

BIOCHEMICAL AND STRUCTURAL STUDIES OF  
HIV-1 MA MEMBRANE INTERACTION  
AND UBIQUITIN RECOGNITION

by

Bin Wang

A dissertation submitted to the faculty of  
The University of Utah  
in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

Department of Biochemistry

The University of Utah

May 2003

Copyright © Bin Wang 2003

All Rights Reserved

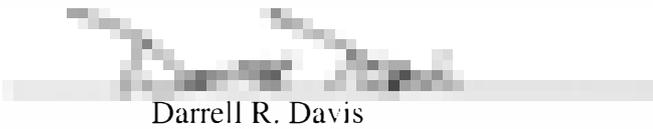
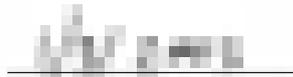
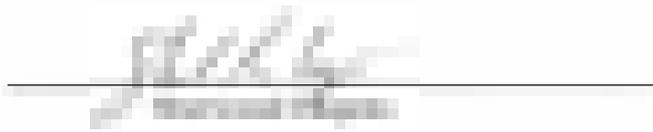
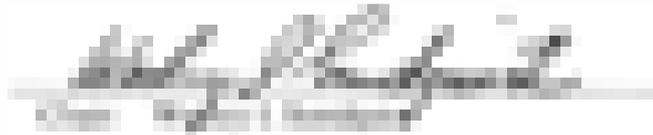
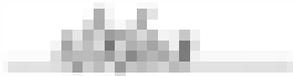
THE UNIVERSITY OF UTAH GRDUATE SCHOOL

## SUPERVISORY COMMITTE APPROVAL

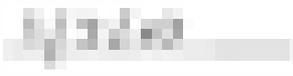
of a dissertation submitted by

Bin Wang

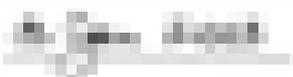
This dissertation has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.



Darrell R. Davis



Christopher Hill



Stephen M. Prescott

THE UNIVERSITY OF UTAH GRDUATE SCHOOL

**FINAL READING APPROVAL**

To the Graduate Council of the University of Utah:

I have read the dissertation of \_\_\_\_\_ Bin \_\_\_\_\_ in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the supervisory committee and is ready for submission to The Graduate School.

\_\_\_\_\_  
Date

\_\_\_\_\_  
Chair: Supervisory Committee

Approved for the Major Department

\_\_\_\_\_  
Dana Carroll  
Chair

Approved for the Graduate Council

\_\_\_\_\_  
David S. Chapman  
Dean of The Graduate School

## ABSTRACT

This dissertation describes biochemical and structural studies of HIV-1 MA membranes interactions (Chapter 2) and two different conserved ubiquitin binding motifs: the Ubiquitin Interaction Motif (UIM, Chapter 3) and the Npl4 Zinc Finger (NZF, Chapter 4).

HIV-1 MA is the N-terminal domain of Gag and is largely responsible for Gag membrane targeting and membrane binding. Both the N-myristoyl group of MA and a highly basic region near the N-terminus are important for MA membrane binding. NMR spectroscopic studies were used to define the sites within MA that change structure upon myristoylation. These studies indicated that the myristoyl group may pack against the N-terminal helices I and II and generally support a myristoyl switch model for MA membrane binding. Biochemical and NMR spectroscopic studies were also performed to map the MA binding site of PtdIns(3,4,5)P<sub>3</sub>, an important regulator for Gag assembly. The binding site is composed of basic residues that also comprise the membrane binding surface of MA and corresponds well with the sulfate binding site observed in the crystal structure of the MA trimer. This interaction may help target Gag to the plasma membrane for viral assembly.

Ubiquitylation is an important posttranslational modification and regulates many important cellular processes. Ubiquitylated substrate proteins are frequently recognized by downstream effector proteins that contain conserved ubiquitin binding motifs. The

UIM and NZF motifs are two such recently identified ubiquitin binding motifs.

Biochemical analyses demonstrated that both UIM and NZF motifs bind directly and specifically to ubiquitin. They both interact with ubiquitin through a hydrophobic patch surrounding residue I44, highlighting the importance of the I44 surface for ubiquitin function. A crystal structure of the Vps27p UIM revealed that the UIM forms an amphipathic helix. The solution structure of the NZF domain was determined using NMR. The structure is comprised of four antiparallel  $\beta$ -strands linked by three well ordered loops. The ubiquitin binding site on NZF is also primarily hydrophobic, and chemical shift mapping and sequence conservation analyses demonstrated that a T $\Phi$  dipeptide sequence in the ubiquitin binding site is highly conserved and forms the major recognition site for ubiquitin.

To my family.

## TABLE OF CONTENTS

ABSTRACT.....	iv
LIST OF FIGURES .....	ix
LIST OF TABLES.....	xi
ACKNOWLEDGEMENTS.....	xii
Chapter	
1. INTRODUCTION .....	1
HIV and AIDS.....	2
HIV-1 Genome Organization and Life Cycle .....	3
HIV-1 MA Protein and Viral Assembly.....	11
Ubiquitin and HIV-1 Budding.....	16
Ubiquitin Recognition and Ubiquitin Binding Domains.....	21
Specific Aims of Dissertation Research.....	24
References .....	25
2. HIV-1 MA MEMBRANE INTERACTION.....	38
Introduction .....	39
Materials and Methods .....	51
Results .....	57
Discussion .....	81
References .....	84
3. STRUCTURE AND UBIQUITIN BINDING OF THE UBIQUITIN INTERACTION MOTIF .....	92
Abstract .....	93
Introduction .....	93
Experimental Procedures.....	97
Results and Discussion.....	102
References .....	124

4. STRUCTURE AND UBIQUITIN INTERACTIONS OF THE CONSERVED NZF DOMAIN OF NPL4 .....	130
Abstract .....	131
Introduction .....	131
Experimental Procedures.....	135
Results .....	142
Discussion .....	159
Acknowledgements .....	165
References .....	165
5. SUMMARY .....	172
References .....	178

## LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1.1 Genome organization and virion structure of HIV-1 .....	4
1.2 Schematic illustration of HIV-1 life cycle. ....	7
2.1 Molecular structures of phosphatidylinositol-3,4,5-triphosphates and inositol-1,3,4,5,6-pentakisphosphate .....	49
2.2 HIV-1 Gag domain organization and a summary of the MA and MACA protein constructs used in this study .....	52
2.3 Purification and characterization of Myr-MA protein. ....	59
2.4 Purification and characterization of Myr-MA <sub>111</sub> (L8A) protein .....	61
2.5 Surface plasmon resonance biosensor analysis of the membrane binding of HIV-1 MA and MACA proteins .....	63
2.6 <sup>1</sup> H/ <sup>15</sup> N HSQC spectrum of the Myr-MA <sub>111</sub> (L8A) protein.....	68
2.7 Strips from the 3D <sup>15</sup> N-edited NOESY-HSQC data of the Myr-MA <sub>111</sub> (L8A) protein, showing the i to i+1 NOE connectivity.....	70
2.8 Strips from the HNCACB data of the Myr-MA <sub>111</sub> (L8A) protein, showing the intra and inter residue C <sub>α</sub> and C <sub>β</sub> correlations .....	72
2.9 Site of MA structural changes upon N-myristoylation .....	74
2.10 Overlay of the <sup>15</sup> N HSQC spectra of MA <sub>108</sub> showing the effect of titrating in PtdIns(3,4,5)P <sub>3</sub> .....	77
2.11 Crystal structure of the HIV-1 MA trimer (Hill et al., 1996), showing the sulfate (and PtdIns(3,4,5)P <sub>3</sub> ) binding sites.....	79
3.1 Quantification of ubiquitin/UIM peptide interactions.....	104

3.2	NMR chemical shift mapping of the interaction surfaces of ubiquitin with different UIM motifs .....	111
3.3	Structure of the Vps27p-2 UIM.....	117
3.4	Equilibrium sedimentation analysis of the oligomeric state of Vps27p-2 in solution .....	120
4.1	Primary sequence, conservation, and zinc binding by the Npl4 NZF domain.....	133
4.2	NMR assignments for the Npl4 NZF domain .....	143
4.3	Structure of the Npl4 NZF domain.....	147
4.4	Structural similarity between the Npl4 NZF Zn site and the rubredoxin knuckles of other metalloproteins.....	149
4.5	Quantitation of ubiquitin binding by NZF domains.....	154
4.6	NMR chemical shift mapping of the interaction surfaces of ubiquitin and Npl4 NZF .....	157

## LIST OF TABLES

<u>Table</u>	<u>Page</u>
3.1 UIM constructs and cloning .....	98
3.2 Ubiquitin binding affinities of various UIM peptides .....	103
3.3 Ubiquitin binding affinities of mutant Hrs UIM peptides .....	109
3.4 X-ray crystallographic data statistics .....	116
4.1 Structure statistics for Npl4 NZF domain .....	146
4.2 Possible structural roles for conserved NZF residues .....	152

## ACKNOWLEDGEMENTS

I am extremely grateful to my advisor Wesley I. Sundquist for his encouragement and guidance in science. I am also very grateful to my committee members, Drs. Sherwood Casjens, Darrell R. Davis, Christopher P. Hill, and Stephen M. Prescott, for their advice and support. I thank my collaborators for their contributions to the work described in this dissertation: Robert D. Fisher and Dan Higginson for the UIM project, Steven L. Alam, Marielle Payne, Timothy L. Stemmler, and Hemmo H. Meyer for the NZF project, Rebecca Rich and David Myszka for BIAcore analyses, and Dennis Winge for atomic absorption spectroscopic analyses. I would also like to thank the members of the Sundquist Lab for their help. Finally, I am forever indebted to my parents and my wife for their endless love and support.

## CHAPTER 1

### INTRODUCTION

## **HIV and AIDS**

The human immunodeficiency virus (HIV) infects cells in the human immune system. Progressive HIV infection leads to their depletion, and ultimately causes acquired immune deficiency syndrome (AIDS), in which the patient's immune system has been destroyed and the patient is vulnerable to opportunistic infections. Since its discovery in the early 1980s, HIV/AIDS has become devastating for humankind. With 25 million people killed by the end of year 2001, HIV/AIDS has become the deadliest infectious disease and the fourth leading cause of death worldwide (UNAIDS, 2001). Moreover, it is estimated that over 40 million people are now living with HIV/AIDS, and most of them are likely to die over the next decade (UNAIDS, 2001). Thus the HIV/AIDS epidemic poses an unprecedented threat to the health, economy, and security of humankind. Indeed, the impact of the HIV/AIDS epidemic can be felt in almost every aspect of human society.

HIV was first discovered in blood samples of AIDS patients in the early 1980s (Barre-Sinoussi et al., 1983; Gallo et al., 1984; Sarngadharan et al., 1984). In the intervening two decades, HIV has been the subject of intensive study in a number of labs. A great deal has been learned about the virus, and two classes of drugs have been developed, which target the viral reverse transcriptase and protease. Combination chemotherapy with these two classes of drugs is now an effective therapy for treating HIV infection (Balter, 1996; Richman, 2001). However, although combination chemotherapy can greatly reduce a patient's viral load and decrease the death rate from AIDS, it does not eliminate the virus. Moreover, drug-resistant viral strains are rapidly emerging, highlighting the importance of new antiviral drug development. Recent

developments in HIV research focused on several new drug targets, including the viral integrase (Hazuda et al., 2000), the zinc finger of the viral nucleocapsid protein (Huang et al., 1998; Rice et al., 1997), the viral regulatory proteins Rev and Tat (Hwang et al., 1999; Qian-Cutrone et al., 1996; Ratmeyer et al., 1996; Tamilarasu et al., 2000; Zapp and Green, 1989), and the viral fusion process (Kilby et al., 1998; Moore and Stevenson, 2000; Root et al., 2001). New antiviral drugs are now being developed against all of these targets. We anticipate that a detailed understanding of the biochemistry of other stages of the HIV replication cycle will provide additional potential drug targets, and this provides a motivation for much of the work in our laboratory.

### **HIV-1 Genome Organization and Life Cycle**

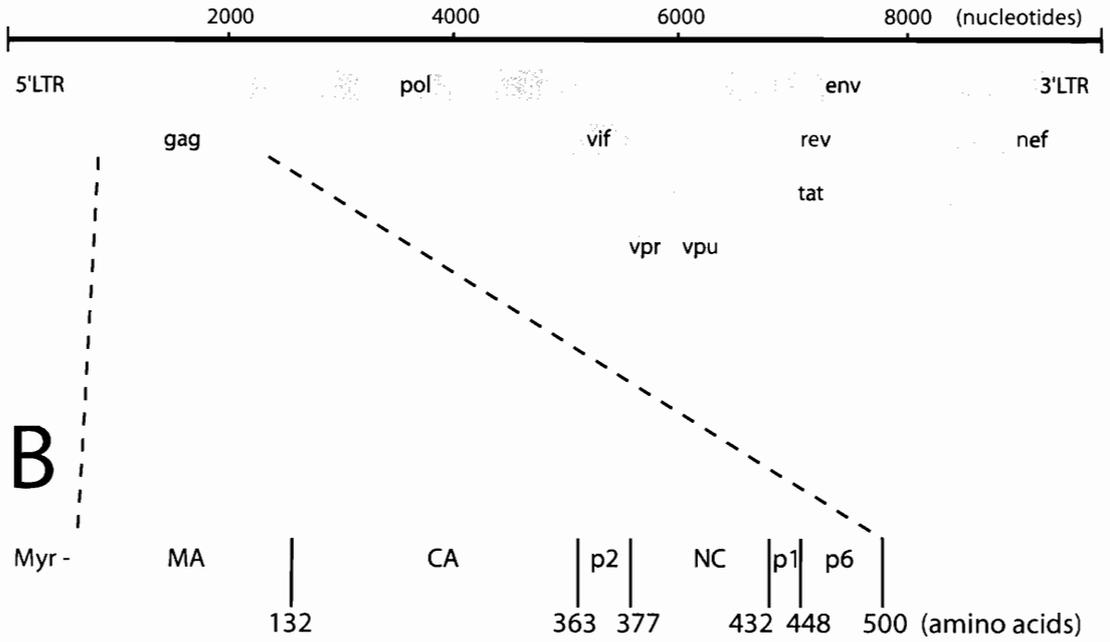
HIV is a retrovirus of the lentivirus genus (Coffin et al., 1997). Two distinct strains of HIV are currently recognized: HIV-1 and HIV-2, and the predominant virus worldwide is HIV-1.

HIV-1 is a single-stranded, positive sense RNA virus (Frankel and Young, 1998). Its ~9.6 kb genome contains 9 open reading frames and encodes 15 proteins (Figure 1.1 A). Three of the open reading frames, Gag, Pol, and Env are common to all retroviruses. These three proteins provide the structural and enzymatic functions of HIV-1, and each of these proteins is synthesized as a polyprotein precursor that is later processed into a series of small proteins.

Gag is the major structural protein of the virus and contains four distinct regions, termed: MA, CA, NC, and p6 (Figure 1.1B), which are ultimately processed into distinct proteins during viral maturation. In the mature virion (Figure 1.1 C), MA associates with the viral membrane and forms a protein shell (the “matrix”) underneath the lipid bilayer.

Figure 1.1. Genome organization and virion structure of HIV-1. A) Genome organization of HIV-1. B) Domain structure of HIV-1 Gag. C) Structure of the mature HIV-1 virion.

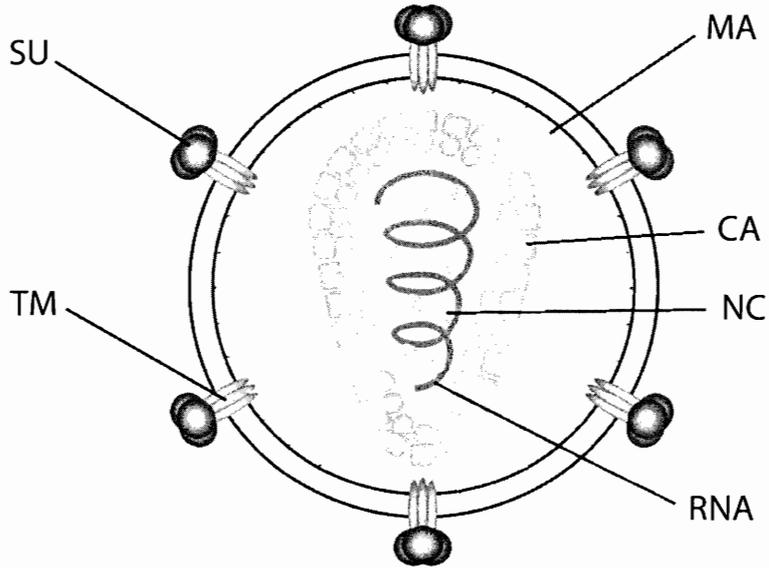
A



B



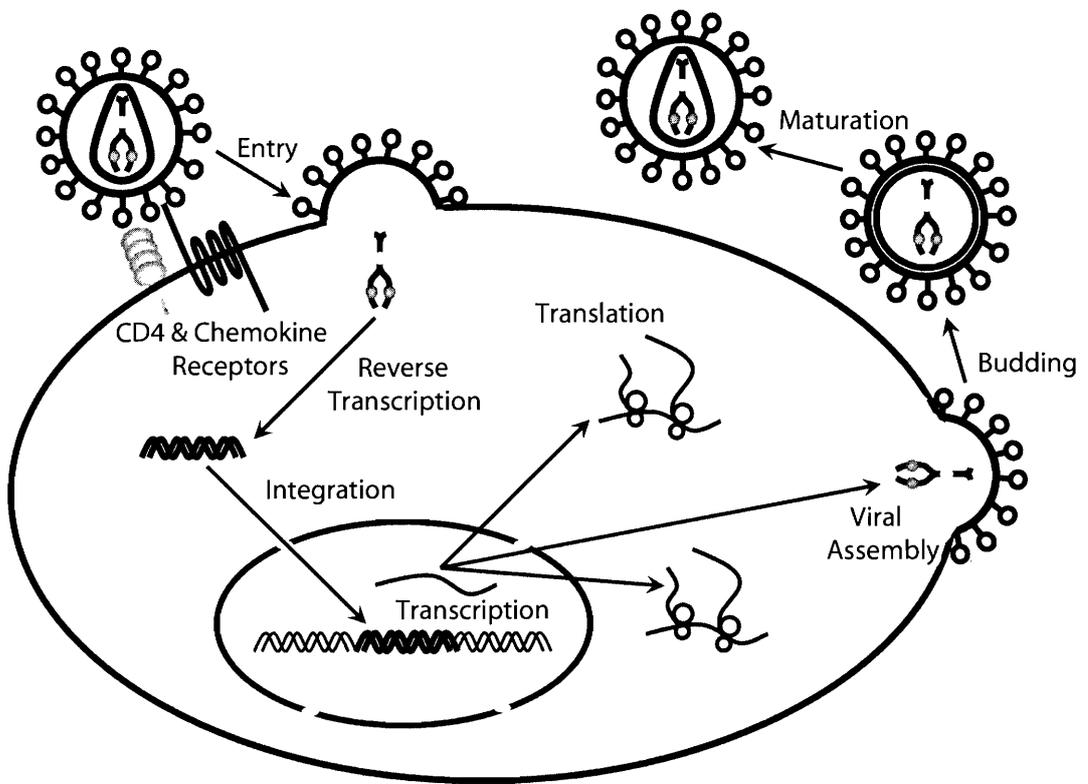
C



CA forms the cone shaped viral “capsid” that surrounds a complex formed by the NC protein and the RNA genome. The enzymatic functions necessary for viral replication are provided by the Pol protein. Pol is initially expressed as a Gag-Pol polyprotein via a -1 ribosomal frame shifting event immediately upstream of the p6 coding region. Pol is then processed by the viral protease to create the three viral enzymes: protease (PR), integrase (IN), and reverse transcriptase (RT). Finally, Env is the precursor of the viral envelope proteins: the surface glycoprotein (SU or gp120) and the transmembrane glycoprotein (TM or gp41). SU and TM are displayed on the virion surface and facilitate the viral entry step. Six additional open reading frames in the HIV-1 genome encode a series of “accessory” proteins: Tat, Rev, Vif, Vpu, Nef, and Vpr. These proteins provide important regulatory functions at different stages of HIV-1 life cycle.

HIV-1 replication begins with the entry of the virus into the host cell (Chan and Kim, 1998) (Figure 1.2). The envelope glycoproteins, SU and TM, provide the receptor recognition and membrane fusion functions necessary for this process. The SU glycoprotein initiates viral entry by binding to the CD4 receptor on the surface of the macrophage or T cell (Lasky et al., 1987). This interaction causes conformational changes within the SU glycoprotein that allow SU to bind to the chemokine coreceptors (typically CCR5 or CXCR4) (Rizzuto et al., 1998; Trkola et al., 1996; Wu et al., 1996). Coreceptor binding loosens the SU/TM interaction, allowing a fusion peptide at the N-terminus of TM to be exposed and inserted into target cell membrane, and this triggers membrane fusion through a mechanism that is not yet well understood (Chan et al., 1998; Chan et al., 1997; Sullivan et al., 1998; Weissenhorn et al., 1997). Upon membrane fusion, the viral core is released into the host cell cytoplasm where the capsid

Figure 1.2. Schematic illustration of HIV-1 life cycle.



disassembles to expose the viral nucleoprotein complex. Within this replication complex, RT uses the packaged tRNA<sup>Lys3</sup> as a primer to produce a double stranded DNA copy of the viral RNA genome (Hsu and Wainberg, 2000). The replicating nucleoprotein complex (now termed the “preintegration complex”) is actively transported into the host cell nucleus (Bukrinsky et al., 1992). Several proteins associated with the preintegration complex (MA, IN, and Vpr) appear to facilitate this active nuclear import process (Bukrinsky et al., 1993; Galloway et al., 1997; Subbramanian et al., 1998). After entering the host cell nucleus, IN catalyzes integration of the viral DNA into the host cell chromosomes in a semirandom fashion (Esposito and Craigie, 1998; Farnet and Bushman, 1996; Miller et al., 1995).

Following integration, HIV-1 uses the transcriptional and translational machinery of the host cell to produce the viral RNAs and proteins encoded by the integrated provirus. As with cellular mRNA, proviral DNA transcription is performed by the RNA polymerase II complex. Viral transcripts are expressed from the HIV-1 proviral promoter, which is located in the U3 region of the 5' long terminal repeat (LTR). Transcription initiation requires a series of cellular transcription factors, including NF- $\kappa$ B, Sp1, Ets-1 and TBP (Jones and Peterlin, 1994). After transcription initiates, efficient transcriptional elongation requires the viral regulatory protein Tat, which binds to the Tat-activation response (TAR) RNA element in the 5' LTR and also to cyclin T, and thereby increases the phosphorylation and processivity of RNA polymerase II (Feng and Holland, 1988; Rittner et al., 1995; Zhou et al., 2000; Zhou and Sharp, 1995). At early stages of viral gene expression, the viral transcripts are fully spliced before being transported into the cytoplasm for translation. The fully spliced viral mRNA encodes only the regulatory

proteins Tat, Rev, and Nef. Both Tat and Rev contain nuclear localization signals and therefore return to the nucleus upon translation. Rev binds to the Rev-response element (RRE) in the Env coding region of the viral mRNA and enhances the nuclear export of unspliced or singly-spliced viral mRNAs (Cullen and Malim, 1991; Hope, 1997). Thus, the accumulation of Rev shifts viral gene expression from the early stage to the late stage, in which the structural and enzymatic proteins, Gag, Gag-Pol, and Env, are translated from the unspliced viral mRNAs.

