0.000	HB3	101
2006	132.305	0.000	N	14		2069	2.164	0.000	HB2	101
2007	4.404	0.000	HA	47		2070	62.917	0.000	CA	101
2009	66.226	0.000	CA	47		2071	31.923	0.000	CB	101
2010	2.326	0.000	HB2	47		2072	3.576	0.000	HD3	101

```
2073   3.763 0.000 HD2  101          2207   2.476 0.000 HB3   18
2074  50.331 0.000 CD   101          2208   2.696 0.000 HB2   18
2075   1.947 0.000 HG3  101          2224   6.614 0.000 HD2   29
2076   1.996 0.000 HG2  101          2239   7.242 0.000 HD1   40
2077  27.325 0.000 CG   101          2249   7.037 0.000 HD1   45
2078 134.908 0.000 N    101          2280   7.298 0.000 HE    63
2079   4.258 0.000 HA    98          2330   6.853 0.000 HD1  104
2080   1.854 0.000 HB3   98          2343   7.081 0.000 HD   116
2081   2.246 0.000 HB2   98          2361   8.134 0.000 H    126
2082  64.020 0.000 CA    98          2362 121.593 0.000 N    126
2083  31.673 0.000 CB    98          2365  67.315 0.000 CA   126
2084   3.633 0.000 HD3   98          2366  31.682 0.000 CB   126
2085   3.762 0.000 HD2   98          2367   3.520 0.000 HA   126
2086  50.471 0.000 CD    98          2368   2.278 0.000 HB   126
2087   1.949 0.000 HG3   98          2369   0.879 0.000 HG1  126
2088   1.998 0.000 HG2   98          2370   1.037 0.000 HG2  126
2089  27.325 0.000 CG    98          2371  24.038 0.000 CG2  126
2090 135.889 0.000 N     98          2372  21.229 0.000 CG1  126
2091   4.604 0.000 HA    11          2387   3.849 0.000 HA2   12
2092   2.100 0.000 HB    11          2394   0.828 0.000 HD1  107
2093  63.331 0.000 CA    11          2395  25.481 0.000 CD1  107
2094  32.313 0.000 CB    11          2396   1.687 0.000 HG   107
2095   3.834 0.000 HD3   11          2397  27.301 0.000 CG   107
2096   3.951 0.000 HD2   11          2403   7.136 0.000 HD    63
2097 138.303 0.000 N     11          2404   2.834 0.000 HB2   63
2103  50.680 0.000 CD    11          2405   6.703 0.000 HE   116
2104   1.914 0.000 HG    11          2406   6.353 0.000 HE1   42
2105  26.067 0.000 CG    11          2407   2.460 0.000 HB2   42
2200   7.484 0.000 H     18          2408   7.156 0.000 HD    84
2201 119.075 0.000 N     18          2409   6.778 0.000 HE    84
2202   4.471 0.000 HA    18          2410   6.740 0.000 HD2   53
2204  57.452 0.000 CA    18          2411   4.232 0.000 HB    57
2206  40.268 0.000 CB    18
```

VITA

Ziwei Zhang was born in Bazhong, Sichuan Province, China in 1979. He completed his elementary and middle school education in the same place. In 1998, Ziwei attended Shanghai Medical University (Merged with Fudan University in 1999), major in Clinical Medicine. In 2003, Ziwei obtained his Bachelor of Medicine in Clinical Medicine with honors. In August 2003, he joined the Department of Molecular Sciences at the University of Tennessee Health Science Center as a graduate student to pursue his Ph.D. degree. Through a joint program, he later joined the lab of Dr. Jie Zheng in the Department of Structural Biology at St. Jude Children's Research Hospital, Memphis TN. His research focused on the application of Nuclear Magnetic Resonance (NMR) spectroscopy in studying the structure and function of important proteins in childhood cancers. He received his Doctor of Philosophy degree from University of Tennessee in May of 2008.