Gag and Gag-Pol are synthesized in the cytoplasm by free ribosomes, and are then transported to the plasma membrane where they initiate viral assembly. A heptanucleotide “slippery” sequence and an RNA hairpin near the end of Gag coding region cause a -1 ribosomal frame shift at a frequency of ~5%, which allows the ribosome to bypass the stop codon at the end of Gag and produce the longer Gag-Pol polyprotein (Wilson et al., 1988). Gag trafficking to the plasma membrane is still not well understood, but recent reports indicate that the cytoskeleton and lipid rafts associated with the plasma membrane may play important roles in Gag trafficking (Lindwasser and Resh, 2001; Ono and Freed, 2001; Ott et al., 2000b). Synthesis of the transmembrane Env protein is carried out by ribosomes on the rough endoplasmic reticulum (RER). Env is glycosylated and cleaved by the cellular protease furin to produce the heterodimeric viral envelope protein (composed of SU and TM), which is then transported to the plasma membrane through the cellular secretory pathway (Moulard and Decroly, 2000). At the same time, HIV-1 regulatory proteins Vpu and Nef promote downregulation of CD4 on the host cell membrane in order to prevent the newly synthesized viral envelope proteins from binding to CD4 receptor molecules on the producer cell (Piguet et al., 1999).

Observable viral assembly initiates when the viral structural and enzymatic proteins are synthesized and transported to common sites on the plasma membrane. During this process, Gag must recruit the viral genomic RNA and make the necessary Gag-Gag interactions that create viral particles (Sakalian and Hunter, 1998). As they assemble, the Gag/Gag-Pol proteins form an electron dense patch underneath the plasma membrane, incorporate Env glycoproteins, induce membrane curvature, and ultimately form a spherical particle that pinches off from the cell plasma membrane. This newly formed viral particle is called an immature virion, and it has a spherical capsid located immediately underneath the viral membrane and an electron-lucent center.

Upon budding, the immature virion must go through a maturation process to become infectious. Viral maturation is characterized by a dramatic morphological change in which the immature viral capsid rearranges and forms the cone-shaped mature capsid (Gelderblom, 1991). Cleavage of the Gag and Gag-Pol polyprotein precursors by the viral protease (PR) is necessary to initiate this process. Upon cleavage, MA remains associated with the inner leaflet of the viral membrane, while the CA protein condenses to form the cone-shaped mature capsid that organizes the genomic RNA / NC protein complex. The mature virion is now infectious and HIV-1 life cycle starts again with the infection of a new target cell.

### **HIV-1 MA Protein and Viral Assembly**

Gag is the major structural protein of HIV-1 virus. As described above, Gag initiates viral assembly at the plasma membrane and proceeds through a series of steps to form the viral capsid. During viral assembly, Gag must also package several other viral components, including the viral RNA genome, viral protein R (Vpr), and cyclophilin A

(CypA) (Franke et al., 1994; Kondo and Gottlinger, 1996; Lu et al., 1995; Paxton et al., 1993). Gag also recruits essential cellular factors to assist viral assembly and budding (Garrus et al., 2001; VerPlank et al., 2001).

As shown in Figure 1.1A, Gag is synthesized as a polyprotein precursor that contains four distinct domains: MA, CA, NC and p6. The N-terminal MA domain is responsible for Gag trafficking and membrane binding. It also facilitates envelope protein incorporation by interacting with the cytoplasmic tail of the TM subunit. CA helps mediate the Gag-Gag interactions required for particle assembly and also recruits the cellular protein CypA. NC functions in RNA packaging through direct contacts between the two zinc fingers within NC and the packaging signal ( $\psi$ ) of the viral genomic RNA. The function of p6 is to recruit Vpr and also cellular factors involved in viral budding.

Gag is synthesized in the cytoplasm and then actively transported to the plasma membrane where viral assembly takes place. Plasma membrane targeting is an important step in the viral assembly process. Although the detailed steps required for Gag membrane targeting are not yet known, both MA and cellular factors are important for this process. For example, when HIV-1 Gag is expressed in murine cells, it fails to reach the plasma membrane and viral assembly is blocked (Mariani et al., 2000). However, plasma membrane targeting and viral assembly can be restored by replacing the MA domain of HIV-1 Gag with the MA domain of Moloney murine leukemia virus (Mo-MuLV) Gag (Chen et al., 2001a). Cellular factors involved in the plasma membrane targeting of Gag appear to include the cytoskeletal system, the ATP binding protein Hp68, and phosphatidylinositol phosphates (PtdInsP). Specifically, HIV-1 Gag and other retroviral Gag proteins have been shown to associate with the cytoskeleton and may

thereby be transported to plasma membrane (Edbauer and Naso, 1983; Edbauer and Naso, 1984; Perrin-Tricaud et al., 1999). Hp68 and PtdInsP were identified as potential viral assembly cofactors through in vitro assembly studies (Campbell et al., 2001; Campbell and Rein, 1999; Lingappa et al., 1997; Zimmerman et al., 2002). Although their functions are not yet fully understood, Hp68 may function as a molecular chaperone during Gag assembly, and PtdInsP may modulate Gag assembly through interactions with the N-terminal MA domain.

The plasma membrane binding functions of MA are mediated by a bipartite signal within MA that is composed of an N-terminal myristoyl group and a highly basic patch located near the protein's N-terminus. The myristoyl group is a 14 carbon saturated fatty acid chain that is co-translationally attached via an amide bond to the N-terminal glycine residue of MA. The myristoyl group is essential for Gag localization, as mutations of the N-terminal glycine residue that abolish the myristoylation of MA also block Gag membrane binding, viral assembly and replication (Bryant and Ratner, 1990; Gheysen et al., 1989; Gottlinger et al., 1989; Pal et al., 1990). In addition to the myristoyl group, mutational analyses have been used to show that a highly basic region near the N-terminus of MA is also important for Gag membrane targeting and membrane binding (Yuan et al., 1993; Zhou et al., 1994). This region presumably interacts with acidic phospholipid head groups on the surface of the plasma membrane.

The lower hydrophobicity of myristic acid as compared with the longer acyl chains of palmitic or stearic acid, means that myristoyl groups alone cannot provide sufficient energy for stable membrane binding (Kim et al., 1994; Peitzsch and McLaughlin, 1993; Sigal et al., 1994; Silvius and l'Heureux, 1994). Indeed, of the 100

proteins known to be myristoylated in yeast, about half are soluble and many bind membranes reversibly (Magee and Courtneidge, 1985; Olson et al., 1985). The myristoyl groups in these proteins are frequently involved in “myristoyl switches,” in which protein membrane binding is controlled by altering exposure of the myristoyl group.

The myristoyl group of HIV-1 MA may also help to control the membrane association of Gag. Myristoylated Gag is synthesized in the cytosol as a soluble protein, and must travel to the plasma membrane before initiating viral assembly. It is likely that the myristoyl group is sequestered by the N-terminal MA domain when Gag is a soluble protein in the cytosol, and then exposed for binding to the plasma membrane for viral assembly. Genetic studies have shown that mutations close to the N-terminus of MA can severely impair Gag membrane binding and viral assembly without affecting the myristoylation (Ono et al., 1997; Paillart and Gottlinger, 1999). These phenotypes are very similar to the phenotype of the myristoylation-defective Gag mutations, suggesting that the myristoyl group may be “locked” in a sequestered position by the former mutations and thereby made unavailable for membrane binding. Interestingly, this membrane binding defect can be rescued by second site mutations in the  $\alpha$ -helical core of MA (Paillart and Gottlinger, 1999), suggesting that the MA core may interact with the myristoyl group. In support of this model, large deletions of MA that remove most or all of the core have been shown to enhance Gag membrane binding and cause the virus to assemble on intracellular membranes (Facke et al., 1993; Spearman et al., 1997). It is likely that, in the absence of the MA core, the myristoyl group is constitutively exposed and this exposure enhances nonspecific Gag membrane binding. Other studies of HIV-1 MA membrane binding also support this myristoyl switch model. For example, Yu et al.

showed that MA is phosphorylated and rapidly translocates to membranes after treatment with phorbol ester (Yu et al., 1995), suggesting that the myristoyl switch in HIV-1 MA may be linked to protein phosphorylation. Other interactions, such as PtdInsP binding, may also play a role in controlling the MA myristoyl switch. However, the detailed mechanism of the MA myristoyl switch is not yet known, and studies of the biochemistry of myristoylated HIV-1 MA are described in Chapter 2 of this dissertation.

In addition to the myristoyl switch, the membrane binding of HIV-1 MA may also be modulated by other cellular factors, and recent studies have shown that PtdInsP can modulate Gag assembly through interactions with the MA domain (Campbell et al., 2001). HIV-1 Gag encodes all the information necessary for viral assembly, and expression of Gag alone in mammalian or insect cells is sufficient to form extracellular viral like particles (VLPs) (Gheysen et al., 1989; Royer et al., 1991; Royer et al., 1992; Shioda and Shibuta, 1990). Cryo-electro microscopy studies have shown that these VLPs have similar morphologies to immature virions (Fuller et al., 1997; Wilk et al., 2001). An in vitro particle assembly system that utilizes pure, recombinant HIV-1 Gag has also been developed (Campbell and Rein, 1999). However, the VLPs assembled from purified HIV-1 Gag protein are much smaller than authentic immature HIV-1 virions. Furthermore, unlike authentic virions, these VLPs are sensitive to high salt (0.5M NaCl), RNase treatment, and trypsin digestion (Campbell et al., 2001). However, these defects can be eliminated by adding mammalian cell lysates to the in vitro assembly mixture, and the active factor(s) in the lysates have been shown to be inositol phosphates (IP) and derivatives of PtdInsP (Campbell et al., 2001).

It is not clear yet how the IP and PtdInsP modulate HIV-1 Gag assembly. However, the region in Gag that is responsible for this modulation has been mapped to the MA domain between residues 16 and 99 (Campbell et al., 2001). Deletion of this region results in the assembly of VLPs that have correct size and are resistant to high salt and RNase treatment, suggesting that binding of IP or PtdInsP to this region of MA may overcome an element that inhibits proper viral assembly. One of the compounds that is active in the in vitro Gag assembly assay is dibutyl PtdIns(3,4,5)P<sub>3</sub>, a water soluble analog of PtdIns(3,4,5)P<sub>3</sub>. PtdIns(3,4,5)P<sub>3</sub> is a key element in many signal transduction pathways and is found in the plasma membrane where MA targets Gag for viral assembly. Although it is not yet clear how (or even if) MA binds to PtdInsP on the plasma membrane, such an interaction could play a role in targeting Gag to the plasma membrane and/or proper viral assembly. There is a precedent for this in other systems, where PtdInsP has been shown to help organize large macromolecular complexes on cell membranes (Ford et al., 2001; Itoh et al., 2001). Thus, we speculate that PtdInsP may play similar functions in HIV-1 viral assembly, and experiments aimed at understanding how PtdInsP binds to HIV-1 MA are also described in Chapter 2 of this dissertation.

### **Ubiquitin and HIV-1 Budding**

Ubiquitin is a small, highly conserved eukaryotic protein that plays a central role in a wide variety of cellular processes (Hershko and Ciechanover, 1998). Ubiquitin can be covalently attached to the lysine side chains of many proteins and serves as a complex posttranslational modification that can regulate the stability, subcellular localization, and function of the target protein. Protein ubiquitylation is catalyzed by a complex three-enzyme cascade, in which a ubiquitin molecule is passed from an

activating enzyme (E1) to a conjugating enzyme (E2), and finally onto a substrate protein through a ligase (E3) (or directly onto a substrate protein with the help of E3) (Ciechanover, 1994; Hershko and Ciechanover, 1992). In the first step, ubiquitin forms a high energy thio-ester bond with a cysteine residue of E1 in a ATP-dependent process. Ubiquitin is then transferred to an active site cysteine residue of an E2. In the final step, ubiquitin is transferred onto a lysine residue of the substrate in a reaction that is catalyzed by E3 either directly (e.g., as stimulated by the RING Finger E3 ligases) or via a covalent ubiquitin-E3 intermediate (e.g., in the HECT-domain E3 ligases). The exposed lysine residues (K29, K48, and K63) of ubiquitin itself can also be ubiquitylated and lead to the formation of poly-ubiquitin chains. The chain length and the linkage type are at least partially responsible for determining which cellular signal is specified by ubiquitylation (Dubiel and Gordon, 1999; Pickart, 2000).

The best characterized function of ubiquitylation is to target proteins for proteasomal degradation, which requires poly-ubiquitin chains of four subunits or longer linked through K48 (Hershko and Ciechanover, 1998). Ubiquitin-mediated protein degradation controls the levels of many important regulatory proteins, including cyclins, cyclin-dependent kinase inhibitors, and transcription factors. This regulated degradation, in turn, controls various cellular processes, such as cell-cycle progression and signal transduction.

Other types of ubiquitylation do not target proteins for proteasomal degradation, but rather regulate activities and target proteins to various cellular processes (Hicke, 2001). K63 linked poly-ubiquitin chains are important for DNA repair and other cellular functions (Pickart, 2000), and it has been shown that Ubc13, an E2 enzyme, can generate

K63 linked poly-ubiquitin chains (Hofmann and Pickart, 1999). Ubc13 together with Rad5 and Rad18 (two RING finger type E3 enzymes) are implicated in DNA repair as shown by genetic studies (Bailly et al., 1997; Ulrich and Jentsch, 2000). L28, a ribosomal subunit, is also a major substrate for K68 linked poly-ubiquitin chains, and the ubiquitylation of L28 enhances translation and is most prominent during the S phase of the cell cycle (Spence et al., 2000). In addition to poly-ubiquitylation, cellular proteins are also found to be modified by a single ubiquitin or by short ubiquitin chains, and this type of ubiquitylation has been shown to be involved in at least three distinct biological functions: histone regulation, endosomal protein trafficking, and retrovirus budding (Hicke, 2001; Pickart, 2001b).

It was first discovered about a decade ago that purified Rous sarcoma virus (RSV) particles contain approximately 100 molecules of ubiquitin, which represent a five-fold enrichment over the concentration of free ubiquitin in the cytosol (Putterman et al., 1990). Since then, similar amounts of ubiquitin have also been found in other retroviruses, including HIV-1, simian immunodeficiency virus (SIV), and murine leukemia virus (Ott et al., 1998). Moreover, a small percentage of the viral Gag proteins are found to be ubiquitylated (Ott et al., 2000a; Ott et al., 1998). Recent studies have established a functional link between ubiquitin and retroviral budding (Patnaik et al., 2000; Schubert et al., 2000; Strack et al., 2000). For example, when the free pool of ubiquitin is depleted by proteasome inhibitor treatment that prevents recycling of the conjugated ubiquitin, RSV and HIV-1 virus budding rates are significantly reduced. The block of RSV budding by proteasome inhibitors can be suppressed by the overexpression of ubiquitin and by the fusion of ubiquitin directly to the C-terminal end of the RSV Gag polyprotein (Patnaik et

al., 2000). In the absence of free ubiquitin, the assembled viral particles remain attached to the plasma membrane and to each other by thin membrane stalks (Patnaik et al., 2000; Schubert et al., 2000). The phenotype of the block to virus budding caused by proteasome inhibitors is very similar to the budding arrest caused by mutations in *cis*-acting signals within retroviral Gag proteins called Late domains (Gottlinger et al., 1991).

Late domains play a critical role in the final stages of retroviral budding, in which the assembled viral particles separate from the plasma membrane (Freed, 2002). Late domains in different retroviruses are composed of one of three highly conserved sequence motifs: PTAP, PPXY, or YXXL (where X can be any amino acid). These motifs are known to interact with cellular proteins and are essential for the Late domain function. Recent studies have also linked the function of Late domain in the retrovirus budding to ubiquitin. The PPXY Late domains of structural proteins of RSV, Ebola virus, and rhabdovirus interact with members of the Nedd4 family of E3 ubiquitin ligases (Harty et al., 2001; Harty et al., 2000; Kikonyogo et al., 2001), and all of these proteins are ubiquitylated at low levels (Ott et al., 2002; Paumet et al., 2000). The function of viral Late domains can also be substituted by a physiological interaction site for Nedd4, suggesting that ubiquitin ligase recruitment by the PPXY Late domain plays an important role in retroviral budding (Strack et al., 2000).

In contrast to PPXY-dependent recruitment of the Nedd4 family of E3 ubiquitin ligases, proteins in the endosomal protein sorting pathway are recruited by PTAP Late domain. Recent studies have shown that the cellular protein Tsg101 interacts with HIV-1 Late domain and is required for HIV-1 budding (Garrus et al., 2001). Originally identified as a tumor suppressor gene, Tsg101 also functions in various cellular processes,

including the sorting of proteins into the endosomal pathway. Cells deficient in the expression of Tsg101 and its yeast homolog Vps23p have defects in the sorting and maturation of vacuolar hydrolases, and in the downregulation of activated G protein coupled receptors and epidermal growth factor receptors (Babst et al., 2000; Katzmann et al., 2001; Li et al., 1999). HIV-1 budding is also dependent on Tsg101 expression (Garrus et al., 2001; Martin-Serrano et al., 2001), and virus budding in cells defective for Tsg101 expression arrests with a phenotype that is similar to the Late domain mutants. Tsg101 contains an ubiquitin E2 variant (UEV) domain that shares sequence similarities to active E2 enzymes but lacks the active site cysteine required for E2 activity. Although the Tsg101 UEV domain cannot conjugate ubiquitin, it can still bind ubiquitin and interact with ubiquitylated protein substrates of endosomal protein sorting (Garrus et al., 2001). Tsg101 and its yeast homolog, Vps23p, are part of a high molecular weight complex, called ESCRT-1, which recognizes monoubiquitylated cargo proteins at the late endosome and helps to target them to the lysosome via the multivesicular body (MVB) pathway (Bishop et al., 2002; Katzmann et al., 2001). As the late endosome matures into the MVB, regions of the limiting membrane invaginate and bud into the lumen of late endosome to form small vesicles. The vesiculated MVB then fuses with the lysosome and delivers these vesicles into the lumen of lysosome where they are degraded. Budding of retroviruses at the plasma membrane is topologically equivalent to the budding of vesicles into the lumen of the late endosome, and it is likely that ubiquitin interaction plays similar roles in MVB protein sorting and retrovirus budding. In both systems, it appears the ubiquitin sensors recognize and sort ubiquitylated substrates (either proteins

destined for the lysosome or Gag proteins destined for virions) and simultaneously recruit the machinery necessary to form MVB vesicles or virus particles.

### **Ubiquitin Recognition and Ubiquitin Binding Domains**

As a wide variety of cellular processes are regulated by ubiquitylation, ubiquitylated substrate proteins must be specifically recognized by a diverse set of downstream receptors and directed into the correct pathway. Ubiquitin is a small protein with well defined 3D structure (Vijay-Kumar et al., 1987; Vijay-Kumar et al., 1985). It has a five-stranded mixed  $\beta$ -sheet and an  $\alpha$ -helix that forms a compact globular domain with a short flexible C-terminal tail. As compared to other posttranslational modifications such as phosphorylation and acetylation, ubiquitylation is a much more complex modification, and cellular proteins may interact and recognize ubiquitin in many different ways. Studies of ubiquitin interaction proteins have led to the identification of several ubiquitin binding motifs, including UBA (ubiquitin associated domain), UEV (ubiquitin E2 variant domain), UIM (ubiquitin interacting motif), and NZF (Npl4 zinc finger domain) (Buchberger, 2002; Hofmann and Bucher, 1996; Hofmann and Falquet, 2001; Madura, 2002; Meyer et al., 2002; Pickart, 2001a).

UBA is a small protein motif of ~40 amino acids that was first identified in E2 and E3 enzymes. UBA domains have subsequently been identified in a number of proteins, including p62 (a ligand of the p56<sup>lck</sup> tyrosine kinase), Rad23/Rhp23 (a DNA repair protein), Ddi1/Mud1 (a DNA damage-induced protein), and Ede1 (an endosomal protein). (Bertolaet et al., 2001; Chen et al., 2001b; Funakoshi et al., 2002; Gagny et al., 2000; Rao and Sastry, 2002; Wilkinson et al., 2001). UBA domains also bind to other proteins in addition to ubiquitin. For example, the C-terminal UBA of the human DNA

repair protein HHR23A interacts with the HIV-1 Vpr protein and 3-methyladenine DNA glycosylase (MPG) (Majoul et al., 1998; Withers-Ward et al., 1997). The UBA is a compact three-helix bundle with several potential interacting sites, including an exposed hydrophobic surface patch (Dieckmann et al., 1998; Mueller and Feigon, 2002). However, the UBA/ubiquitin interaction is yet not understood in molecular detail.

UEV domains share significant sequence similarity to ubiquitin conjugating enzyme E2, but are missing the active site cysteine required for E2's enzymatic activity. Although UEV domains cannot conjugate ubiquitin, some UEV's retain the ability to bind ubiquitin and act as cofactors for ubiquitylation and ubiquitin sensors. The UEV protein, Mms2, forms a complex with E2 Ubc13, and this complex catalyzes the assembly of K63 linked polyubiquitin chains (Hofmann and Pickart, 1999). The UEV domain of Tsg101 interacts with ubiquitin, functions as a ubiquitin sensor in the endosomal pathway, and plays a similar role in HIV-1 virus budding (Garrus et al., 2001; Katzmann et al., 2001). Tsg101 is a component of the ESCRT-I protein complex, which is essential for sorting ubiquitylated protein substrates into the MVB. The structure of several UEV domains have been characterized; they share similar E2 fold but interact with ubiquitin in different ways (Moraes et al., 2001; Pornillos et al., 2002; VanDemark et al., 2001).

The UIM motif is based on a sequence of the S5a/Rpn10 subunit of the proteasome known to recognize ubiquitin (Young et al., 1998). Database searches have identified UIM motifs in a wide variety of proteins, including proteins involved in ubiquitylation, ubiquitin metabolism, and receptor mediated endocytosis (Hofmann and Falquet, 2001). Recently, the UIMs in endocytic proteins, such as Eps15, eps15R, epsins,

and Hrs, have been shown to bind ubiquitin and also to be responsible for the monoubiquitylation of these proteins (Polo et al., 2002). Thus, it appears that in this context the UIM may serve to attract ubiquitin as a substrate for conjugation onto the protein. In functional terms, the Epsin UIMs are important for internalization of receptors into clathrin-coated vesicles at the plasma membrane, and the Hrs UIM is required for targeting ubiquitylated proteins into MVB (Raiborg et al., 2002; Shih et al., 2002). These endocytic proteins may use their UIMs to recognize ubiquitylated cargo proteins and target them into the endocytosis pathway. The UIM is a short motif of ~20 amino acid residues that is unlikely to form an independent folding domain like UBA. Instead, it has been proposed that the motif forms a short  $\alpha$ -helix that can be embedded into different protein folds (Hofmann and Falquet, 2001). It is not clear yet how UIM recognizes and binds ubiquitin.

The NZF domain is a novel class of Zn fingers that also interact with ubiquitin (Meyer et al., 2002). NZF domains are about 30 amino acid residues long and contain four highly conserved cysteine residues. Vps36p, a protein involved in endosomal protein sorting, is one protein that contains an NZF domain. Vps36p is a component of ESCRT-II, an endosomal protein complex that is recruited by the ESCRT-I complex during MVB formation (Babst et al., 2002). The NZF domain in Vps36p may function as the ubiquitin sensor for ESCRT-II and thereby help to target ubiquitylated cargo protein into MVB. In addition to Vps36p, NZF domains are also found in a number of other proteins, including mammalian protein Npl4. Npl4 is a subunit of the heterodimeric UN complex (together with Ufd1), which functions in ER-associated degradation (ERAD) (Bays et al., 2001; Meyer et al., 2000). ERAD regulates the level of many ER resident proteins, such as

HMG-CoA reductase, and also serves as a quality control mechanism to remove newly synthesized ER proteins that fail to fold or assemble correctly (Brodsky and McCracken, 1999). Mutations of Npl4 are deficient in ER-associated degradation, but still capable of cytosolic protein degradation. In Npl4 mutant cells, the ER membrane protein HMG-CoA reductase is fully ubiquitylated but not degraded, indicating that the Npl4 functions following ubiquitylation of ER target proteins (Bays et al., 2001). In addition to interacting with ubiquitin through its NZF domain, the UN complex also recruits the AAA-ATPase p97/Cdc48, which has been shown to associate with the 26S proteasome in an ATP dependent manner (Dai et al., 1998; Verma et al., 2000). Although the detailed role of the UN complex in ERAD is not known, it is likely that UN functions by recognizing ubiquitylated target proteins through the NZF domain of Npl4, then recruiting p97/Cdc48, which in turn uses the energy of ATP hydrolysis to pull target proteins out of the membrane to be degraded or processed by the 26S proteasome.

Since both ubiquitin and the MVB pathway play important roles in HIV-1 budding, gaining an understanding of the molecular interactions between the UIM, NZF domains and ubiquitin is also expected to provide insight to the detailed functions of ubiquitin in HIV-1 budding. Biochemical and structural studies of UIM and NZF domains are described in Chapter 3 and 4.

### **Specific Aims of Dissertation Research**

The research described in this dissertation is focused in two areas: 1) the interaction between HIV-1 MA protein and the lipid membrane bilayer, and 2) ubiquitin recognition by UIM and NZF domains.

HIV-1 MA targets the Gag polyprotein to the plasma membrane for viral assembly. The membrane targeting and binding functions of MA involve a highly basic region close to its N-terminus, the N-myristoyl group, and possibly PtdInsP binding. Chapter 2 describes the biochemical studies of MA membrane binding, structural studies of myristoylated MA, and mapping of the binding sites of PtdInsP on MA.

Ubiquitin is involved in a wide variety of biological processes, including retroviral budding. UIM and NZF domains are found in many proteins known to interact with ubiquitin and appear to be responsible for ubiquitin recognition in these proteins. Chapter 3 describes studies of the UIM/ubiquitin interaction and mapping of the UIM binding site of UIM on ubiquitin. Chapter 4 describes studies of the NZF/ubiquitin interaction and the structure of the Npl4 NZF domain.

Chapters 3 and 4 are written as independent manuscripts. Work described in Chapter 3 was in collaboration with Robert D. Fisher of Christopher P. Hill's laboratory. I performed the work on ubiquitin/UIM interaction and the NMR chemical shift mapping studies and Robert D. Fisher determined the Vps27p crystal structure and performed the studies of UIM oligomerization.

## References

- Babst, M., Katzmann, D. J., Snyder, W. B., Wendland, B., and Emr, S. D. (2002). Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body. *Dev. Cell* 3, 283-289.
- Babst, M., Odorizzi, G., Estepa, E. J., and Emr, S. D. (2000). Mammalian Tumor Susceptibility Gene 101 (TSG101) and the Yeast Homologue, Vps23p, Both Function in Late Endosomal Trafficking. *Traffic* 1, 248-258.
- Bailly, V., Lauder, S., Prakash, S., and Prakash, L. (1997). Yeast DNA repair proteins Rad6 and Rad18 form a heterodimer that has ubiquitin conjugating, DNA binding, and ATP hydrolytic activities. *J. Biol. Chem.* 272, 23360-23365.

Balter, M. (1996). New hope in HIV disease. *Science* 274, 1988-1989.

Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., *et al.* (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220, 868-871.

Bays, N. W., Wilhovsky, S. K., Goradia, A., Hodgkiss-Harlow, K., and Hampton, R. Y. (2001). HRD4/NPL4 is required for the proteasomal processing of ubiquitinated ER proteins. *Mol. Biol. Cell.* 12, 4114-4128.

Bertolaet, B. L., Clarke, D. J., Wolff, M., Watson, M. H., Henze, M., Divita, G., and Reed, S. I. (2001). UBA domains of DNA damage-inducible proteins interact with ubiquitin. *Nat. Struct. Biol.* 8, 417-422.

Bishop, N., Horman, A., and Woodman, P. (2002). Mammalian class E vps proteins recognize ubiquitin and act in the removal of endosomal protein-ubiquitin conjugates. *J. Cell. Biol.* 157, 91-101

Brodsky, J. L., and McCracken, A. A. (1999). ER protein quality control and proteasome-mediated protein degradation. *Semin. Cell. Dev. Biol.* 10, 507-513.

Bryant, M., and Ratner, L. (1990). Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc. Natl. Acad. Sci. U S A* 87, 523-527.

Buchberger, A. (2002). From UBA to UBX: new words in the ubiquitin vocabulary. *Trends Cell. Biol.* 12, 216-221.

Bukrinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubel, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., and Stevenson, M. (1993). A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. *Nature* 365, 666-669.

Bukrinsky, M. I., Sharova, N., Dempsey, M. P., Stanwick, T. L., Bukrinskaya, A. G., Haggerty, S., and Stevenson, M. (1992). Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. *Proc. Natl. Acad. Sci. U S A* 89, 6580-6584.

Campbell, S., Fisher, R. J., Towler, E. M., Fox, S., Issaq, H. J., Wolfe, T., Phillips, L. R., and Rein, A. (2001). Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proc. Natl. Acad. Sci. U S A* 98, 10875-10879.

Campbell, S., and Rein, A. (1999). In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. *J. Virol.* 73, 2270-2279.

- Chan, D. C., Chutkowski, C. T., and Kim, P. S. (1998). Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Proc. Natl. Acad. Sci. U S A* *95*, 15613-15617.
- Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997). Core structure of gp41 from the HIV envelope glycoprotein. *Cell* *89*, 263-273.
- Chan, D. C., and Kim, P. S. (1998). HIV entry and its inhibition. *Cell* *93*, 681-684.
- Chen, B. K., Rousso, I., Shim, S., and Kim, P. S. (2001a). Efficient assembly of an HIV-1/MLV Gag-chimeric virus in murine cells. *Proc. Natl. Acad. Sci. U S A* *98*, 15239-15244.
- Chen, L., Shinde, U., Ortolan, T. G., and Madura, K. (2001b). Ubiquitin-associated (UBA) domains in Rad23 bind ubiquitin and promote inhibition of multi-ubiquitin chain assembly. *EMBO Rep.* *2*, 933-938.
- Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. *Cell* *79*, 13-21.
- Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997). *Retroviruses* (Plainview, New York, Cold Spring Harbor Press).
- Cullen, B. R., and Malim, M. H. (1991). The HIV-1 Rev protein: prototype of a novel class of eukaryotic post-transcriptional regulators. *Trends Biochem. Sci.* *16*, 346-350.
- Dai, R. M., Chen, E., Longo, D. L., Gorbea, C. M., and Li, C. C. (1998). Involvement of valosin-containing protein, an ATPase Co-purified with IkappaBalpha and 26 S proteasome, in ubiquitin-proteasome-mediated degradation of IkappaBalpha. *J. Biol. Chem.* *273*, 3562-3573.
- Dieckmann, T., Withers-Ward, E. S., Jarosinski, M. A., Liu, C. F., Chen, I. S., and Feigon, J. (1998). Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr. *Nat. Struct. Biol.* *5*, 1042-1047.
- Dubiel, W., and Gordon, C. (1999). Ubiquitin pathway: another link in the polyubiquitin chain? *Curr. Biol.* *9*, R554-557.
- Edbauer, C. A., and Naso, R. B. (1983). Cytoskeleton-associated Pr65gag and retrovirus assembly. *Virology* *130*, 415-426.
- Edbauer, C. A., and Naso, R. B. (1984). Cytoskeleton-associated Pr65gag and assembly of retrovirus temperature-sensitive mutants in chronically infected cells. *Virology* *134*, 389-397.
- Esposito, D., and Craigie, R. (1998). Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.* *17*, 5832-5843.

- Facke, M., Janetzko, A., Shoeman, R. L., and Krausslich, H. G. (1993). A large deletion in the matrix domain of the human immunodeficiency virus gag gene redirects virus particle assembly from the plasma membrane to the endoplasmic reticulum. *J. Virol.* *67*, 4972-4980.
- Farnet, C. M., and Bushman, F. D. (1996). HIV cDNA integration: molecular biology and inhibitor development. *AIDS* *10*, S3-11.
- Feng, S., and Holland, E. C. (1988). HIV-1 tat trans-activation requires the loop sequence within tar. *Nature* *334*, 165-167.
- Ford, M. G., Pearse, B. M., Higgins, M. K., Vallis, Y., Owen, D. J., Gibson, A., Hopkins, C. R., Evans, P. R., and McMahon, H. T. (2001). Simultaneous binding of PtdIns(4,5)P<sub>2</sub> and clathrin by AP180 in the nucleation of clathrin lattices on membranes. *Science* *291*, 1051-1055.
- Franke, E. K., Yuan, H. E., and Luban, J. (1994). Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* *372*, 359-362.
- Frankel, A. D., and Young, J. A. (1998). HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* *67*, 1-25.
- Freed, E. O. (2002). Viral late domains. *J. Virol.* *76*, 4679-4687.
- Fuller, S. D., Wilk, T., Gowen, B. E., Krausslich, H. G., and Vogt, V. M. (1997). Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. *Curr. Biol.* *7*, 729-738.
- Funakoshi, M., Sasaki, T., Nishimoto, T., and Kobayashi, H. (2002). Budding yeast Dsk2p is a polyubiquitin-binding protein that can interact with the proteasome. *Proc. Natl. Acad. Sci. U S A* *99*, 745-750.
- Gagny, B., Wiederkehr, A., Dumoulin, P., Winsor, B., Riezman, H., and Haguenaer-Tsapis, R. (2000). A novel EH domain protein of *saccharomyces cerevisiae*, ede1p, involved in endocytosis. *J. Cell. Sci.* *113*, 3309-3319.
- Gallay, P., Hope, T., Chin, D., and Trono, D. (1997). HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. *Proc. Natl. Acad. Sci. U S A* *94*, 9825-9830.
- Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., and et al. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* *224*, 500-503.
- Garrus, J. E., von Schwedler, U. K., Pornillos, O. W., Morham, S. G., Zavitz, K. H., Wang, H. E., Wettstein, D. A., Stray, K. M., Cote, M., Rich, R. L., et al. (2001). Tsg101

and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* 107, 55-65.

Gelderblom, H. R. (1991). Assembly and morphology of HIV: potential effect of structure on viral function. *AIDS* 5, 617-637.

Gheysen, D., Jacobs, E., de Foresta, F., Thiriart, C., Francotte, M., Thines, D., and De Wilde, M. (1989). Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell* 59, 103-112.

Gottlinger, H. G., Dorfman, T., Sodroski, J. G., and Haseltine, W. A. (1991). Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc. Natl. Acad. Sci. U S A* 88, 3195-3199.

Gottlinger, H. G., Sodroski, J. G., and Haseltine, W. A. (1989). Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U S A* 86, 5781-5785.

Harty, R. N., Brown, M. E., McGettigan, J. P., Wang, G., Jayakar, H. R., Huibregtse, J. M., Whitt, M. A., and Schnell, M. J. (2001). Rhabdoviruses and the cellular ubiquitin-proteasome system: a budding interaction. *J. Virol.* 75, 10623-10629.

Harty, R. N., Brown, M. E., Wang, G., Huibregtse, J., and Hayes, F. P. (2000). A PPxY motif within the VP40 protein of ebola virus interacts physically and functionally with a ubiquitin ligase: implications for filovirus budding. *Proc. Natl. Acad. Sci. U S A* 97, 13871-13876.

Hazuda, D. J., Felock, P., Witmer, M., Wolfe, A., Stillmock, K., Grobler, J. A., Espeseth, A., Gabryelski, L., Schleif, W., Blau, C., and Miller, M. D. (2000). Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* 287, 646-650.

Hershko, A., and Ciechanover, A. (1992). The ubiquitin system for protein degradation. *Annu. Rev. Biochem.* 61, 761-807.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* 67, 425-479.

Hicke, L. (2001). Protein regulation by monoubiquitin. *Nat. Rev. Mol. Cell. Biol.* 2, 195-201.

Hofmann, K., and Bucher, P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem. Sci.* 21, 172-173.

Hofmann, K., and Falquet, L. (2001). A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem. Sci.* 26, 347-350.

Hofmann, R. M., and Pickart, C. M. (1999). Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. *Cell* *96*, 645-653.

Hope, T. J. (1997). Viral RNA export. *Chem. Biol.* *4*, 335-344.

Hsu, M., and Wainberg, M. A. (2000). Interactions between human immunodeficiency virus type 1 reverse transcriptase, tRNA primer, and nucleocapsid protein during reverse transcription. *J. Hum. Virol.* *3*, 16-26.

Huang, M., Maynard, A., Turpin, J. A., Graham, L., Janini, G. M., Covell, D. G., and Rice, W. G. (1998). Anti-HIV agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *J. Med. Chem.* *41*, 1371-1381.

Hwang, S., Tamilarasu, N., Ryan, K., Huq, I., Richter, S., Still, W. C., and Rana, T. M. (1999). Inhibition of gene expression in human cells through small molecule-RNA interactions. *Proc. Natl. Acad. Sci. U S A* *96*, 12997-13002.

Itoh, T., Koshiba, S., Kigawa, T., Kikuchi, A., Yokoyama, S., and Takenawa, T. (2001). Role of the ENTH domain in phosphatidylinositol-4,5-bisphosphate binding and endocytosis. *Science* *291*, 1047-1051.

Jones, K. A., and Peterlin, B. M. (1994). Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* *63*, 717-743.

Katzmann, D. J., Babst, M., and Emr, S. D. (2001). Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell* *106*, 145-155.

Kikonyogo, A., Bouamr, F., Vana, M. L., Xiang, Y., Aiyar, A., Carter, C., and Leis, J. (2001). Proteins related to the Nedd4 family of ubiquitin protein ligases interact with the L domain of Rous sarcoma virus and are required for gag budding from cells. *Proc. Natl. Acad. Sci. U S A* *98*, 11199-11204.

Kilby, J. M., Hopkins, S., Venetta, T. M., DiMassimo, B., Cloud, G. A., Lee, J. Y., Alldredge, L., Hunter, E., Lambert, D., Bolognesi, D., *et al.* (1998). Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.* *4*, 1302-1307.

Kim, J., Shishido, T., Jiang, X., Aderem, A., and McLaughlin, S. (1994). Phosphorylation, high ionic strength, and calmodulin reverse the binding of MARCKS to phospholipid vesicles. *J. Biol. Chem.* *269*, 28214-28219.

Kondo, E., and Gottlinger, H. G. (1996). A conserved LXXLF sequence is the major determinant in p6gag required for the incorporation of human immunodeficiency virus type 1 Vpr. *J. Virol.* *70*, 159-164.

- Lasky, L. A., Nakamura, G., Smith, D. H., Fennie, C., Shimasaki, C., Patzer, E., Berman, P., Gregory, T., and Capon, D. J. (1987). Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. *Cell* 50, 975-985.
- Li, Y., Kane, T., Tipper, C., Spatrack, P., and Jenness, D. D. (1999). Yeast mutants affecting possible quality control of plasma membrane proteins. *Mol. Cell. Biol.* 19, 3588-3599.
- Lindwasser, O. W., and Resh, M. D. (2001). Multimerization of human immunodeficiency virus type 1 Gag promotes its localization to barges, raft-like membrane microdomains. *J. Virol.* 75, 7913-7924.
- Lingappa, J. R., Hill, R. L., Wong, M. L., and Hegde, R. S. (1997). A multistep, ATP-dependent pathway for assembly of human immunodeficiency virus capsids in a cell-free system. *J. Cell. Biol.* 136, 567-581.
- Lu, Y. L., Bennett, R. P., Wills, J. W., Gorelick, R., and Ratner, L. (1995). A leucine triplet repeat sequence (LXX)<sub>4</sub> in p6gag is important for Vpr incorporation into human immunodeficiency virus type 1 particles. *J. Virol.* 69, 6873-6879.
- Madura, K. (2002). The ubiquitin-associated (UBA) domain: on the path from prudence to prurience. *Cell Cycle* 1, 235-244.
- Magee, A. I., and Courtneidge, S. A. (1985). Two classes of fatty acid acylated proteins exist in eukaryotic cells. *EMBO J.* 4, 1137-1144.
- Majoul, I., Sohn, K., Wieland, F. T., Pepperkok, R., Pizza, M., Hillemann, J., and Soling, H. D. (1998). KDEL receptor (Erd2p)-mediated retrograde transport of the cholera toxin A subunit from the Golgi involves COPI, p23, and the COOH terminus of Erd2p. *J. Cell. Biol.* 143, 601-612.
- Mariani, R., Rutter, G., Harris, M. E., Hope, T. J., Krausslich, H. G., and Landau, N. R. (2000). A block to human immunodeficiency virus type 1 assembly in murine cells. *J. Virol.* 74, 3859-3870.
- Martin-Serrano, J., Zang, T., and Bieniasz, P. D. (2001). HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nat. Med.* 7, 1313-1319.
- Meyer, H. H., Shorter, J. G., Seemann, J., Pappin, D., and Warren, G. (2000). A complex of mammalian ufd1 and npl4 links the AAA-ATPase, p97, to ubiquitin and nuclear transport pathways. *EMBO J.* 19, 2181-2192.
- Meyer, H. H., Wang, Y., and Warren, G. (2002). Direct binding of ubiquitin conjugates by the mammalian p97 adaptor complexes, p47 and Ufd1-Npl4. *EMBO J.* 21, 5645-5652.

- Miller, M. D., Bor, Y. C., and Bushman, F. (1995). Target DNA capture by HIV-1 integration complexes. *Curr. Biol.* *5*, 1047-1056.
- Moore, J. P., and Stevenson, M. (2000). New targets for inhibitors of HIV-1 replication. *Nat. Rev. Mol. Cell Biol.* *1*, 40-49.
- Moraes, T. F., Edwards, R. A., McKenna, S., Pastushok, L., Xiao, W., Glover, J. N., and Ellison, M. J. (2001). Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2-hUbc13. *Nat. Struct. Biol.* *8*, 669-673.
- Moulard, M., and Decroly, E. (2000). Maturation of HIV envelope glycoprotein precursors by cellular endoproteases. *Biochim. Biophys. Acta.* *1469*, 121-132.
- Mueller, T. D., and Feigon, J. (2002). Solution structures of UBA domains reveal a conserved hydrophobic surface for protein-protein interactions. *J. Mol. Biol.* *319*, 1243-1255.
- Olson, E. N., Towler, D. A., and Glaser, L. (1985). Specificity of fatty acid acylation of cellular proteins. *J. Biol. Chem.* *260*, 3784-3790.
- Ono, A., and Freed, E. O. (2001). Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proc. Natl. Acad. Sci. U S A* *98*, 13925-13930.
- Ono, A., Huang, M., and Freed, E. O. (1997). Characterization of human immunodeficiency virus type 1 matrix revertants: effects on virus assembly, Gag processing, and Env incorporation into virions. *J. Virol.* *71*, 4409-4418.
- Ott, D. E., Coren, L. V., Chertova, E. N., Gagliardi, T. D., and Schubert, U. (2000a). Ubiquitination of HIV-1 and MuLV Gag. *Virology* *278*, 111-121.
- Ott, D. E., Coren, L. V., Copeland, T. D., Kane, B. P., Johnson, D. G., Sowder, R. C., 2nd, Yoshinaka, Y., Oroszlan, S., Arthur, L. O., and Henderson, L. E. (1998). Ubiquitin is covalently attached to the p6Gag proteins of human immunodeficiency virus type 1 and simian immunodeficiency virus and to the p12Gag protein of Moloney murine leukemia virus. *J. Virol.* *72*, 2962-2968.
- Ott, D. E., Coren, L. V., Johnson, D. G., Kane, B. P., Sowder, R. C., 2nd, Kim, Y. D., Fisher, R. J., Zhou, X. Z., Lu, K. P., and Henderson, L. E. (2000b). Actin-binding cellular proteins inside human immunodeficiency virus type 1. *Virology* *266*, 42-51.
- Ott, D. E., Coren, L. V., Sowder, R. C., 2nd, Adams, J., Nagashima, K., and Schubert, U. (2002). Equine infectious anemia virus and the ubiquitin-proteasome system. *J. Virol.* *76*, 3038-3044.
- Paillart, J. C., and Gottlinger, H. G. (1999). Opposing effects of human immunodeficiency virus type 1 matrix mutations support a myristyl switch model of gag membrane targeting. *J. Virol.* *73*, 2604-2612.

- Pal, R., Reitz, M. S., Jr., Tschachler, E., Gallo, R. C., Sarngadharan, M. G., and Veronese, F. D. (1990). Myristoylation of gag proteins of HIV-1 plays an important role in virus assembly. *AIDS Res. Hum. Retroviruses* 6, 721-730.
- Patnaik, A., Chau, V., and Wills, J. W. (2000). Ubiquitin is part of the retrovirus budding machinery. *Proc. Natl. Acad. Sci. U S A* 97, 13069-13074.
- Paumet, F., Le Mao, J., Martin, S., Galli, T., David, B., Blank, U., and Roa, M. (2000). Soluble NSF attachment protein receptors (SNAREs) in RBL-2H3 mast cells: functional role of syntaxin 4 in exocytosis and identification of a vesicle-associated membrane protein 8-containing secretory compartment. *J. Immunol.* 164, 5850-5857.
- Paxton, W., Connor, R. I., and Landau, N. R. (1993). Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of *gag* and mutational analyses. *J. Virol.* 67, 7229.
- Peitzsch, R. M., and McLaughlin, S. (1993). Binding of acylated peptides and fatty acids to phospholipid vesicles: pertinence to myristoylated proteins. *Biochemistry* 32, 10436-10443.
- Perrin-Tricaud, C., Davoust, J., and Jones, I. M. (1999). Tagging the human immunodeficiency virus gag protein with green fluorescent protein. Minimal evidence for colocalisation with actin. *Virology* 255, 20-25.
- Pickart, C. M. (2000). Ubiquitin in chains. *Trends Biochem. Sci.* 25, 544-548.
- Pickart, C. M. (2001a). Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* 70, 503-533.
- Pickart, C. M. (2001b). Ubiquitin enters the new millennium. *Mol. Cell* 8, 499-504.
- Piguet, V., Schwartz, O., Le Gall, S., and Trono, D. (1999). The downregulation of CD4 and MHC-I by primate lentiviruses: a paradigm for the modulation of cell surface receptors. *Immunol. Rev.* 168, 51-63.
- Polo, S., Sigismund, S., Faretta, M., Guidi, M., Capua, M. R., Bossi, G., Chen, H., De Camilli, P., and Di Fiore, P. P. (2002). A single motif responsible for ubiquitin recognition and monoubiquitination in endocytic proteins. *Nature* 416, 451-455.
- Pornillos, O., Alam, S. L., Davis, D. R., and Sundquist, W. I. (2002). Structure of the Tsg101 UEV domain in complex with the PTAP motif of the HIV-1 p6 protein. *Nat. Struct. Biol.* 9, 812-817.
- Putterman, D., Pepinsky, R. B., and Vogt, V. M. (1990). Ubiquitin in avian leukosis virus particles. *Virology* 176, 633-637.
- Qian-Cutrone, J., Huang, S., Chang, L. P., Pirnik, D. M., Klohr, S. E., Dalterio, R. A., Hugill, R., Lowe, S., Alam, M., and Kadow, K. F. (1996). Harziphilone and fleephilone,

two new HIV REV/RRE binding inhibitors produced by *Trichoderma harzianum*. *J. Antibiot. (Tokyo)* *49*, 990-997.

Raiborg, C., Bache, K. G., Gillooly, D. J., Madhus, I. H., Stang, E., and Stenmark, H. (2002). Hrs sorts ubiquitinated proteins into clathrin-coated microdomains of early endosomes. *Nat. Cell Biol.* *4*, 394-398.

Rao, H., and Sastry, A. (2002). Recognition of specific ubiquitin conjugates is important for the proteolytic functions of the ubiquitin-associated domain proteins Dsk2 and Rad23. *J. Biol. Chem.* *277*, 11691-11695.

Ratmeyer, L., Zapp, M. L., Green, M. R., Vinayak, R., Kumar, A., Boykin, D. W., and Wilson, W. D. (1996). Inhibition of HIV-1 Rev-RRE interaction by diphenylfuran derivatives. *Biochemistry* *35*, 13689-13696.

Rice, W. G., Turpin, J. A., Huang, M., Clanton, D., Buckheit, R. W., Jr., Covell, D. G., Wallqvist, A., McDonnell, N. B., DeGuzman, R. N., Summers, M. F., *et al.* (1997). Azodicarbonamide inhibits HIV-1 replication by targeting the nucleocapsid protein. *Nat. Med.* *3*, 341-345.

Richman, D. D. (2001). HIV chemotherapy. *Nature* *410*, 995-1001.

Rittner, K., Churcher, M. J., Gait, M. J., and Karn, J. (1995). The human immunodeficiency virus long terminal repeat includes a specialised initiator element which is required for Tat-responsive transcription. *J. Mol. Biol.* *248*, 562-580.

Rizzuto, C. D., Wyatt, R., Hernandez-Ramos, N., Sun, Y., Kwong, P. D., Hendrickson, W. A., and Sodroski, J. (1998). A conserved HIV gp120 glycoprotein structure involved in chemokine receptor binding. *Science* *280*, 1949-1953.

Root, M. J., Kay, M. S., and Kim, P. S. (2001). Protein design of an HIV-1 entry inhibitor. *Science* *291*, 884-888.

Royer, M., Cerutti, M., Gay, B., Hong, S. S., Devauchelle, G., and Boulanger, P. (1991). Functional domains of HIV-1 gag-polyprotein expressed in baculovirus-infected cells. *Virology* *184*, 417-422.

Royer, M., Hong, S. S., Gay, B., Cerutti, M., and Boulanger, P. (1992). Expression and extracellular release of human immunodeficiency virus type 1 Gag precursors by recombinant baculovirus-infected cells. *J. Virol.* *66*, 3230-3235.

Sakalian, M., and Hunter, E. (1998). Molecular events in the assembly of retrovirus particles. *Adv. Exp. Med. Biol.* *440*, 329-339.

Sarnagadharan, M. G., DeVico, A. L., Bruch, L., Schupbach, J., and Gallo, R. C. (1984). HTLV-III: the etiologic agent of AIDS. *Princess Takamatsu Symp.* *15*, 301-308.

- Schubert, U., Ott, D. E., Chertova, E. N., Welker, R., Tessmer, U., Princiotta, M. F., Bennink, J. R., Krausslich, H. G., and Yewdell, J. W. (2000). Proteasome inhibition interferes with gag polyprotein processing, release, and maturation of HIV-1 and HIV-2. *Proc. Natl. Acad. Sci. U S A* *97*, 13057-13062.
- Shih, S. C., Katzmann, D. J., Schnell, J. D., Sutanto, M., Emr, S. D., and Hicke, L. (2002). Epsins and Vps27p/Hrs contain ubiquitin-binding domains that function in receptor endocytosis. *Nat. Cell Biol.* *4*, 389-393.
- Shioda, T., and Shibuta, H. (1990). Production of human immunodeficiency virus (HIV)-like particles from cells infected with recombinant vaccinia viruses carrying the gag gene of HIV. *Virology* *175*, 139-148.
- Sigal, C. T., Zhou, W., Buser, C. A., McLaughlin, S., and Resh, M. D. (1994). Amino-terminal basic residues of Src mediate membrane binding through electrostatic interaction with acidic phospholipids. *Proc. Natl. Acad. Sci. U S A* *91*, 12253-12257.
- Silvius, J. R., and l'Heureux, F. (1994). Fluorimetric evaluation of the affinities of isoprenylated peptides for lipid bilayers. *Biochemistry* *33*, 3014-3022.
- Spearman, P., Horton, R., Ratner, L., and Kuli-Zade, I. (1997). Membrane binding of human immunodeficiency virus type 1 matrix protein in vivo supports a conformational myristyl switch mechanism. *J. Virol.* *71*, 6582-6592.
- Spence, J., Gali, R. R., Dittmar, G., Sherman, F., Karin, M., and Finley, D. (2000). Cell cycle-regulated modification of the ribosome by a variant multiubiquitin chain. *Cell* *102*, 67-76.
- Strack, B., Calistri, A., Accola, M. A., Palu, G., and Gottlinger, H. G. (2000). A role for ubiquitin ligase recruitment in retrovirus release. *Proc. Natl. Acad. Sci. U S A* *97*, 13063-13068.
- Subbramanian, R. A., Yao, X. J., Dilhuydy, H., Rougeau, N., Bergeron, D., Robitaille, Y., and Cohen, E. A. (1998). Human immunodeficiency virus type 1 Vpr localization: nuclear transport of a viral protein modulated by a putative amphipathic helical structure and its relevance to biological activity. *J. Mol. Biol.* *278*, 13-30.
- Sullivan, N., Sun, Y., Sattentau, Q., Thali, M., Wu, D., Denisova, G., Gershoni, J., Robinson, J., Moore, J., and Sodroski, J. (1998). CD4-Induced conformational changes in the human immunodeficiency virus type 1 gp120 glycoprotein: consequences for virus entry and neutralization. *J. Virol.* *72*, 4694-4703.
- Tamilarasu, N., Huq, I., and Rana, T. M. (2000). Design, synthesis, and biological activity of a cyclic peptide: an inhibitor of HIV-1 tat-TAR interactions in human cells. *Bioorg. Med. Chem. Lett.* *10*, 971-974.
- Trkola, A., Dragic, T., Arthos, J., Binley, J. M., Olson, W. C., Allaway, G. P., Cheng-Mayer, C., Robinson, J., Maddon, P. J., and Moore, J. P. (1996). CD4-dependent,

antibody-sensitive interactions between HIV-1 and its co-receptor CCR-5. *Nature* 384, 184-187.

Ulrich, H. D., and Jentsch, S. (2000). Two RING finger proteins mediate cooperation between ubiquitin-conjugating enzymes in DNA repair. *EMBO J.* 19, 3388-3397.

UNAIDS (2001). United Nations Program on HIV/AIDS. AIDS epidemic update: December, 2001.

VanDemark, A. P., Hofmann, R. M., Tsui, C., Pickart, C. M., and Wolberger, C. (2001). Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell* 105, 711-720.

Verma, R., Chen, S., Feldman, R., Schieltz, D., Yates, J., Dohmen, J., and Deshaies, R. J. (2000). Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol. Biol. Cell* 11, 3425-3439.

VerPlank, L., Bouamr, F., LaGrassa, T. J., Agresta, B., Kikonyogo, A., Leis, J., and Carter, C. A. (2001). Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55Gag. *Proc. Natl. Acad. Sci. U S A* 98, 7724-7729.

Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* 194, 531-544.

Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., and Cook, W. J. (1985). Three-dimensional structure of ubiquitin at 2.8 Å resolution. *Proc. Natl. Acad. Sci. U S A* 82, 3582-3585.

Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997). Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 387, 426-430.

Wilk, T., Gross, I., Gowen, B. E., Rutten, T., de Haas, F., Welker, R., Krausslich, H. G., Boulanger, P., and Fuller, S. D. (2001). Organization of immature human immunodeficiency virus type 1. *J. Virol.* 75, 759-771.

Wilkinson, C. R., Seeger, M., Hartmann-Petersen, R., Stone, M., Wallace, M., Semple, C., and Gordon, C. (2001). Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. *Nat. Cell. Biol.* 3, 939-943.

Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M., and Kingsman, A. J. (1988). HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. *Cell* 55, 1159-1169.

Withers-Ward, E. S., Jowett, J. B., Stewart, S. A., Xie, Y. M., Garfinkel, A., Shibagaki, Y., Chow, S. A., Shah, N., Hanaoka, F., Sawitz, D. G., *et al.* (1997). Human immunodeficiency virus type 1 Vpr interacts with HHR23A, a cellular protein implicated in nucleotide excision DNA repair. *J. Virol.* 71, 9732-9742.

Wu, L., Gerard, N. P., Wyatt, R., Choe, H., Parolin, C., Ruffing, N., Borsetti, A., Cardoso, A. A., Desjardin, E., Newman, W., *et al.* (1996). CD4-induced interaction of primary HIV-1 gp120 glycoproteins with the chemokine receptor CCR-5. *Nature* 384, 179-183.

Young, P., Deveraux, Q., Beal, R. E., Pickart, C. M., and Rechsteiner, M. (1998). Characterization of two polyubiquitin binding sites in the 26 S protease subunit 5a. *J. Biol. Chem.* 273, 5461-5467.

Yu, X. F., Matsuda, Z., Yu, Q. C., Lee, T. H., and Essex, M. (1995). Role of the C terminus Gag protein in human immunodeficiency virus type 1 virion assembly and maturation. *J. Gen. Virol.* 76, 3171-3179.

Yuan, X., Yu, X., Lee, T. H., and Essex, M. (1993). Mutations in the N-terminal region of human immunodeficiency virus type 1 matrix protein block intracellular transport of the Gag precursor. *J. Virol.* 67, 6387-6394.

Zapp, M. L., and Green, M. R. (1989). Sequence-specific RNA binding by the HIV-1 Rev protein. *Nature* 342, 714-716.

Zhou, M., Halanski, M. A., Radonovich, M. F., Kashanchi, F., Peng, J., Price, D. H., and Brady, J. N. (2000). Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription. *Mol. Cell. Biol.* 20, 5077-5086.

Zhou, Q., and Sharp, P. A. (1995). Novel mechanism and factor for regulation by HIV-1 Tat. *EMBO J.* 14, 321-328.

Zhou, W., Parent, L. J., Wills, J. W., and Resh, M. D. (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J. Virol.* 68, 2556-2569.

Zimmerman, C., Klein, K. C., Kiser, P. K., Singh, A. R., Firestein, B. L., Riba, S. C., and Lingappa, J. R. (2002). Identification of a host protein essential for assembly of immature HIV-1 capsids. *Nature* 415, 88-92.

## CHAPTER 2

### HIV-1 MA MEMBRANE INTERACTION

## Introduction

Gag is the major structural protein of the Human Immunodeficiency Virus Type I (HIV-1) (Freed, 1998). Late in the infectious cycle, Gag is synthesized in the cytoplasm as a polyprotein precursor and is co-translationally N-myristoylated. Newly synthesized Gag molecules are transported to the plasma membrane where they assemble into immature virions that bud from the cell. In addition to its structural roles, Gag also recruits other viral components to the assembling virion, including the genomic RNA, the Gag-Pol fusion protein, the Env glycoproteins (Dorfman et al., 1994; Freed and Martin, 1996; Ono et al., 1997), viral protein R (Vpr) (Kondo and Gottlinger, 1996; Lu et al., 1995; Paxton et al., 1993), and the cellular proteins Cyclophilin A and Tsg101 (Franke et al., 1994; Garrus et al., 2001; Thali et al., 1994; VerPlank et al., 2001). As the virus buds, the Gag polyprotein precursor is proteolytically cleaved by viral protease (PR) to generate a series of smaller proteins: MA, CA, NC and p6, which then rearrange to form the mature infectious virus.

The plasma membrane targeting function of Gag protein is essential for viral assembly, and the N-terminal MA domain is largely responsible for membrane targeting (Yuan et al., 1993; Zhou et al., 1994). For example, mutations in MA domain can mislocalize Gag to intracellular membrane compartments, including the endoplasmic reticulum (ER) and Golgi (Facke et al., 1993; Ono et al., 2000; Reil et al., 1998). Moreover, when Gag is expressed in murine cells, the protein fails to accumulate at the plasma membrane, suggesting that Gag transport may be mediated by cellular factor(s) that binds the MA domain (Mariani et al., 2000). Consistent with this model, replacing the MA domain of HIV-1 Gag with the MA domain of Moloney murine leukemia virus

(Mo-MuLV) Gag restored plasma membrane targeting and efficient viral assembly (Chen et al., 2001).

Recent studies have also indicated that “lipid rafts” in the plasma membrane have a critical role in HIV-1 virus assembly (Nguyen and Hildreth, 2000; Ono and Freed, 2001). Lipid rafts are microdomains in the plasma membrane that are highly enriched in sphingolipid and cholesterol and are resistant to non-ionic detergent treatment (Anderson and Jacobson, 2002; Brown and London, 2000; Simons and Toomre, 2000). HIV-1 Gag has been shown to associate specifically with these cholesterol-enriched lipid rafts. As expected, raft binding is mediated by the N-terminus of Gag (Nguyen and Hildreth, 2000; Ono and Freed, 2001). Disruption of the lipid rafts by cholesterol depletion significantly reduced the HIV-1 particle production, indicating that this interaction is likely to be functionally important (Ono and Freed, 2001).

Although the processes of Gag transport and assembly are not yet well understood, several cellular factors have been implicated in these processes. HIV-1 Gag and other retroviral Gag proteins have been shown to associate with the cytoskeletal system, suggesting that Gag may be transported to the plasma membrane through interactions with the cytoskeletal system (Edbauer and Naso, 1983; Edbauer and Naso, 1984; Perrin-Tricaud et al., 1999). More recently, the ATP-binding protein, Hp68, has been reported as a Gag binding protein and shown to be important for the assembly of immature viral capsid in a cell-free system (Lingappa et al., 1997; Zimmerman et al., 2002), suggesting that Hp68 may function as a molecular chaperone during Gag assembly. Finally, phosphatidyl inositol phosphates (PtdInsP) have been identified as cellular factors that can modulate Gag assembly in vitro (Campbell et al., 2001; Campbell

and Rein, 1999), and the possible roles of PtdInsP in HIV-1 Gag assembly are discussed below.

#### Gag/MA myristoylation

HIV-1 Gag protein is cotranslationally modified by myristic acid on its N-terminus (Veronese et al., 1988). Myristic acid is a 14-carbon, saturated fatty acid that is covalently attached to the N-terminal glycine residues of substrate proteins via an amide linkage by the enzyme N-myristoyl transferase (Deichaite et al., 1988; Towler et al., 1988). N-myristoylation is required for Gag membrane targeting, and mutations of the N-terminal glycine of Gag that prevent myristoylation also block viral assembly and replication (Bryant and Ratner, 1990; Gottlinger et al., 1989; Pal et al., 1990). In analogy to other systems, it is assumed that the myristoyl group of Gag inserts into the lipid bilayer and provides hydrophobic interactions that help anchor the protein on the plasma membrane. However, unlike the longer acyl chains of palmitic or stearic acid, the energy provided by inserting the 14-carbon myristoyl group into the lipid bilayer is barely sufficient to attach a protein to the membrane surface (Kim et al., 1994b; Peitzsch and McLaughlin, 1993; Sigal et al., 1994; Silvius and l'Heureux, 1994). Therefore, myristoylated proteins typically require additional interactions to bind membrane stably. Some proteins, such as p56<sup>lck</sup> - a member of the Src family, use a second hydrophobic modification to stabilize membrane binding. Thus, in addition to the N-terminal myristoylation, p56<sup>lck</sup> is also palmitoylated via thio-ester bond at a cysteine residue near the protein's N-terminus (Paige et al., 1993). The palmitoyl group significantly enhances the protein's membrane binding affinity. Other proteins, such as the myristoylated alanine-rich C-kinase substrate (MARCKS), use electrostatic interactions. In addition to

the N-terminal myristoylated domain that mediates membrane binding, MARCKS also has a basic effector domain containing PKC phosphorylation sites as well as calmodulin and actin binding sites. Peptides corresponding to the basic region in this domain bind strongly to membranes containing acidic lipids (Kim et al., 1994a; Taniguchi and Manenti, 1993). Moreover, phosphorylation of this basic domain, which weakens the electrostatic interaction, causes MARCKS to leave phospholipid vesicles containing acidic lipids (Kim et al., 1994b). In the case of HIV-1 MA, a highly basic region near the N-terminus of MA also contributes to membrane binding (Hill et al., 1996; Yuan et al., 1993; Zhou et al., 1994). This basic region likely provides the additional energy required for stable membrane binding through electrostatic interactions with the negatively charged phospholipid head groups on the membrane surface.

A “bipartite” membrane binding model in which both the hydrophobic interaction of the myristoyl group and electrostatic interaction of the basic residues contribute to membrane binding agrees well with structural models for MA membrane binding (Hill et al., 1996). The structure of unmyristoylated HIV-1 MA protein has been determined by both NMR and X-ray crystallography (Hill et al., 1996; Massiah et al., 1994; Matthews et al., 1994), and the structures are in good agreement. MA contains five  $\alpha$ -helices and a three stranded mixed  $\beta$ -sheet. Helices I, II, III, and V pack against the long central helix IV to form a single globular domain. Helix V projects away from the packed helix bundle. Basic residues near the N-terminus of MA required for membrane binding are localized to the N-terminal tail, the first two  $\beta$ -strands and the exposed face of helix II. Although MA is monomeric in solution even at high concentrations, both HIV-1 and SIV MA formed similar trimers in the crystal. A model for MA membrane binding has been

proposed based on the trimer structures (Hill et al., 1996; Rao et al., 1995). In this model, the appropriate basic residues come together to form a putative membrane binding surface across the top of the trimer, and helix V points toward the center of the virus. The basic residues could therefore make electrostatic interactions with the negatively charged lipid membrane surface, and three N-myristoyl groups are also close enough to insert into the lipid membrane and stabilize trimer membrane binding.

About 100 proteins are known to be myristoylated. Fractionation studies in yeast have shown that many myristoylated proteins bind to membranes, including the plasma and ER membranes, but also that as many as 50% exist in the cytosol as soluble proteins (Magee and Courtneidge, 1985; Olson et al., 1985). Moreover, at least some of the myristoylated proteins exhibit multiple chemical or conformational states that help regulate membrane binding. For example, the membrane binding of GAP-43, a guanine nucleotide exchange protein, is regulated by reversible palmitoylation (Sudo et al., 1992). In other cases, reversible membrane binding is regulated by what have come to be called “myristoyl switches.” In general, myristoyl switch proteins adopt two distinct conformations: a soluble conformation in which the hydrophobic myristoyl group is sequestered within the protein, and a membrane-bound conformation in which the myristoyl group extrudes and inserts into the lipid bilayer.

Three general classes of myristoyl switches have now been identified in various proteins (Ames et al., 1996). The first class is controlled by  $\text{Ca}^{2+}$  binding and includes recoverin and its family members, GCAP (guanylate cyclase activator proteins) and YCP (yeast  $\text{Ca}^{2+}$  binding protein) (Dizhoor et al., 1995; Subbaraya et al., 1994). The second class is controlled by GTP/GDP exchange and includes ARFs (ADP-ribosylation factor)

and transducin ( $\alpha$ -subunit). The third class of myristoyl switch is controlled by phosphorylation and includes MARCKS (myristoylated alanine-rich C kinase substrate), Src and possibly also the HIV-1 MA protein (McLaughlin and Aderem, 1995; Resh, 1994; Zhou and Resh, 1996). The structural mechanisms of the first two classes of myristoyl switch are now understood in molecular detail, because both conformations of recoverin and ARF have been determined (Ames et al., 1997; Ames et al., 1996; Goldberg, 1998). However, the structural basis for the third class of myristoyl switches remains to be determined.

Recoverin is a  $\text{Ca}^{2+}$  sensor in retinal rod cells, and  $\text{Ca}^{2+}$  binding promotes recoverin membrane binding (Zozulya and Stryer, 1992). The structure of recoverin is composed of four copies of the small “EF hand” calcium binding motif (Flaherty et al., 1993). In the  $\text{Ca}^{2+}$  free form, the myristoyl group is buried in a hydrophobic pocket formed by helices from three of the different EF hands. In the presence of  $\text{Ca}^{2+}$ , two EF hands in the N-terminal domain each bind one calcium ion and rotate 45 degrees relative to the two EF hands in the C-terminal domain, leading to the extrusion of the myristoyl group (Ames et al., 1997; Ames et al., 1996).

ARF functions in Golgi vesicular trafficking (Rothman and Wieland, 1996). GTP binding activates ARF and causes it to translocate from the cytosol to the membrane surface. Once at the membrane, ARF recruits the coatamer subunit  $\beta$ -COP, leading to the formation of coatamer-coated transport vesicles (Zhao et al., 1997). ARF has a G-protein fold, which includes a 7-stranded mixed  $\beta$ -sheet surrounded by five  $\alpha$ -helices (Amor et al., 1994). In the GDP-bound structure, hydrophobic residues Tyr58, Asn60, Ile61, and Phe63 from loop  $\lambda$ 3 between  $\beta$ -strand  $\beta$ 2 and  $\beta$ 3 form a binding site for the myristoylated

N-terminal helix  $\alpha 1$ . GTP binding causes a large shift of  $\beta$ -strand 2 and 3 and eliminates this binding site. Thus, the myristoyl group is exposed upon GTP binding and becomes available for membrane binding (Goldberg et al., 1996).

One final example of a structurally characterized protein that may use a myristoyl switch mechanism is the reovirus membrane-penetration protein,  $\mu 1$ . The structure of the activated viral shell suggests that exposure of the myristoyl group is induced by the autolytic cleavage of  $\mu 1$ . Exposure of the myristoyl group and subsequent molecular rearrangements are postulated to be important steps in the viral penetration mechanism (Liemann et al., 2002).

The myristoylated HIV-1 Gag protein also has at least two different membrane binding states: it is a soluble protein when it is translated in the cytosol, and it then binds to the plasma membrane to initiate viral assembly. The different membrane binding states of Gag have led several groups to suggest that its membrane binding properties are regulated, possibly by a myristoyl switch in its N-terminal MA domain (Freed, 1998). These models suggest that the globular domain of MA may control the exposure of the N-terminal myristoyl group and thereby control the membrane binding of MA and Gag. In support of these models, extensive deletions that remove most or all of the entire globular core of MA enhance Gag membrane binding and viral particle formation, suggesting that the N-myristoyl group may be exposed in the absence of MA global core (Reil et al., 1998; Spearman et al., 1997). The membrane binding of MA can also be increased by deletion of the final helix (Zhou and Resh, 1996). Moreover, MA mutations that appear to alter the equilibrium between the myristoyl “in” and “out” conformations have been identified. For example, point mutations and a large deletion in MA have been

shown to redirect viral assembly to intracellular membrane compartments (Facke et al., 1993; Freed et al., 1994), suggesting that the myristoyl group may be constitutively exposed in these mutants and the membrane binding can no longer be regulated. The single mutation (L8A) in the N-terminal helix I of MA dramatically impairs Gag membrane binding without affecting N-myristoylation, possibly by keeping the myristoyl group sequestered (Paillart and Gottlinger, 1999). Interestingly, several second-site mutations in the  $\alpha$ -helical core of MA can suppress this phenotype and enhance Gag membrane binding (Paillart and Gottlinger, 1999). However, the detailed mechanism of the myristoyl switch is not yet known, nor is it clear what controls the switching mechanism. It has been shown that phorbol ester treatment can both phosphorylate MA and induce its rapid translocation to membrane (Yu et al., 1995), suggesting that the myristoyl switch may be controlled by phosphorylation. However, other mechanisms, including PtdInsP binding, could also be the control mechanisms of this myristoyl switch. Detailed understanding of this myristoyl switch will likely come from structural studies of the myristoylated HIV-1 MA protein, and this is a major focus of this research described herein.

#### Phosphatidylinositol Phosphates and HIV-1 Gag assembly

Like other retroviral Gag proteins, HIV-1 Gag appears to encode all the information required for viral particle formation, as expression of Gag protein alone in mammalian or insect cells is sufficient to produce virus like particles (VLPs) (Gheysen et al., 1989; Royer et al., 1991; Royer et al., 1992; Shioda and Shibuta, 1990). Although these VLPs lack other viral components, such as genomic RNA and envelope proteins, they are otherwise structurally analogous to the immature viral particles, and have similar

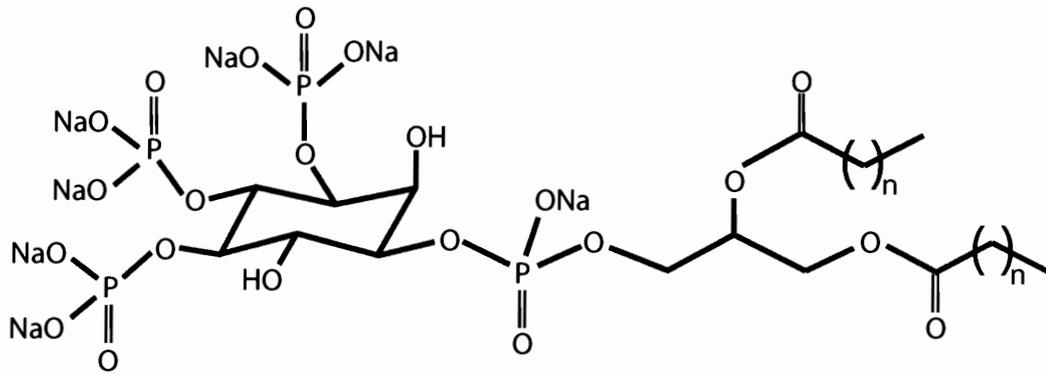
morphologies and sizes (Fuller et al., 1997; Wilk et al., 2001). Although the arrangement of Gag proteins in VLPs and immature viral particles lacks strict icosahedral symmetry, the Gag subunits are locally ordered. Radial density profile analyses have shown that Gag molecules span about 270Å and are composed of separately folded MA, CA, and NC/p6 domains. MA is responsible for Gag membrane binding, and CA and NC domains contain the Gag-Gag interaction sites required for viral particle assembly. The Gag proteins are arranged radially in both VLPs and immature viral particles with the N-terminal MA domain immediately beneath the membrane and the C-terminal NC domain projecting toward the center of the particles.

In vitro assembly studies of retroviral Gag proteins and fragments have shown that pure recombinant proteins are capable of the molecular interactions required for viral particle assembly. Bacterially expressed and purified Rous sarcoma virus (RSV), Mason-Pfizer Monkey virus (MPMV), and MoMuLV Gag proteins can assemble into VLPs that resemble authentic immature viral particles (Campbell and Vogt, 1997; Klikova et al., 1995). In the case of HIV-1, the purified HIV-1 CA protein can assemble both rod-shaped and tubular structures in vitro (Campbell and Rein, 1999; Ehrlich et al., 1992; Gross et al., 1997; Li et al., 2000). Spherical particles of heterogeneous sizes are typically formed by constructs that contain N-terminal MA extensions before CA (Gross et al., 1998; von Schwedler et al., 1998). C-terminal extensions beyond CA that include NC or NC-p6 frequently form tubular structures. Moreover, constructs that contain the NC domain assemble at lower protein concentrations if RNA is provided in the assembly reaction mixture (Campbell and Vogt, 1997; Gross et al., 1997).

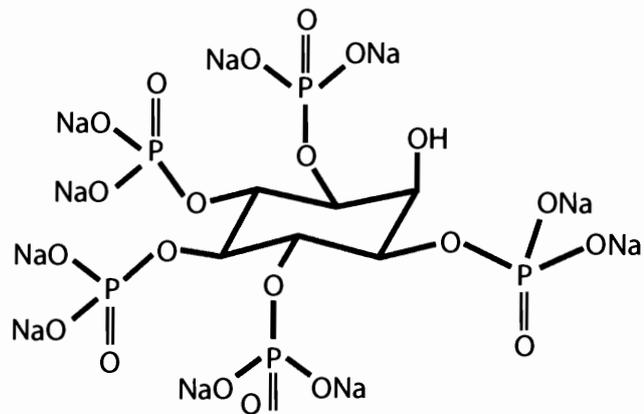
An important difference between VLPs formed by HIV-1 Gag and other retroviral Gag proteins is that the HIV-1 Gag VLPs differ from authentic HIV-1 virions in several respects: 1) they are much smaller, with diameters about 25-30 nm, in contrast to the ~100 nm diameter of authentic HIV-1 virions, 2) they are sensitive to high salt (0.5 M NaCl) or RNase treatment, and 3) they are sensitive to proteolysis (Campbell et al., 2001). These defects can be eliminated by the addition of mammalian cell lysates to the *in vitro* assembly system, suggesting that cellular factor(s) are involved in viral assembly.

Extensive fractionation of rabbit reticulocyte lysates has shown that the “active factors” required for accurate *in vitro* assembly of HIV-1 Gag are inositol phosphates (IPs) and phosphatidylinositol phosphates (PtdInsPs) (Campbell et al., 2001). The most active IPs and PtdInsPs include inositol-1,3,4,5,6-pentakisphosphate (IP<sub>5</sub>) and phosphatidylinositol-3,4,5-triphosphates (PtdIns(3,4,5)P<sub>3</sub>). The molecular structures of these compounds are illustrated in Figure 2.1. Addition of pure IPs or PtdInsPs into *in vitro* assembly reactions of HIV-1 Gag results in larger VLPs which are resistant to high salt, RNase treatment and trypsin digestion. A region in Gag required for these effects has been mapped by deletion analysis to MA, between residues 16 and 99 (Campbell et al., 2001). Intriguingly, HIV-1 Gag proteins lacking residues 16 to 99 assemble into VLPs of correct size (which are also resistant to high salt and RNase treatment) *without* the addition of IPs or PtdInsPs. These results suggest that IPs or PtdInsPs may bind to the globular MA domain of Gag and overcome some inhibitory signal that would otherwise result in aberrant particle assembly. Although the detailed mechanism of this modulation is not yet known, inositol derivatives have been shown to help organize other large macromolecular complexes in cells. For example, adaptor protein AP180 binds to

Figure 2.1. Molecular structures of phosphatidylinositol-3,4,5-triphosphates and inositol-1,3,4,5,6-pentakisphosphate.



phosphatidylinositol-3,4,5-triphosphate



inositol-1,3,4,5,6-pentakisphosphate

phosphatidylinositol-4,5-phosphates (PtdIns(4,5)P<sub>2</sub>) through a lysine-rich motif on its N-terminal domain, and this recruits AP180 to the membrane and initiates the assembly process of clathrin coated vesicles (Ford et al., 2001; Itoh et al., 2001). PtdInsPs could, in principle, perform a similar function in HIV-1 Gag assembly. Consistent with this idea, the HIV-1  $\Delta$ 16-99 Gag construct is also incorrectly targeted to intracellular membranes when expressed in cells (Facke et al., 1993), suggesting that IPs and PtdInsPs may play an important role in targeting Gag to their assembly sites on the plasma membrane. Studies of the interaction between HIV-1 MA and IPs / PtdInsPs are therefore expected to increase our understanding of the mechanism of Gag assembly, and this is another major focus of the research described herein.

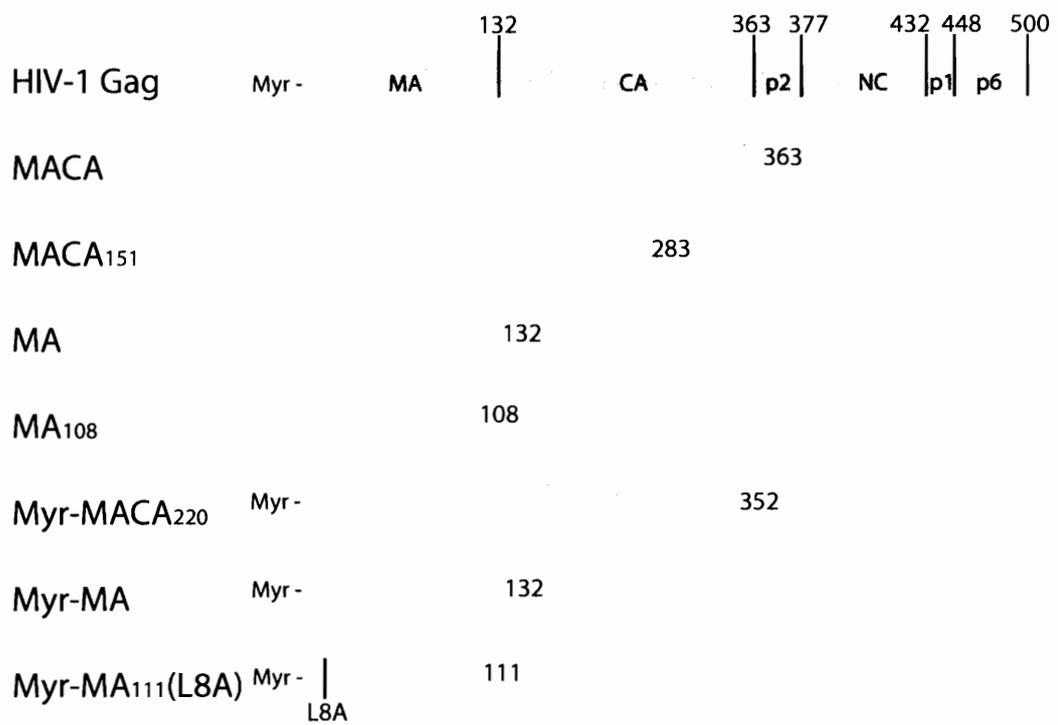
## **Materials and Methods**

### Expression and purification of recombinant HIV-1 MA and MA-CA proteins

The MA and MA-CA constructs used in this study are shown in Figure 2.2. We have previously described the cloning, expression, and purification of unmyristoylated HIV-1<sub>NL4-3</sub> MA protein (Massiah et al., 1994). Analogous procedures were used to express and purify unmyristoylated MA<sub>108</sub>, MA, MACA<sub>151</sub>, and MACA. All recombinant proteins were confirmed by mass spectrometry (not shown).

To make the recombinant, myristoylated HIV-1 MA proteins described in this study (Myr-MA, Myr-MA<sub>111</sub>(L8A), and Myr-MACA<sub>220</sub>), the proteins were co-expressed in *E. coli* with yeast N-myristoyltransferase. This dual plasmid expression system reconstitutes the eukaryotic protein modification and was originally developed by Gordon and coworkers for myristoylation of other recombinant proteins (Duronio et al., 1990).

Figure 2.2. HIV-1 Gag domain organization and a summary of the MA and MACA protein constructs used in this study.



A full description of the expression and purification protocol is given for the Myr-MA<sub>111</sub>(L8A) protein, and the other myristoylated HIV-1 Gag proteins were expressed and purified using analogous methods.

The gene encoding MA<sub>111</sub>(L8A) was amplified by PCR and subcloned into the NdeI/BamHI sites of the pET11a expression vector (Novagen). Primers were designed to introduce the L8A mutation, a His<sub>10</sub> tag on the C-terminal of MA<sub>111</sub>(L8A) with a Factor Xa cleavage site, and the appropriate cloning sites. This plasmid (WISP0041) was cotransformed into *E. coli* BL21(DE3) together with the pBB131 plasmid (which expresses yeast N-myristoyltransferase), and the cells were grown under selection for both ampicillin and kanamycin resistance. Expression of both MA<sub>111</sub>(L8A) and the yeast N-myristoyltransferase was induced by addition of IPTG (0.5 mM) when the culture reached OD<sub>600</sub> of 0.5-0.6. One hundred mg/L myristic acid was also added into the culture at the time of induction. Cells were harvested 4 hours after induction.

Myr-MA<sub>111</sub>(L8A) was purified as described below, with all protein purification steps performed at 4°C except where specifically noted. The cell pellet from a 6 L culture was resuspended in 45 mL lysis buffer (20 mM Tris pH 7.4, 50 mM imidazole) with 1 tablet of protease inhibitor cocktail (Roche Diagnostics GmbH), and lysed by two passes through a French press at 20,000 PSI and 4×30 seconds sonication with duty cycle setting at 50% and output setting at 5. The cell lysate was centrifuged for 1 hour at 39,000×g to remove insoluble debris, and 5 M NaCl added to a final concentration of 500 mM. The Myr-MA<sub>111</sub>(L8A) protein was then affinity purified on a Ni<sup>2+</sup> affinity column (Pharmacia) using a linear gradient from 50 mM to 500 mM imidazole in 20 mM Tris pH 7.4, 500 mM NaCl. Myr-MA<sub>111</sub>(L8A) eluted at ~500 mM imidazole. The purified protein

fractions were pooled and dialyzed against 20 mM Tris pH 7.4, 300 mM NaCl, 5 mM EDTA, and then concentrated to ~15 mL in an Amicon concentrator. The salt was lowered by dialysis against 20 mM Tris pH7.4, 200 mM NaCl, 2 mM Ca<sup>2+</sup>, 5 mM βME, and then against 20 mM Tris pH 7.4, 100 mM NaCl, 2 mM Ca<sup>2+</sup>, 5 mM βME. 100 μg Factor Xa (New England BioLabs) was then added to remove the His<sub>10</sub> affinity tag (16 hours at 23°C). After removal of the affinity tag, the protein was dialyzed against 25 mM MOPS pH 6.8, 50 mM NaCl, 5 mM βME and purified to homogeneity by a S Sepharose chromatography (Pharmacia) using a linear gradient of 50 mM to 1 M NaCl with 25 mM MOPS pH 6.8 and 5 mM βME. Myr-MA<sub>111</sub>(L8A) eluted at ~700 mM NaCl. Typical yields were about 1 mg pure protein per liter of culture. The purity and full myristoylation of purified proteins were confirmed by SDS-PAGE and electrospray mass spectrometry (M.W.<sub>cal</sub> = 13103 g/mol and M.W.<sub>obs</sub> = 13104 g/mol).

Myr-MA and Myr-MACA<sub>220</sub> were cloned, expressed, and purified using similar procedures, except that the plasmid (WISP0066) used for Myr-MACA<sub>220</sub> expression was pET32a- based and had a C-terminal His<sub>6</sub> tag with an enterokinase cleavage site.

Isotopically labeled proteins for NMR studies were expressed and purified from *E. coli* grown in M9 minimal media with <sup>15</sup>NH<sub>4</sub>Cl or <sup>15</sup>NH<sub>4</sub>Cl / <sup>13</sup>C<sub>6</sub>-glucose as the sole nitrogen and carbon sources, respectively.

### Biosensor quantitation of the interaction between

#### HIV-1 MA proteins and liposomes

Lipids consisting of phosphatidylcholine (PC; Avanti #840053) or a 2:1 mixture of PC and phosphatidylserine (PS; Avanti #840032) were suspended in 20 mM *bis*-Tris pH 7.0, 100 mM NaCl, 5 mM βME, put through five freeze-thaw cycles, and extruded 17

times through two 50 nm polycarbonate filter membranes using the mini-extruder kit from Avanti. Binding experiments were performed using a BIACORE 2000 biosensor. Liposomes were immobilized on a L1 pioneer sensor chip (BIAcore) with a flow rate of 2-5  $\mu\text{L}/\text{min}$  until a surface response of 6000-7000 was achieved. Surfaces were stabilized by injecting 50  $\mu\text{L}$  of 50 mM NaOH followed by 50  $\mu\text{L}$  of 10 mM glycine pH 2.0 at a flow rate of 100  $\mu\text{L}/\text{min}$ . Prior to injection of MA or MACA proteins, 0.1 mg/mL BSA was injected over the surfaces for 5 minutes at 10  $\mu\text{L}/\text{min}$ . Following the injection of BSA, 100  $\mu\text{L}$  of recombinant HIV-1 MA proteins (0.1 mg/mL) was injected over the surface. The surfaces were regenerated with 50  $\mu\text{L}$  50 mM NaOH and 50  $\mu\text{L}$  10 mM glycine pH 2.0.

#### NMR spectroscopy of HIV-1 Myr-MA<sub>111</sub>(L8A)

$^{15}\text{N}$  or  $^{15}\text{N}/^{13}\text{C}$  labeled Myr-MA<sub>111</sub>(L8A) samples (0.5 mM to 1.0 mM) for NMR experiments were prepared in 90%  $^1\text{H}_2\text{O}/10\%$   $^2\text{H}_2\text{O}$  containing 25 mM NaPi pH 5.5, 100 mM NaCl, 1 mM  $\beta\text{ME}$ . NMR data were collected at 23°C on a Varian Inova 600 MHz spectrometer equipped with a triple-resonance  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  probe and z-axis pulsed field gradient capability. The backbone amides of Myr-MA<sub>111</sub>(L8A) were assigned using the following NMR experiments:  $^{15}\text{N}/^1\text{H}$  HSQC (Mori et al., 1995), 3D  $^{15}\text{N}$ -edited NOESY-HSQC (Mori et al., 1995; Zhang et al., 1994), 3D  $^{15}\text{N}$ -edited TOCSY-HSQC (Zhang et al., 1994), HNCA, CACONH (Grzesiek et al., 1993), HNCACB (Wittekind, 1993), CCONH (Grzesiek et al., 1993), and HNCO (Kay et al., 1994). All spectra were processed using FELIX (MSI) and analyzed using XEASY (Bartels et al., 1995).

## NMR titration of HIV-1 MA<sub>108</sub> and phosphatidylinositol-

### 3, 4, 5-triphosphate

<sup>15</sup>N labeled MA<sub>108</sub> (0.3 mM) was prepared in 90% <sup>1</sup>H<sub>2</sub>O / 10% <sup>2</sup>H<sub>2</sub>O containing 25 mM *bis*-Tris D<sub>19</sub> (Cambridge Isotope Lab) pH 6.2, 100 mM NaCl. <sup>15</sup>N/<sup>1</sup>H HSQC spectra (Mori et al., 1995) were collected at 30°C on a Varian Inova 600 MHz spectrometer. To identify the binding site of phosphatidylinositol 3,4,5-triphosphate (PtdIns(3,4,5)P<sub>3</sub>) on the MA structure, dibutyryl PtdIns(3,4,5)P<sub>3</sub> (Echelon Research Lab), a water soluble derivative of PtdIns(3,4,5)P<sub>3</sub>, was titrated into MA<sub>1-108</sub> and <sup>15</sup>N/<sup>1</sup>H HSQC spectra were collected. NMR data were processed using FELIX (MSI) and analyzed using XEASY (Bartels et al., 1995).

## **Results**

### Expression and purification of myristoylated HIV-1 MA proteins

As reviewed above, HIV-1 Gag is cotranslationally myristoylated on its N-terminus. In this process, the N-terminal methionine is removed and a myristoyl group is covalently attached to the amino group of the subsequent glycine residue. N-myristoylation is an eukaryotic protein modification that is normally carried out cotranslationally by N-myristoyl transferase (NMT) (Towler et al., 1988). However, protein myristoylation can be reconstituted in *E. coli* using a dual plasmid expression system and exogenous myristic acid (Duronio et al., 1990). In this system, one plasmid expresses protein of interest, which contains a myristoylation signal, and the other plasmid expresses yeast NMT. When both proteins are induced, NMT utilizes the myristic acid supplied in the medium to myristoylate the target protein. This system was used to make all of the myristoylated MA proteins described in this study.

The expression and purification of the HIV-1 Myr-MA and Myr-MA<sub>111</sub>(L8A) proteins are shown in Figures 2.3A and 2.4A. The N-myristoylation of the purified proteins was confirmed by electrospray mass spectrometry (Figures 2.3B and 2.4B). The purified proteins migrated as single species with molecular masses corresponding to the fully myristoylated proteins.

#### Membrane binding of HIV-1 MA and MACA proteins

The MA domain of HIV-1 Gag is responsible for membrane targeting and membrane binding. Genetic analyses have shown that both the N-myristoylation and a highly basic region near the N-terminal of MA are important for this function (Bryant and Ratner, 1990; Gottlinger et al., 1989; Pal et al., 1990; Yuan et al., 1993; Zhou et al., 1994), but pure recombinant myristoylated proteins have not been prepared previously, nor have biochemical studies been performed on myristoylated MA proteins. We therefore tested the membrane binding of HIV-1 MA and MACA proteins using surface plasmon resonance biosensor (Figure 2.5).

A series of different proteins was tested in order to determine what elements are important for the MA/Gag membrane binding interaction. Binding sensorgrams exhibited complex kinetics, we therefore only analyzed the data qualitatively, by comparing maximal response values (RU) at equilibrium. These comparisons are valid because the binding reactions were performed under identical conditions. The first set of experiments evaluated the importance of both the N-myristoyl group and the ionic interaction. As seen in Figure 2.5, Myr-MACA<sub>220</sub> had a higher binding affinity to lipid membrane than did unmyristoylated MA-CA (maximal equilibrium binding was 700 RU (Myr-MACA<sub>220</sub>) vs. 500 RU (MACA<sub>220</sub>) on phosphatidylcholine (PC) lipid surface and 2700 RU (Myr-

Figure 2.3. Purification and characterization of Myr-MA protein. A) Expression and purification of the Myr-MA protein. Lane 1, molecular weight standards; lane 2, total cellular BL21(DE3) *E. coli* proteins prior to induced expression of the Myr-MA protein; lane 3, total cellular BL21(DE3) *E. coli* proteins following induction of the Myr-MA expression; lane 4, fully purified Myr-MA protein. B) Electrospray mass spectrum of the purified Myr-MA protein, showing that the protein is fully myristoylated. The calculated mass of the unmyristoylated MA protein (14712 g/mol) is indicated by an arrow.

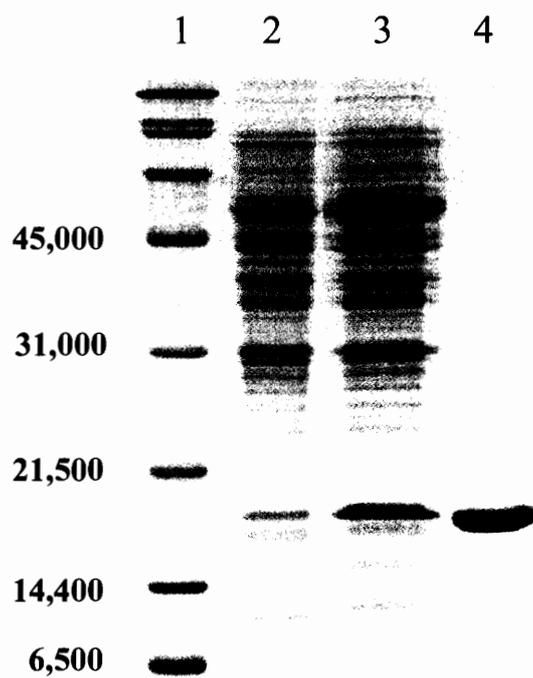
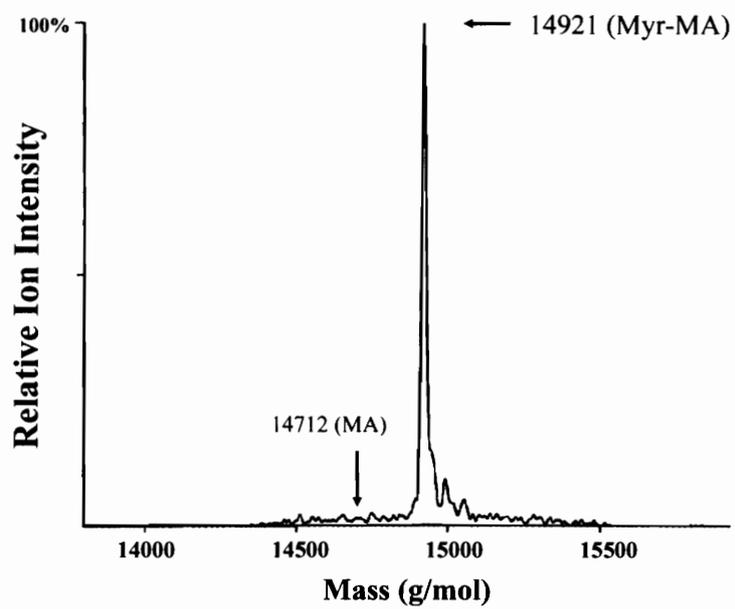
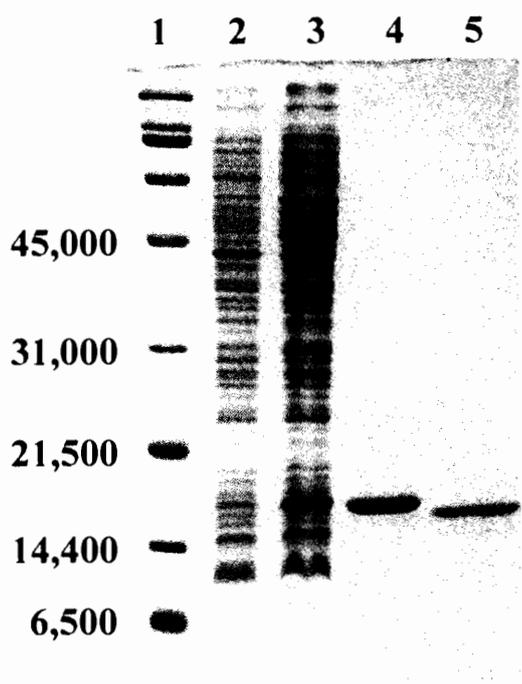
**A****B**

Figure 2.4. Purification and characterization of Myr-MA<sub>111</sub>(L8A) protein. A) Expression and purification of the Myr-MA<sub>111</sub>(L8A) protein. Lane 1, molecular weight standards; lane 2, total cellular BL21(DE3) *E. coli* proteins prior to induced expression of the Myr-MA<sub>111</sub>(L8A) protein; lane 3, total cellular BL21(DE3) *E. coli* proteins following induction of the Myr-MA<sub>111</sub>(L8A) expression; lane 4, Ni<sup>2+</sup> affinity purified Myr-MA<sub>111</sub>(L8A) protein; lane 5, Factor Xa processed and fully purified Myr-MA<sub>111</sub>(L8A) protein. B) Electrospray mass spectrum of the purified Myr-MA<sub>111</sub>(L8A) protein, showing that the protein is fully myristoylated. The calculated mass of the unmyristoylated MA<sub>111</sub>(L8A) protein (12891.63 g/mol) is indicated by an arrow.

**A**



**B**

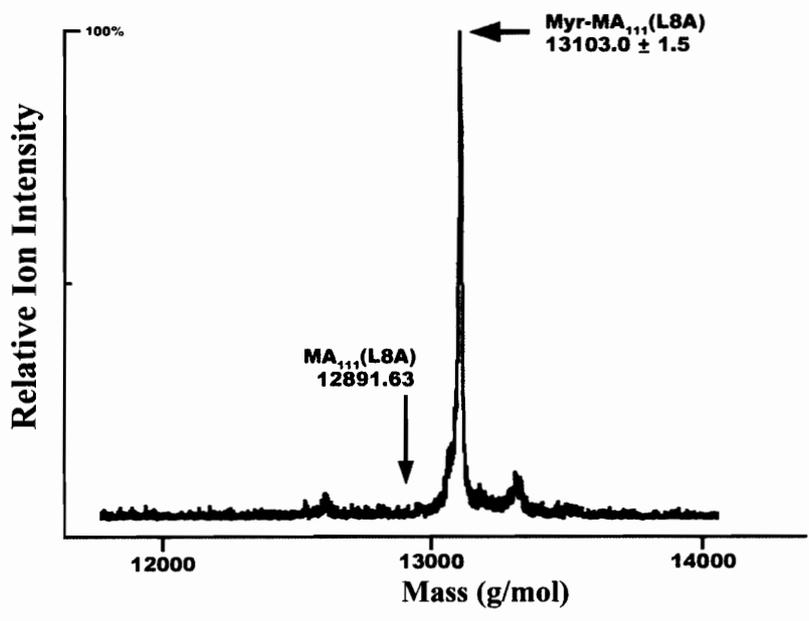
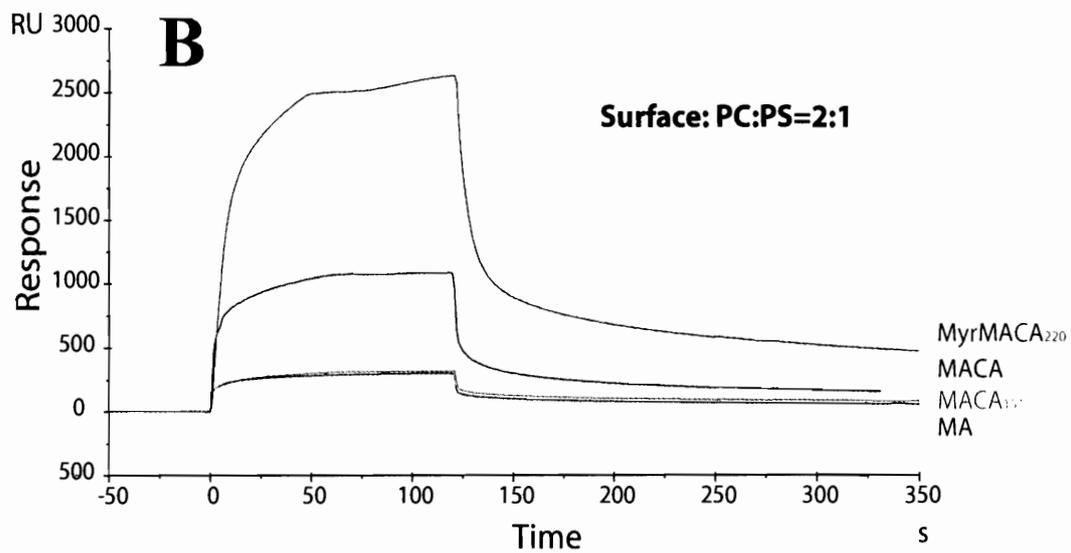
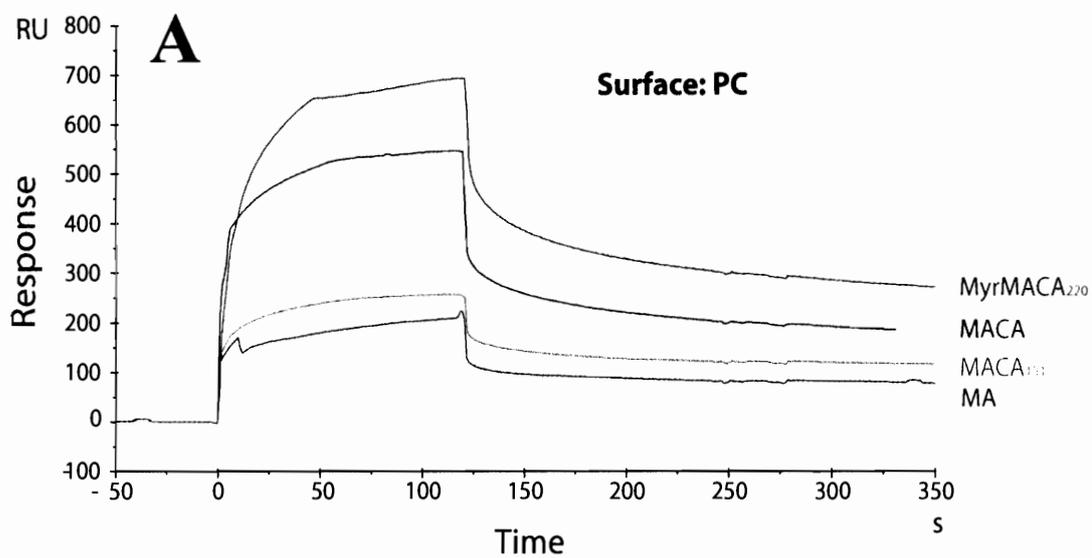


Figure 2.5. Surface plasmon resonance biosensor analysis of the membrane binding of HIV-1 MA and MACA proteins. A) Proteins binding to a phosphatidylcholine (PC) lipid surface. B) Proteins binding to a phosphatidylcholine / phosphatidylserine (PC/PS) lipid surface. Lipid surfaces were prepared by flow liposomes consisting of PC or a 2:1 mixture of PC and PS over a L1 pioneer sensor chip. Proteins were allowed to flow over the lipid surface at a constant concentration of 0.1 mg/mL at time 0. The binding levels were recorded in SPR response unit (RU) following injection of MA proteins.



MACA<sub>220</sub>) vs. 1000 RU (MACA<sub>220</sub>) on phosphatidylcholine / phosphatidylserine (PC/PS) lipid surface). Note that both proteins have significant higher binding affinities for the negatively charged PC/PS lipid surface vs. the neutral PC lipid proteins migrated as single species with molecular masses corresponding to the fully myristoylated proteins.

#### Membrane binding of HIV-1 MA and MACA proteins

The MA domain of HIV-1 Gag is responsible for membrane targeting and membrane binding. Genetic analyses have shown that both the N-myristoylation and a highly basic region near the N-terminal of MA are important for this function (Bryant and Ratner, 1990; Gottlinger et al., 1989; Pal et al., 1990; Yuan et al., 1993; Zhou et al., 1994), but pure recombinant myristoylated proteins have not been prepared previously, nor have biochemical studies been performed on myristoylated MA proteins. We therefore tested the membrane binding of HIV-1 MA and MACA proteins using surface plasmon resonance biosensor (Figure 2.5).

A series of different proteins was tested in order to determine what elements are important for the MA/Gag membrane binding interaction. Binding sensorgrams exhibited complex kinetics, we therefore only analyzed the data qualitatively, by comparing maximal response values (RU) at equilibrium. These comparisons are valid because the binding reactions were performed under identical conditions. The first set of experiments evaluated the importance of both the N-myristoyl group and the ionic interaction. As seen in Figure 2.5, Myr-MACA<sub>220</sub> had a higher binding affinity to lipid membrane than did unmyristoylated MA-CA (maximal equilibrium binding was 700 RU (Myr-MACA<sub>220</sub>) vs. 500 RU (MACA<sub>220</sub>) on phosphatidylcholine (PC) lipid surface and 2700 RU (Myr-MACA<sub>220</sub>) vs. 1000 RU (MACA<sub>220</sub>) on phosphatidylcholine / phosphatidylserine (PC/PS)

lipid surface). Note that both proteins have significant higher binding affinities for the negatively charged PC/PS lipid surface vs. the neutral PC lipid surface. These results are all consistent with the model that both the N-myristoylation of HIV-1 Gag and electrostatic interaction between the basic amino acids on the N-terminal of Gag and negatively charged lipid membrane contribute to membrane binding.

A second set of experiments tested the importance of protein-protein interactions for Gag membrane binding. As shown in Figure 2.5, the longer Gag constructs consistently had higher binding affinities to both lipid membranes. On the PC lipid surface, MACA bound with a maximal equilibrium response of 700 RU while MACA<sub>151</sub> and MA bound with 250 RU and 150 RU, respectively. Similarly, on the mixed PC/PS lipid surface, the binding affinity order was again MACA<sub>220</sub> > MACA<sub>151</sub> > MA, although in this case the CA<sub>151</sub> domain made only a small contribution to membrane binding affinity. HIV-1 CA makes important Gag-Gag interactions and this result is consistent with the idea that Gag-Gag interactions increase the avidity of membrane binding.

#### NMR studies of HIV-1 MA<sub>111</sub>(L8A) and structural changes within HIV-1 MA upon myristoylation

We performed a series of NMR studies aimed at determining the structural consequences of N-myristoylation of the HIV-1 MA protein. Our initial studies were performed using the full length wild type Myr-MA protein. However, <sup>1</sup>H/<sup>15</sup>N HSQC spectra indicated that the C-terminal tail of Myr-MA is disordered. These residues are also disordered in the NMR structure of the unmyristoylated MA (Massiah et al., 1994) and they exhibited no significant chemical shift changes upon myristoylation. The unstructured, C-terminal tail of MA was therefore removed to decrease the protein's

correlation time and thereby decrease line widths and increase the quality of the NMR data. This change did, in fact, reduce line width by ~4Hz (from 28Hz to 24Hz). However, the truncated wild type Myr-MA protein was still poorly soluble and precipitated during the course of NMR data collection. This is presumably due to the dynamic exchange of the protein between the myristoyl “in” and “out” conformation. To reduce problems with dynamic exchange, the L8A mutation was introduced into MA. Genetic studies have shown that this mutation blocks Gag membrane binding without affecting Gag myristoylation (Paillart and Gottlinger, 1999), and the mutation presumably functions by helping to keep the myristoyl group sequestered. The L8A mutation was therefore expected to increase the solubility of the myristoylated MA protein and decrease exchange between the myristoyl “in” and “out” conformations. This mutation did improve spectral quality without causing significant global chemical shift changes, and the mutant protein Myr-MA<sub>111</sub>(L8A) was therefore used in subsequent studies.

We were able to assign 100 of the 111 backbone amide resonances (114 amino acids in the protein, 3 of which are prolines) in the <sup>15</sup>N HSQC spectrum of Myr-MA<sub>111</sub>(L8A) (Figure 2.6). The data used to make sequential assignments for residues S67 to L75 are illustrated in Figures 2.7 and 2.8. Figure 2.7 shows strips of the 3D <sup>15</sup>N-edited NOESY-HSQC data with the *i* to *i*+1 NOEs that help establish sequential connectivities. Figure 2.8 shows intra and inter residue C<sub>α</sub> correlations and C<sub>β</sub> correlations in the HNCACB data. These data are all of high quality. However, it was not possible to assign 11 backbone amide resonances owing to a lack of correlation data for these spin systems (Figure 2.9). This problem presumably reflects dynamic exchange processes associated with the N-myristoyl group.

Figure 2.6.  $^1\text{H}/^{15}\text{N}$  HSQC spectrum of the Myr-MA<sub>111</sub>(L8A) protein.

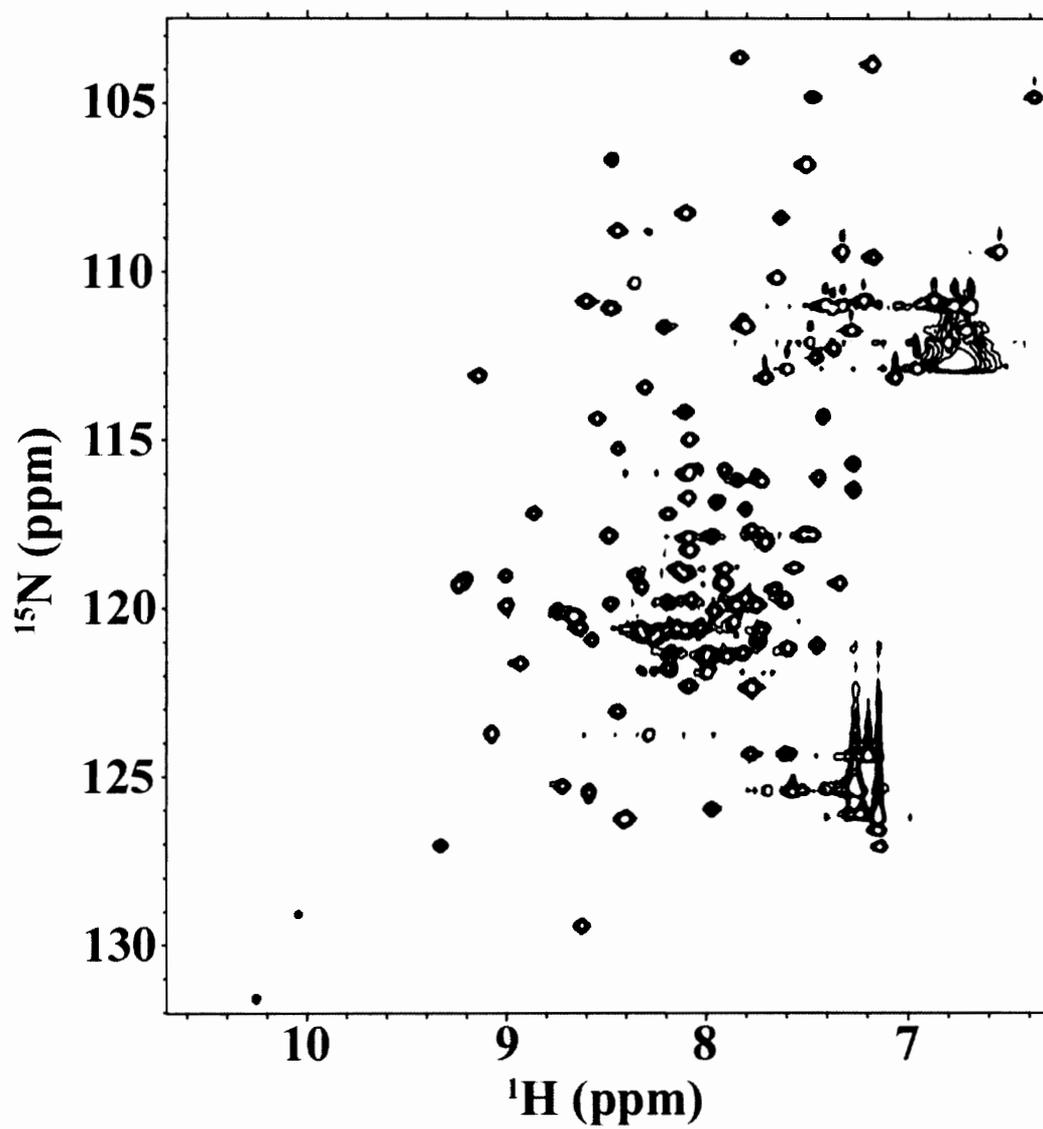


Figure 2.7. Strips from the 3D  $^{15}\text{N}$ -edited NOESY-HSQC data of the Myr-MA<sub>111</sub>(L8A) protein, showing the  $i$  to  $i+1$  NOE connectivity.

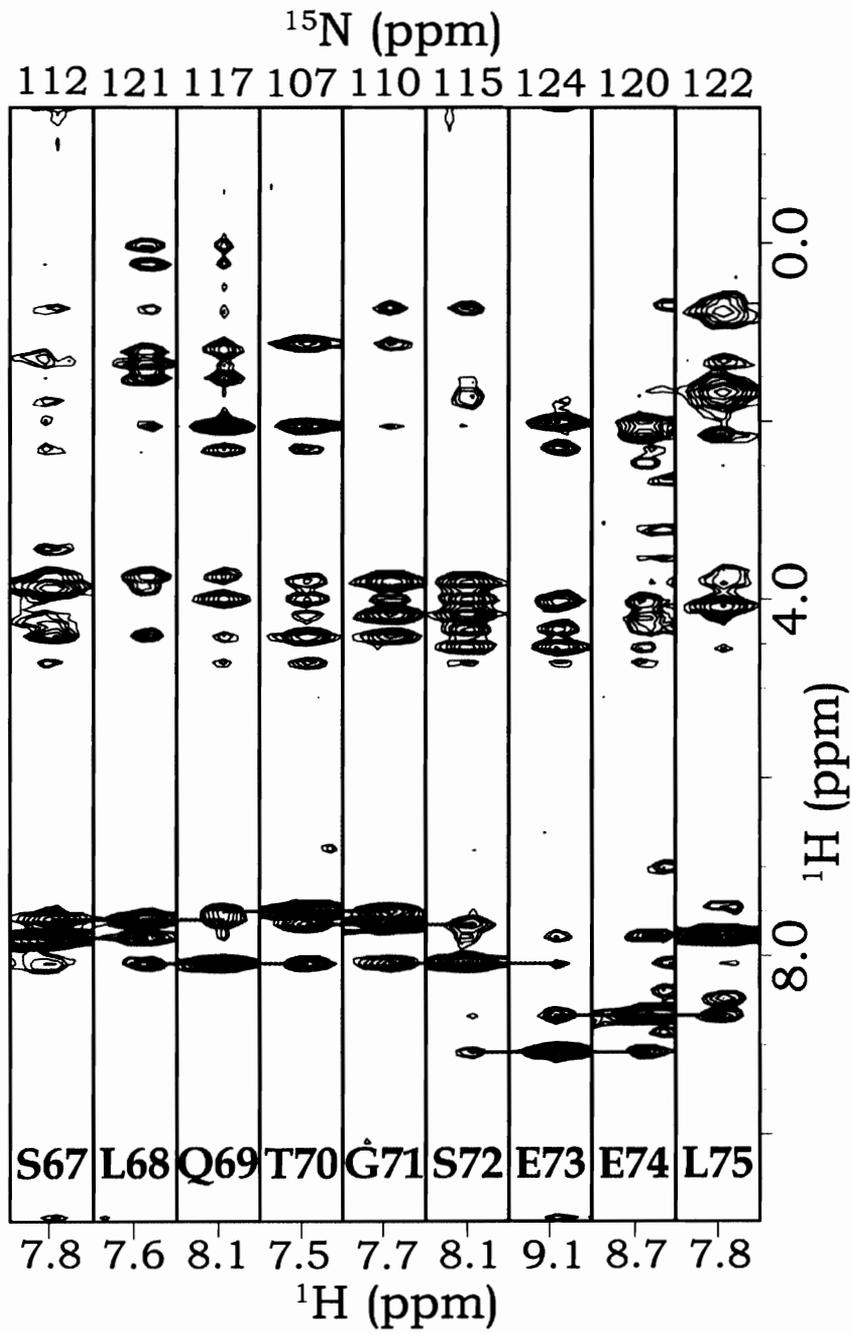


Figure 2.8. Strips from the HNCACB data of the Myr-MA<sub>111</sub>(L8A) protein, showing the intra and inter residue C<sub>α</sub> and C<sub>β</sub> correlations.

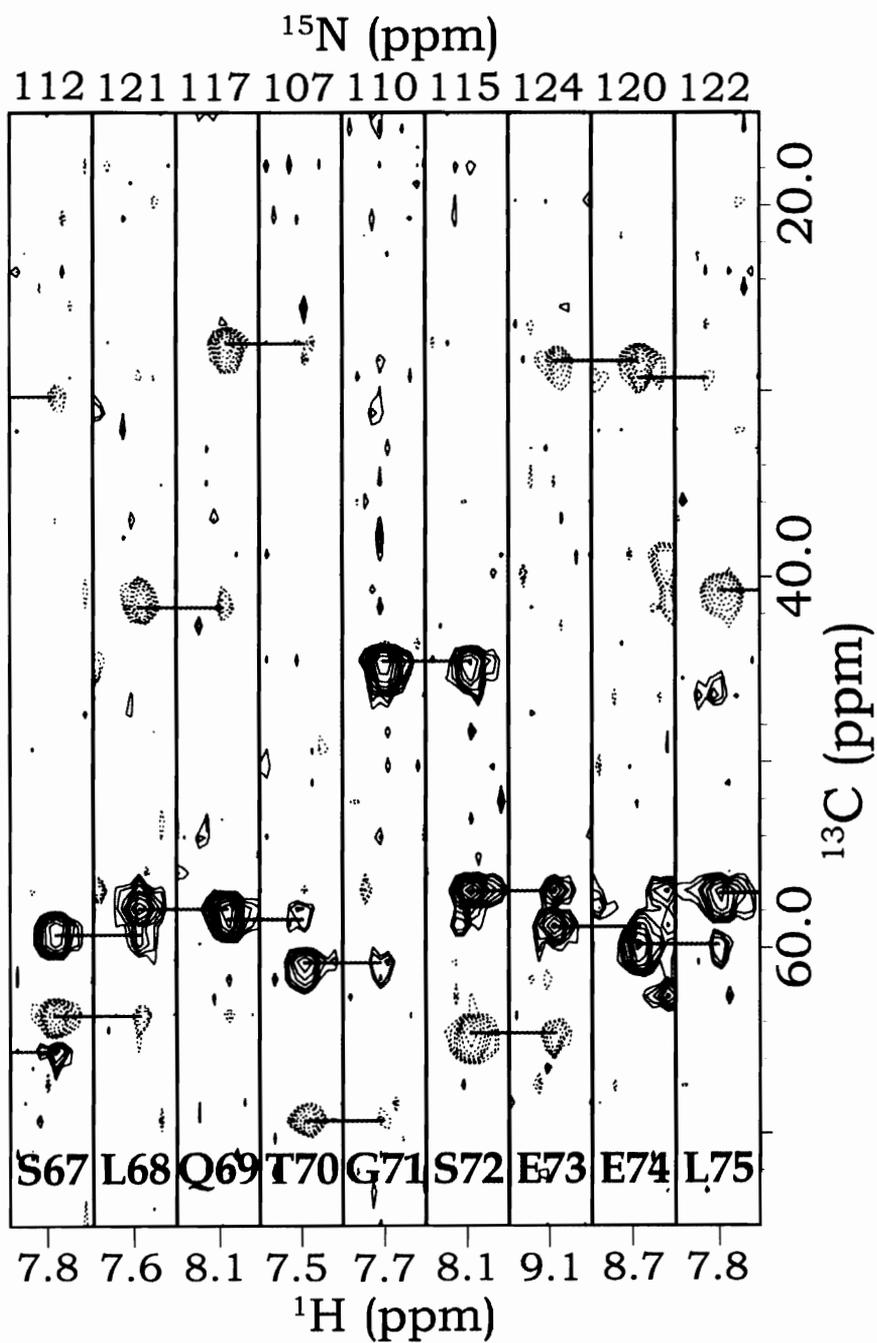
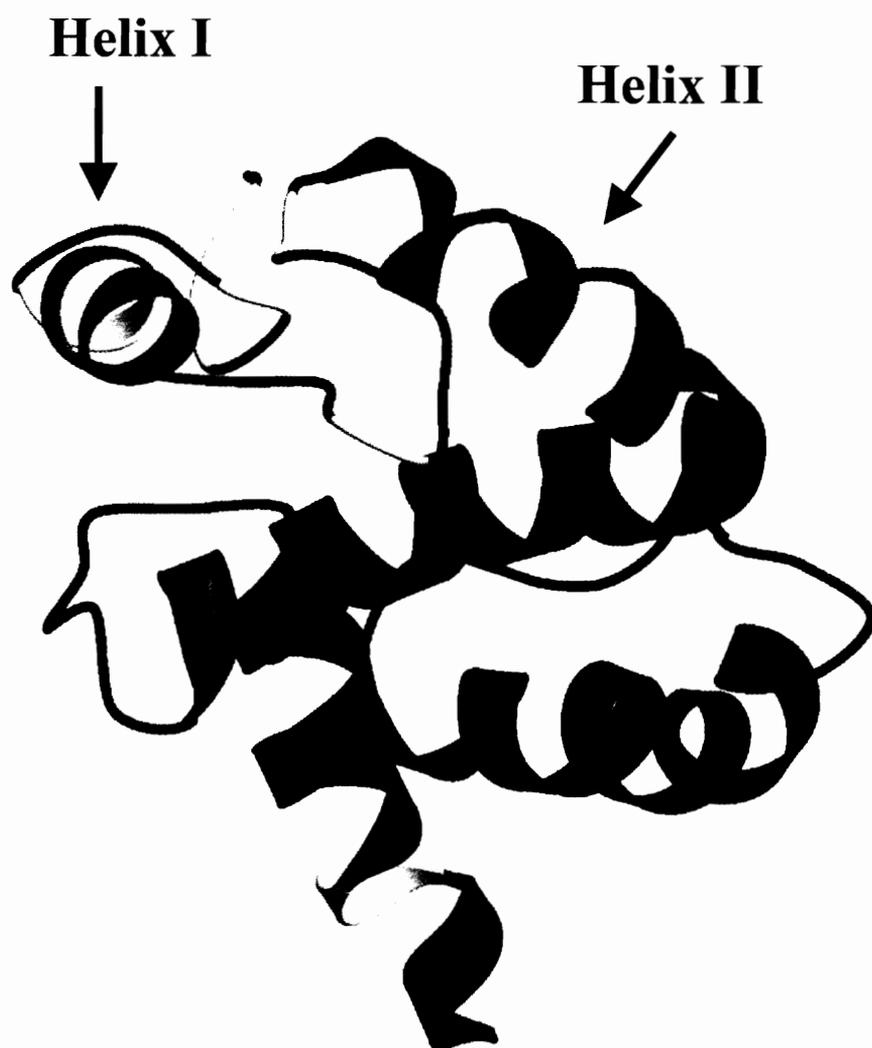


Figure 2.9. Site of MA structural changes upon N-myristoylation. Residues with significant chemical shift changes in  $^1\text{H}/^{15}\text{N}$  HSQC spectrum ( $>0.2\text{ppm}$  in  $^1\text{H}$  or  $>0.3\text{ppm}$  in  $^{15}\text{N}$ ) are shown in red, and residues with no significant chemical shift changes are shown in blue. Unassigned residues are shown in yellow.



Upon myristoylation of MA, 38 backbone amide resonances show significant chemical shift changes ( $>0.2\text{ppm}$  in  $^1\text{H}$  or  $>0.3\text{ppm}$  in  $^{15}\text{N}$ ). The majority of these residues are localized to the N-terminal of the MA protein, especially helices 1 and 2 (Figure 2.9). These results suggest that the N-myristoylation changes the local structural of the N-terminus of the MA protein, possibly by packing against the protein, or alternatively by causing a conformational change that alters the local magnetic environment of the shifted backbone amides.

#### Binding site of phosphatidylinositol-3, 4, 5-triphosphate on MA

Recent biochemical studies on HIV-1 Gag assembly *in vitro* have suggested that PtdInsPs may bind MA and play a role in Gag membrane targeting and assembly (Campbell et al., 2001). However, the binding site of PtdInsPs has not been mapped. We therefore used NMR spectroscopy to map the binding site of a water-soluble dibutanoyl form of PtdIns(3,4,5)P<sub>3</sub> on MA<sub>108</sub>. When PtdIns(3,4,5)P<sub>3</sub> was titrated into MA<sub>108</sub>, several backbone amides exhibited significant chemical shift changes (Figure 2.10). These residues were highly clustered around the sulfate binding sites observed in the crystal structure of MA (Figure 2.11) (Hill et al., 1996). These residues included the basic residues R22, K26, K27 and R39 that interact directly with the sulfate anions in the crystal structure. In the membrane binding model of MA (Hill et al., 1996), these sites are on the membrane binding surface of the MA trimer and should therefore be able to bind to the phosphate head group of PtdInsPs. Although this remains an attractive hypothesis, we were unable to demonstrate any specificity in PtdInsPs (or IPs) binding. For example, a  $^{15}\text{N}$  HSQC titration experiment showed that IS<sub>6</sub>, an analog of IP<sub>6</sub> that is not active in

Figure 2.10. Overlay of the  $^{15}\text{N}$  HSQC spectra of MA<sub>108</sub> showing the effect of titrating in PtdIns(3,4,5)P<sub>3</sub>. Residues with significant chemical shift changes are labeled and highlighted with black boxes.

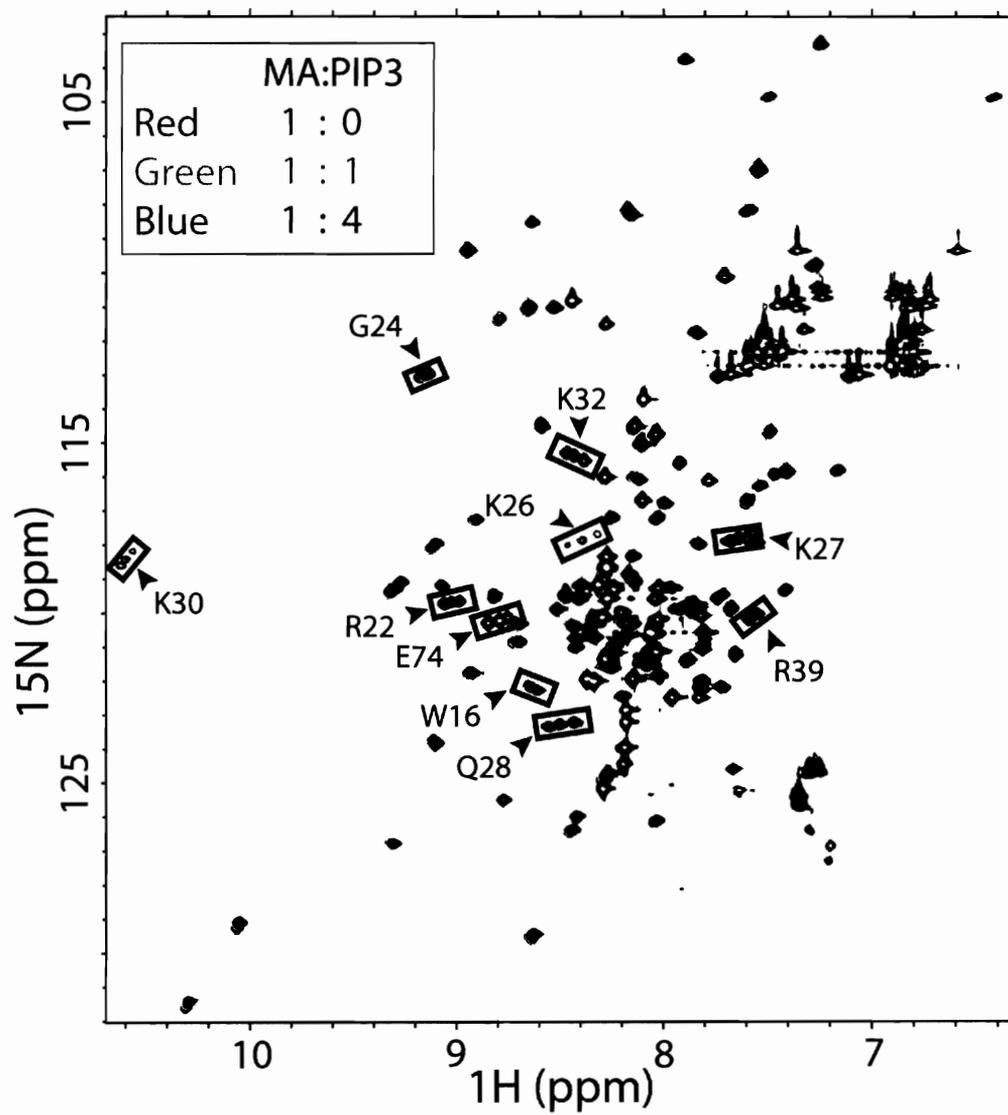


Figure 2.11. Crystal structure of the HIV-1 MA trimer (Hill et al., 1996), showing the sulfate (and PtdIns(3,4,5)P<sub>3</sub>) binding sites. Residues with significant chemical shift changes upon addition of PtdIns(3,4,5)P<sub>3</sub> are labeled and highlighted with red arrows.



modulating Gag assembly, also binds to precisely the same binding site as PtdIns(3,4,5)P<sub>3</sub>.

## **Discussion**

### Membrane binding studies

MA is the N-terminal domain of the Gag polyprotein and is responsible for Gag membrane targeting and membrane binding, both of which are essential steps in viral assembly and replication. The MA protein is myristoylated on its N-terminus and has a highly basic patch near its N-terminus. Genetic studies have shown that both elements are important for MA membrane targeting and binding function (Bryant and Ratner, 1990; Gottlinger et al., 1989; Pal et al., 1990; Yuan et al., 1993; Zhou et al., 1994). The crystal structure of the MA trimer has provided a model for MA membrane binding, which agrees well with the genetic studies (Hill et al., 1996).

However, no biochemical systems have previously been available for studying the MA/membrane interaction, and this was one goal of my studies. Surface plasmon resonance biosensor experiments were used to test the importance of the MA myristoyl group, the basic patch, and Gag-Gag interactions for membrane binding. These experiments showed that both the myristoyl group and electrostatic interactions contribute to MA membrane binding affinities. They also showed that longer Gag constructs, which included the protein-protein interaction sites within the CA domain of Gag have higher membrane binding affinities than MA alone, suggesting that the Gag-Gag interactions add cooperativity to membrane binding. HIV-1 CA protein is composed of two distinct domains, the N-terminal domain (NTD) and the C-terminal domain (CTD). The CTD dimerizes in solution and crystal, contributes significantly to Gag-Gag

interactions, and is essential for virion assembly (Gamble et al., 1997; Zhang et al., 1996).

### MA myristoylation

Myristic acid is less hydrophobic than other fatty acids (e.g., palmitic acid), and for this reason the membrane binding of myristoylated proteins is usually reversible. Furthermore, controlled exposure of the myristoyl group is often used as a mechanism to regulate the membrane binding. Such “myristoyl switches” have been well characterized in several proteins, including recoverin and ARF (Ames et al., 1997; Goldberg, 1998). Several studies on HIV-1 MA have suggested that the membrane binding of MA is also regulated by such a myristoyl switch mechanism. For example, deletions of the globular core of MA have been shown to dramatically increase the association of the MA protein with membranes (Spearman et al., 1997), indicating that the  $\alpha$ -helical core is involved in the sequestration of the myristoyl group. Furthermore, point mutations that expose the myristoyl group can redirect Gag membrane binding and viral particle formation to all intracellular membrane compartments and thereby severely impair the extracellular viral particle production (Freed et al., 1994). Conversely, single amino acid substitutions (e.g., L8A) can also block Gag membrane binding and viral assembly without affecting the Gag myristoylation (Paillart and Gottlinger, 1999), possibly by keeping the myristoyl group sequestered. These observations indicate that MA can adopt distinct myristoyl “in” and myristoyl “out” conformations, with the later conformation used to bind membrane. However, there has not, to date, been any direct evidence for such a conformation change.

By comparing the  $^{15}\text{N}$  HSQC spectrum of MyrMA<sub>111</sub>(L8A) and wild type unmyristoylated MA protein, we have shown that MA does indeed undergo structural changes upon myristoylation and that these changes are mainly localized to the N-terminal of the protein, especially helices I and II, suggesting the myristoyl group may pack against this region of MA in forming the myristoyl “in” conformation. The NMR signals from the side chain carbons and protons are very weak in this region, however. This is apparently due to dynamic rearrangements (Kay, 1998) and suggests that the myristoyl group in the protein is in intermediate exchange between myristoyl “in” and “out” conformations. These results are consistent with the myristoyl switch model, in which the myristoyl group is sequestered by the  $\alpha$ -helical core of MA when the protein is in solution and then exposed for insertion into a lipid bilayer when the protein binds to membrane. Unfortunately, I was unable to completely overcome the problems with Myr-MA conformational dynamics despite of intensive effort (e.g., varying protein construct, ionic strength, and temperature). Further structural studies on the myristoylated HIV-1 MA protein are therefore required to determine the detailed molecular mechanism of this myristoyl switch.

#### PtdInsPs binding

There is also evidence that cellular factors, including IPs and PtdInsPs, are involved HIV-1 viral assembly (Campbell et al., 2001). We have used NMR chemical shift perturbation experiments to map the binding site of PtdIns(3,4,5)P<sub>3</sub> on the MA structure. The binding site is composed of basic residues on the membrane binding surface of MA and corresponds well with the sulfate binding sites observed in the crystal structure of the HIV-1 MA trimer. It is tempting to speculate that MA can use these sites

to bind the phosphate head group of PtdIns(3,4,5)P<sub>3</sub> and that this binding may help target Gag protein to the assembly sites on the plasma membrane. This model accommodates both the electrostatic interaction between the basic residues and the negatively charged lipid head group as well as the hydrophobic interaction of the myristoyl group in lipid bilayer. Indeed, the binding of MA to the lipid head groups could even be linked to the exposure of the myristoyl group for lipid insertion. Thus, PtdIns(3,4,5)P<sub>3</sub> binding could form part of the myristoyl switching mechanism. However, we were unable to show specificity of PtdIns(3,4,5)P<sub>3</sub> binding using either surface plasmon resonance biosensor experiments (not shown) or NMR spectroscopy. Thus, the biological function of the PtdIns(3,4,5)P<sub>3</sub> binding interactions that we observe in vitro remains to be established.

## References

- Ames, J. B., Ishima, R., Tanaka, T., Gordon, J. I., Stryer, L., and Ikura, M. (1997). Molecular mechanics of calcium-myristoyl switches. *Nature* 389, 198-202.
- Ames, J. B., Tanaka, T., Stryer, L., and Ikura, M. (1996). Portrait of a myristoyl switch protein. *Curr. Opin. Struct. Biol.* 6, 432-438.
- Amor, J. C., Harrison, D. H., Kahn, R. A., and Ringe, D. (1994). Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature* 372, 704-708.
- Anderson, R. G., and Jacobson, K. (2002). A role for lipid shells in targeting proteins to caveolae, rafts, and other lipid domains. *Science* 296, 1821-1825.
- Bartels, C., Xia, T.-h., Billeter, M., Guntert, P., and Wuthrich, K. (1995). The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* 5, 1-10.
- Brown, D. A., and London, E. (2000). Structure and function of sphingolipid- and cholesterol-rich membrane rafts. *J. Biol. Chem.* 275, 17221-17224.
- Bryant, M., and Ratner, L. (1990). Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc. Natl. Acad. Sci. U S A* 87, 523-527.

Campbell, S., Fisher, R. J., Towler, E. M., Fox, S., Issaq, H. J., Wolfe, T., Phillips, L. R., and Rein, A. (2001). Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proc. Natl. Acad. Sci. U S A* *98*, 10875-10879.

Campbell, S., and Rein, A. (1999). In vitro assembly properties of human immunodeficiency virus type 1 Gag protein lacking the p6 domain. *J. Virol.* *73*, 2270-2279.

Campbell, S., and Vogt, V. M. (1997). In vitro assembly of virus-like particles with Rous sarcoma virus Gag deletion mutants: identification of the p10 domain as a morphological determinant in the formation of spherical particles. *J. Virol.* *71*, 4425-4435.

Chen, B. K., Rousso, I., Shim, S., and Kim, P. S. (2001). Efficient assembly of an HIV-1/MLV Gag-chimeric virus in murine cells. *Proc. Natl. Acad. Sci. U S A* *98*, 15239-15244.

Deichaite, I., Casson, L. P., Ling, H. P., and Resh, M. D. (1988). In vitro synthesis of pp60v-src: myristylation in a cell-free system. *Mol. Cell. Biol.* *8*, 4295-4301.

Dizhoor, A. M., Olshevskaya, E. V., Henzel, W. J., Wong, S. C., Stults, J. T., Ankoudinova, I., and Hurley, J. B. (1995). Cloning, sequencing, and expression of a 24-kDa Ca(2+)-binding protein activating photoreceptor guanylyl cyclase. *J. Biol. Chem.* *270*, 25200-25206.

Dorfman, T., Mammano, F., Haseltine, W. A., and Gottlinger, H. G. (1994). Role of the matrix protein in the virion association of the human immunodeficiency virus type 1 envelope glycoprotein. *J. Virol.* *68*, 1689-1696.

Duronio, R. J., Jackson-Machelski, E., Heuckeroth, R. O., Olins, P. O., Devine, C. S., Yonemoto, W., Slice, L. W., Taylor, S. S., and Gordon, J. I. (1990). Protein N-myristoylation in *Escherichia coli*: reconstitution of a eukaryotic protein modification in bacteria. *Proc. Natl. Acad. Sci. U S A* *87*, 1506-1510.

Edbauer, C. A., and Naso, R. B. (1983). Cytoskeleton-associated Pr65gag and retrovirus assembly. *Virology* *130*, 415-426.

Edbauer, C. A., and Naso, R. B. (1984). Cytoskeleton-associated Pr65gag and assembly of retrovirus temperature-sensitive mutants in chronically infected cells. *Virology* *134*, 389-397.

Ehrlich, L. S., Agresta, B. E., and Carter, C. A. (1992). Assembly of recombinant human immunodeficiency virus type 1 capsid protein in vitro. *J. Virol.* *66*, 4874-4883.

Facke, M., Janetzko, A., Shoeman, R. L., and Krausslich, H. G. (1993). A large deletion in the matrix domain of the human immunodeficiency virus gag gene redirects virus particle assembly from the plasma membrane to the endoplasmic reticulum. *J. Virol.* *67*, 4972-4980.

- Flaherty, K. M., Zozulya, S., Stryer, L., and McKay, D. B. (1993). Three-dimensional structure of recoverin, a calcium sensor in vision. *Cell* 75, 709-716.
- Ford, M. G., Pearse, B. M., Higgins, M. K., Vallis, Y., Owen, D. J., Gibson, A., Hopkins, C. R., Evans, P. R., and McMahon, H. T. (2001). Simultaneous binding of PtdIns(4,5)P<sub>2</sub> and clathrin by AP180 in the nucleation of clathrin lattices on membranes. *Science* 291, 1051-1055.
- Franke, E. K., Yuan, H. E., and Luban, J. (1994). Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* 372, 359-362.
- Freed, E. O. (1998). HIV-1 gag proteins: diverse functions in the virus life cycle. *Virology* 251, 1-15.
- Freed, E. O., and Martin, M. A. (1996). Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. *J. Virol.* 70, 341-351.
- Freed, E. O., Orenstein, J. M., Buckler-White, A. J., and Martin, M. A. (1994). Single amino acid changes in the human immunodeficiency virus type 1 matrix protein block virus particle production. *J. Virol.* 68, 5311-5320.
- Fuller, S. D., Wilk, T., Gowen, B. E., Krausslich, H. G., and Vogt, V. M. (1997). Cryo-electron microscopy reveals ordered domains in the immature HIV-1 particle. *Curr. Biol.* 7, 729-738.
- Gamble, T. R., Yoo, S., Vajdos, F. F., von Schwedler, U. K., Worthylake, D. K., Wang, H., McCutcheon, J. P., Sundquist, W. I., and Hill, C. P. (1997). Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* 278, 849-853.
- Garrus, J. E., von Schwedler, U. K., Pornillos, O. W., Morham, S. G., Zavitz, K. H., Wang, H. E., Wettstein, D. A., Stray, K. M., Cote, M., Rich, R. L., *et al.* (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* 107, 55-65.
- Gheysen, D., Jacobs, E., de Foresta, F., Thiriart, C., Francotte, M., Thines, D., and De Wilde, M. (1989). Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell* 59, 103-112.
- Goldberg, J. (1998). Structural basis for activation of ARF GTPase: mechanisms of guanine nucleotide exchange and GTP-myristoyl switching. *Cell* 95, 237-248.
- Goldberg, J., Nairn, A. C., and Kuriyan, J. (1996). Structural basis for the autoinhibition of calcium/calmodulin-dependent protein kinase I. *Cell* 84, 875-887.
- Gottlinger, H. G., Sodroski, J. G., and Haseltine, W. A. (1989). Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U S A* 86, 5781-5785.

- Gross, I., Hohenberg, H., Huckhagel, C., and Krausslich, H. G. (1998). N-Terminal extension of human immunodeficiency virus capsid protein converts the in vitro assembly phenotype from tubular to spherical particles. *J. Virol.* *72*, 4798-4810.
- Gross, I., Hohenberg, H., and Krausslich, H. G. (1997). In vitro assembly properties of purified bacterially expressed capsid proteins of human immunodeficiency virus. *Eur. J. Biochem.* *249*, 592-600.
- Grzesiek, S., Anglister, J., and Bax, A. (1993). Correlation of backbone amide and aliphatic side-chain resonances in <sup>13</sup>C/<sup>15</sup>N-enriched proteins by isotropic mixing of <sup>13</sup>C magnetization. *J. Magn. Reson. B.* *101*, 114-119.
- Hill, C. P., Worthylake, D., Bancroft, D. P., Christensen, A. M., and Sundquist, W. I. (1996). Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proc. Natl. Acad. Sci. U S A* *93*, 3099-3104.
- Itoh, T., Koshihara, S., Kigawa, T., Kikuchi, A., Yokoyama, S., and Takenawa, T. (2001). Role of the ENTH domain in phosphatidylinositol-4,5-bisphosphate binding and endocytosis. *Science* *291*, 1047-1051.
- Kay, L. E. (1998). Protein dynamics from NMR. *Nat. Struct. Biol.* *5 Suppl*, 513-517.
- Kay, L. E., Xu, G. Y., and Yamazaki, T. (1994). Enhanced-sensitivity triple-resonance spectroscopy with minimal H<sub>2</sub>O saturation. *J. Mag. Res.* *109*, 129-133.
- Kim, J., Blackshear, P. J., Johnson, J. D., and McLaughlin, S. (1994a). Phosphorylation reverses the membrane association of peptides that correspond to the basic domains of MARCKS and neuromodulin. *Biophys. J.* *67*, 227-237.
- Kim, J., Shishido, T., Jiang, X., Aderem, A., and McLaughlin, S. (1994b). Phosphorylation, high ionic strength, and calmodulin reverse the binding of MARCKS to phospholipid vesicles. *J. Biol. Chem.* *269*, 28214-28219.
- Klikova, M., Rhee, S. S., Hunter, E., and Ruml, T. (1995). Efficient in vivo and in vitro assembly of retroviral capsids from Gag precursor proteins expressed in bacteria. *J. Virol.* *69*, 1093-1098.
- Kondo, E., and Gottlinger, H. G. (1996). A conserved LXXLF sequence is the major determinant in p6Gag required for the incorporation of human immunodeficiency virus type 1 Vpr. *J. Virol.* *70*, 159-164.
- Li, S., Hill, C. P., Sundquist, W. I., and Finch, J. T. (2000). Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* *407*, 409-413.
- Liemann, S., Chandran, K., Baker, T. S., Nibert, M. L., and Harrison, S. C. (2002). Structure of the reovirus membrane-penetration protein, Mu1, in a complex with its protector protein, Sigma3. *Cell* *108*, 283-295.

- Lingappa, J. R., Hill, R. L., Wong, M. L., and Hegde, R. S. (1997). A multistep, ATP-dependent pathway for assembly of human immunodeficiency virus capsids in a cell-free system. *J. Cell. Biol.* *136*, 567-581.
- Lu, Y. L., Bennett, R. P., Wills, J. W., Gorelick, R., and Ratner, L. (1995). A leucine triplet repeat sequence (LXX)<sub>4</sub> in p6Gag is important for Vpr incorporation into human immunodeficiency virus type 1 particles. *J. Virol.* *69*, 6873-6879.
- Magee, A. I., and Courtneidge, S. A. (1985). Two classes of fatty acid acylated proteins exist in eukaryotic cells. *EMBO J.* *4*, 1137-1144.
- Mariani, R., Rutter, G., Harris, M. E., Hope, T. J., Krausslich, H. G., and Landau, N. R. (2000). A block to human immunodeficiency virus type 1 assembly in murine cells. *J. Virol.* *74*, 3859-3870.
- Massiah, M. A., Starich, M. R., Paschall, C., Summers, M. F., Christensen, A. M., and Sundquist, W. I. (1994). Three-dimensional structure of the human immunodeficiency virus type 1 matrix protein. *J. Mol. Biol.* *244*, 198-223.
- Matthews, S., Barlow, P., Boyd, J., Barton, G., Russell, R., Mills, H., Cunningham, M., Meyers, N., Burns, N., Clark, N., and et al. (1994). Structural similarity between the p17 matrix protein of HIV-1 and interferon-gamma. *Nature* *370*, 666-668.
- McLaughlin, S., and Aderem, A. (1995). The myristoyl-electrostatic switch: a modulator of reversible protein-membrane interactions. *Trends Biochem. Sci.* *20*, 272-276.
- Mori, S., Abeygunawardana, C., Johnson, M. O., and van Zijl, P. C. (1995). Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B.* *108*, 94-98.
- Nguyen, D. H., and Hildreth, J. E. (2000). Evidence for budding of human immunodeficiency virus type 1 selectively from glycolipid-enriched membrane lipid rafts. *J. Virol.* *74*, 3264-3272.
- Olson, E. N., Towler, D. A., and Glaser, L. (1985). Specificity of fatty acid acylation of cellular proteins. *J. Biol. Chem.* *260*, 3784-3790.
- Ono, A., and Freed, E. O. (2001). Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proc. Natl. Acad. Sci. U S A* *98*, 13925-13930.
- Ono, A., Huang, M., and Freed, E. O. (1997). Characterization of human immunodeficiency virus type 1 matrix revertants: effects on virus assembly, Gag processing, and Env incorporation into virions. *J. Virol.* *71*, 4409-4418.
- Ono, A., Orenstein, J. M., and Freed, E. O. (2000). Role of the Gag matrix domain in targeting human immunodeficiency virus type 1 assembly. *J. Virol.* *74*, 2855-2866.

- Paige, L. A., Nadler, M. J., Harrison, M. L., Cassady, J. M., and Geahlen, R. L. (1993). Reversible palmitoylation of the protein-tyrosine kinase p56lck. *J. Biol. Chem.* *268*, 8669-8674.
- Paillart, J. C., and Gottlinger, H. G. (1999). Opposing effects of human immunodeficiency virus type 1 matrix mutations support a myristyl switch model of gag membrane targeting. *J. Virol.* *73*, 2604-2612.
- Pal, R., Reitz, M. S., Jr., Tschachler, E., Gallo, R. C., Sarngadharan, M. G., and Veronese, F. D. (1990). Myristoylation of gag proteins of HIV-1 plays an important role in virus assembly. *AIDS Res. Hum. Retroviruses* *6*, 721-730.
- Paxton, W., Connor, R. I., and Landau, N. R. (1993). Incorporation of Vpr into human immunodeficiency virus type 1 virions: requirement for the p6 region of gag and mutational analyses. *J. Virol.* *67*, 7229.
- Peitzsch, R. M., and McLaughlin, S. (1993). Binding of acylated peptides and fatty acids to phospholipid vesicles: pertinence to myristoylated proteins. *Biochemistry* *32*, 10436-10443.
- Perrin-Tricaud, C., Davoust, J., and Jones, I. M. (1999). Tagging the human immunodeficiency virus gag protein with green fluorescent protein. Minimal evidence for colocalisation with actin. *Virology* *255*, 20-25.
- Rao, Z., Belyaev, A. S., Fry, E., Roy, P., Jones, I. M., and Stuart, D. I. (1995). Crystal structure of SIV matrix antigen and implications for virus assembly. *Nature* *378*, 743-747.
- Reil, H., Bukovsky, A. A., Gelderblom, H. R., and Gottlinger, H. G. (1998). Efficient HIV-1 replication can occur in the absence of the viral matrix protein. *EMBO J.* *17*, 2699-2708.
- Resh, M. D. (1994). Myristylation and palmitoylation of Src family members: the fats of the matter. *Cell* *76*, 411-413.
- Rothman, J. E., and Wieland, F. T. (1996). Protein sorting by transport vesicles. *Science* *272*, 227-234.
- Royer, M., Cerutti, M., Gay, B., Hong, S. S., Devauchelle, G., and Boulanger, P. (1991). Functional domains of HIV-1 gag-polyprotein expressed in baculovirus-infected cells. *Virology* *184*, 417-422.
- Royer, M., Hong, S. S., Gay, B., Cerutti, M., and Boulanger, P. (1992). Expression and extracellular release of human immunodeficiency virus type 1 Gag precursors by recombinant baculovirus-infected cells. *J. Virol.* *66*, 3230-3235.

- Shioda, T., and Shibuta, H. (1990). Production of human immunodeficiency virus (HIV)-like particles from cells infected with recombinant vaccinia viruses carrying the gag gene of HIV. *Virology* 175, 139-148.
- Sigal, C. T., Zhou, W., Buser, C. A., McLaughlin, S., and Resh, M. D. (1994). Amino-terminal basic residues of Src mediate membrane binding through electrostatic interaction with acidic phospholipids. *Proc. Natl. Acad. Sci. U S A* 91, 12253-12257.
- Silvius, J. R., and l'Heureux, F. (1994). Fluorimetric evaluation of the affinities of isoprenylated peptides for lipid bilayers. *Biochemistry* 33, 3014-3022.
- Simons, K., and Toomre, D. (2000). Lipid rafts and signal transduction. *Nat. Rev. Mol. Cell Biol.* 1, 31-39.
- Spearman, P., Horton, R., Ratner, L., and Kuli-Zade, I. (1997). Membrane binding of human immunodeficiency virus type 1 matrix protein in vivo supports a conformational myristyl switch mechanism. *J. Virol.* 71, 6582-6592.
- Subbaraya, I., Ruiz, C. C., Helekar, B. S., Zhao, X., Gorczyca, W. A., Pettenati, M. J., Rao, P. N., Palczewski, K., and Baehr, W. (1994). Molecular characterization of human and mouse photoreceptor guanylate cyclase-activating protein (GCAP) and chromosomal localization of the human gene. *J. Biol. Chem.* 269, 31080-31089.
- Sudo, Y., Valenzuela, D., Beck-Sickinger, A. G., Fishman, M. C., and Strittmatter, S. M. (1992). Palmitoylation alters protein activity: blockade of G(o) stimulation by GAP-43. *EMBO J.* 11, 2095-2102.
- Taniguchi, H., and Manenti, S. (1993). Interaction of myristoylated alanine-rich protein kinase C substrate (MARCKS) with membrane phospholipids. *J. Biol. Chem.* 268, 9960-9963.
- Thali, M., Bukovsky, A., Kondo, E., Rosenwirth, B., Walsh, C. T., Sodroski, J., and Gottlinger, H. G. (1994). Functional association of cyclophilin A with HIV-1 virions. *Nature* 372, 363-365.
- Towler, D. A., Gordon, J. I., Adams, S. P., and Glaser, L. (1988). The biology and enzymology of eukaryotic protein acylation. *Annu. Rev. Biochem.* 57, 69-99.
- Veronese, F. D., Copeland, T. D., Oroszlan, S., Gallo, R. C., and Sarngadharan, M. G. (1988). Biochemical and immunological analysis of human immunodeficiency virus gag gene products p17 and p24. *J. Virol.* 62, 795-801.
- VerPlank, L., Bouamr, F., LaGrassa, T. J., Agresta, B., Kikonyogo, A., Leis, J., and Carter, C. A. (2001). Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55Gag. *Proc. Natl. Acad. Sci. U S A* 98, 7724-7729.

von Schwedler, U. K., Stemmler, T. L., Klishko, V. Y., Li, S., Albertine, K. H., Davis, D. R., and Sundquist, W. I. (1998). Proteolytic refolding of the HIV-1 capsid protein amino-terminus facilitates viral core assembly. *EMBO J.* *17*, 1555-1568.

Wilk, T., Gross, I., Gowen, B. E., Rutten, T., de Haas, F., Welker, R., Krausslich, H. G., Boulanger, P., and Fuller, S. D. (2001). Organization of immature human immunodeficiency virus type 1. *J. Virol.* *75*, 759-771.

Wittekind, M. (1993). HNCACB, A HIGH-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha and beta carbon resonances in proteins. *J. Magn. Reson. B.* *101*, 201-205.

Yu, G., Shen, F. S., Sturch, S., Aquino, A., Glazer, R. I., and Felsted, R. L. (1995). Regulation of HIV-1 gag protein subcellular targeting by protein kinase C. *J. Biol. Chem.* *270*, 4792-4796.

Yuan, X., Yu, X., Lee, T. H., and Essex, M. (1993). Mutations in the N-terminal region of human immunodeficiency virus type 1 matrix protein block intracellular transport of the Gag precursor. *J. Virol.* *67*, 6387-6394.

Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994). Backbone <sup>1</sup>H and <sup>15</sup>N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR* *4*, 845-858.

Zhang, W. H., Hockley, D. J., Nermut, M. V., Morikawa, Y., and Jones, I. M. (1996). Gag-Gag interactions in the C-terminal domain of human immunodeficiency virus type 1 p24 capsid antigen are essential for Gag particle assembly. *J. Gen. Virol.* *77*, 743-751.

Zhao, L., Helms, J. B., Brugger, B., Harter, C., Martoglio, B., Graf, R., Brunner, J., and Wieland, F. T. (1997). Direct and GTP-dependent interaction of ADP ribosylation factor 1 with coatamer subunit beta. *Proc. Natl. Acad. Sci. U S A* *94*, 4418-4423.

Zhou, W., Parent, L. J., Wills, J. W., and Resh, M. D. (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J. Virol.* *68*, 2556-2569.

Zhou, W., and Resh, M. D. (1996). Differential membrane binding of the human immunodeficiency virus type 1 matrix protein. *J. Virol.* *70*, 8540-8548.

Zimmerman, C., Klein, K. C., Kiser, P. K., Singh, A. R., Firestein, B. L., Riba, S. C., and Lingappa, J. R. (2002). Identification of a host protein essential for assembly of immature HIV-1 capsids. *Nature* *415*, 88-92.

Zozulya, S., and Stryer, L. (1992). Calcium-myristoyl protein switch. *Proc. Natl. Acad. Sci. U S A* *89*, 11569-11573.

## CHAPTER 3

### STRUCTURE AND UBIQUITIN BINDING OF THE UBIQUITIN INTERACTING MOTIF

Robert D. Fisher<sup>1</sup>, Bin Wang<sup>1</sup>, Steven L. Alam<sup>1</sup>, Dan Higginson<sup>1</sup>, Rebecca Rich<sup>2</sup>, David Myszka<sup>2</sup>, Wesley I. Sundquist<sup>1\*</sup>, and Christopher P. Hill<sup>1\*</sup>

<sup>1</sup>*Department of Biochemistry, University of Utah, Salt Lake City, UT 84132 USA and*

<sup>2</sup>*University of Utah Protein Interactions Core Facility*

\*Corresponding authors.

## Abstract

Ubiquitylation is used to target proteins into a large number of different biological processes including proteasomal degradation, endocytosis, virus budding, and vacuolar protein sorting (Vps). Ubiquitylated proteins are typically recognized using one of several different conserved ubiquitin binding modules. Here, we report the crystal structure and ubiquitin binding properties of one such module, the Ubiquitin Interacting Motif (UIM). UIM peptides from proteins involved in endocytosis and vacuolar protein sorting including Hrs, Vps27p, Stam1, and Eps15 bound specifically, but with modest affinity ( $K_d=0.1-1$  mM), to free ubiquitin. The UIM peptides all bound to the “144” surface of ubiquitin as analyzed by NMR chemical shift perturbation mapping experiments. Full affinity ubiquitin binding required the presence of conserved acidic patches at the N- and C-terminus of the UIM, as well as highly conserved central alanine and serine residues. The 1.7 Å resolution crystal structure of the second yeast Vps27p UIM (Vps27p-2) revealed that the ubiquitin interacting motif forms an amphipathic helix. Although Vps27p-2 is monomeric in solution, the motif unexpectedly crystallized as an antiparallel four helix bundle, and the potential biological implications of UIM oligomerization are therefore discussed.

## Introduction

Ubiquitylation plays a critical role in maintaining protein homeostasis in the cell because it serves to target proteins for both proteasomal and lysosomal degradation (Dupre et al., 2001; Hershko and Ciechanover, 1998; Hicke, 1999; Katzmann et al., 2002; Pickart, 2001a; Varshavsky et al., 2000). It has recently become clear that protein ubiquitylation also plays important roles in a large number of other biological processes,

including DNA repair, transcription, translation, signal transduction, organelle assembly, protein trafficking, and virus budding (Ben-Neriah, 2002; Conaway et al., 2002; Hicke, 2001; Jason et al., 2002; Pickart, 2001b). Thus, it is of general importance to understand how cells recognize and sort ubiquitylated proteins.

The recognition of ubiquitylated proteins is frequently mediated by one of several conserved ubiquitin binding modules, which include the UIM (Ubiquitin Interacting Motif), UBA (Ubiquitin-Associated domain), UEV (Ubiquitin E2 Variant/UBC-like domain) (Buchberger, 2002; Hofmann and Bucher, 1996; Hofmann and Falquet, 2001; Madura, 2002; Pickart, 2001a), NZF (Np14 Zinc Finger domain) (Meyer et al., 2002), and CUE (Coupling of Ubiquitin Conjugation to ER Degradation domain) (Ponting, 2000). Each of these motifs forms an independent folding domain that can bind ubiquitin in vitro, and the motifs are used in a modular fashion to add ubiquitin binding activities to a large variety of multifunctional proteins (Buchberger, 2002). Thus, understanding how ubiquitylated proteins are recognized and sorted within cells will require a full description of the structures and ubiquitin interactions of the different conserved ubiquitin binding modules.

The UIM motif was originally identified based upon studies of the S5a subunit of the 19S regulator in the human 26S proteasome (hS5a). Biochemical and mutational analyses have revealed that hS5a contains two copies of an ~30 residue sequence motif (initially denoted pUbS) that can bind both ubiquitylated proteins as well as poly(Ub) chains (Beal et al., 1996; Deveraux et al., 1994; van Nocker et al., 1996; Young et al., 1998). The pUbS motifs from hS5a have hydrophobic core sequences composed of alternating large and small residues (Leu-Ala-Leu-Ala-Leu), which are flanked on both

sides by patches of acidic residues. Similar ubiquitin binding sequences are found in the S5a proteins of all higher eukaryotes. Sequence analyses and iterative database searches based upon the original pUBS motif have been used to define a more general ubiquitin interacting motif (UIM), which is found in a number of different proteins that function in a variety of biological pathways (Hofmann and Falquet, 2001). These sequence analyses have also provided a more precise definition of the UIM as a 20 residue sequence corresponding to the consensus: X-**Ac-Ac-Ac-X-Φ-X-X-Ala-X-X-X-Ser-X-X-Ac-X-X-X-X**, where **Φ** represents a large hydrophobic residue (typically Leu), **Ac** represents an acidic residue (Glu, Asp), and **X** represents residues that are less well conserved (Hofmann and Falquet, 2001).

UIM motifs are particularly prevalent in proteins that function in the pathways of endocytosis and vacuolar protein sorting (Hofmann and Falquet, 2001). These two linked pathways serve to sort membrane-associated proteins and their cargos from the plasma membrane (or Golgi) for eventual destruction (or localization) in the lysosome (yeast vacuole). Unlike proteasomal protein targeting, which requires at least a tetraubiquitin chain (Thrower et al., 2000), monoubiquitylation is sufficient to mark proteins for both endocytosis and lysosomal trafficking (Katzmann et al., 2001; Polo et al., 2002; Raiborg et al., 2002; Reggiori and Pelham, 2001; Shih et al., 2002; Shih et al., 2000; Terrell et al., 1998; Urbanowski and Piper, 2001).

Endocytic proteins that contain UIM motifs include the epsins (yeast Ent1p, Ent2p), Eps15, and Eps15R (yeast Ede1p) (Hofmann and Falquet, 2001). These proteins are all required for endocytosis of receptor/ligand complexes like the one formed by the epidermal growth factor (EGF) with its receptor (EGFR) (Chen et al., 1998; Shih et al.,

2002). Recent work from several laboratories has demonstrated that the UIM motifs in these proteins can bind ubiquitin *in vitro* and play essential roles *in vivo*, as deletion or mutation of the UIM sequences blocks receptor internalization (Polo et al., 2002; Shih et al., 2002). Upon EGFR stimulation, the EGFR is ubiquitylated, and the Eps15 and Eps15R proteins are also phosphorylated and monoubiquitylated (Polo et al., 2002; van Delft et al., 1997; Wiley and Burke, 2001). Strikingly, Polo and coworkers found that mutating either of the two UIM motifs found in Eps15 prevented Eps15 monoubiquitylation (Polo et al., 2002). Thus, it appears that UIM motifs can both bind ubiquitin and also direct protein ubiquitylation, although the relationship between these two activities is not yet fully understood.

Upon internalization, ubiquitylated receptors can then be sorted through the endosomal system to the lysosome via the vacuolar protein sorting pathway (reviewed in Katzmann et al., 2002). UIM-containing proteins required for vacuolar protein sorting include Hrs (yeast Vps27p), Stam1 (yeast Hse1), and Stam2 (Hofmann and Falquet, 2001). There appear to be strong parallels between the requirements for endocytosis and vacuolar protein sorting because the protein substrates of both pathways are ubiquitylated, the UIM domains of proteins in the pathway are required for proper sorting of substrates, and because Hrs, like Eps15, also becomes ubiquitylated in a process that again depends upon the integrity of its own UIM motif (Bilodeau et al., 2002; Polo et al., 2002; Raiborg et al., 2002; Shih et al., 2002). Interestingly, vesicle formation during cellular vacuolar protein sorting also appears to be intimately related to another ubiquitin-dependent process; the budding of many enveloped viruses (reviewed in Pornillos et al., 2002b). For example, we have recently shown that Hrs protein fragments, when fused to

the C-terminal end of the structural HIV-1 Gag protein, can rescue the budding of virus-like particles that lack their cis-acting signals normally required for efficient virus budding (Higginson et al. in preparation).

The prevalence of the UIM domain and its important role in the monoubiquitin-dependent processes of endocytosis and vacuolar protein sorting has led us, and others (Shekhtman and Cowburn, 2002; Walters et al., 2002), to study the detailed biochemical and structural basis for Ub/UIM interactions. Toward this end, we have characterized the ubiquitin binding properties of UIM peptides from Eps15, Vps27p, Hrs, Stam-1, and Stam-2, and determined the high resolution crystal structure of the second UIM motif from Vps27p.

## **Experimental Procedures**

### Protein preparation

To create GST-UIM fusion proteins for biosensor binding experiments, complementary synthetic oligodeoxynucleotides (5' phosphorylated) encoding UIM peptides were heat annealed and cloned into the complementary NdeI/BamHI sites of vector WISP94-18 (Yoo et al., 1997). WISP94-18 has been modified from the parental pGEX-2T vector (Pharmacia Amersham Biosciences) to create these cloning sites and allows expression of the UIM (and other) peptides as fusions at the C-terminal end of glutathione *S*-transferase (GST). Sequences and cloning constructs are given in Table 3.1, and the final constructs were all confirmed by DNA sequencing.

GST-UIM fusion peptides were expressed in DH5 $\alpha$  E.coli cells. Protein expression was induced with 0.5 mM IPTG (OD<sub>600</sub>=0.4), and after 4 h at 23 °C the cells were harvested by centrifugation (4 °C), resuspended in lysis buffer (20 mM Tris:HCl pH

Table 3.1. UIM constructs and cloning

UIM peptide	SwissProt Database Entry <sup>a</sup>	Plasmid (WISP) and oligodeoxynucleotide (WISO) designations	Oligodeoxynucleotide Sequence (sense strand)
Hrs-A	O14964	WISP 02-6 WISO 02-3, 02-4	5' TATGCAGGAAGAAGAAGAACTGCAGCTG GCGCTGGCGCTGAGCCAGAGCGAAGCGGAA GAAAAATAAG3'
Hrs-B	O14964	WISP 02-5 WISO 02-1, 02-2	5' TATGAAACGCGATGAAACCGCGCTGCAG GAAGAAGAAGAACTGCAGCTGGCGCTGGCG CTGAGCCAGAGCGAAGCGGAAGAAAAAGAA CGCCTGCGCCAGAAAAGCACCTATTAAG3'
Vps27p -1	P40343	WISP 02-61 WISO 02-199, 02-199	5' TATGGATGAAGAAGAACTGATCCGCAAA GCGATCGAACTGAGCCTGAAAGAAAGCCGC AACAGCTAAG3'
Vps27p -2	P40343	WISP 02-62 WISO 02-200, 02-201	5' TATGGAAGAAGATCCCGATCTGAAAGCG GCGATCCAGGAAAGCCTGCGCGAAGCGGAA GAAGCGTAAG3'
STAM 1-A	Q92783	WISP 02-9 WISO 02-9, 02-10	5' TATGCAGGAAGAAGAAGAACTGCAGCTG GCGCTGGCGCTGAGCCAGAGCGAAGCGGAA GAAAAATAAG3'
STAM 1-B	Q92783	WISP 02-39 WISO 02-133, 02-134	5' TATGACCGTGGCGAACAAAAAGAAGAA GAAGATCTGGCGAAAGCGATCGAACTGAGC CTGAAAGAACAGCGCCAGCAGAGCACCACC CTGAGCACCTAAG3'
STAM 2-B	075886	WISP 02-40 WISO 02-135, 02-136	5' TATGAGCAGCAACAAAAACAAAGAAGAT GAAGATATCGCGAAAGCGATCGAACTGAGC CTGCAGGAACAGAAAACAGCAGCATACCGAA ACAAAAAGCCTGTATCCGAGCTAAG3'
Eps15- 1	P42567	WISP 02-10 WISO 02-11, 02-12	5' TATGAGCGAAGAAGATATGATCGAATGG GCGAAACGCGAAAGCGAAGAAGAACAGCGC CAGCGCTAAG3'
Eps15- 2	P42567	WISP 02-11 WISO 02-13, 02-14	5' TATGCAGGAACAGGAAGATCTGGAAGT GCGATCGCGCTGAGCAAAAGCGAAATCAGC CGCGCGTAAG3'
Hrs-B 259EE EE/AA AA	O14964	WISP 02-38 WISO 02-131, 02-132	5' TATGAAACGCGATGAAACCGCGCTGCAG GCGGCGGCGGCGCTGCAGCTGGCGCTGGCG CTGAGCCAGAGCGAAGCGGAAGAAAAAGAA CGCCTGCGCCAGAAAAGCACCTATTAAG3'
Hrs-B 266A/G	O14964	WISP 02-37 WISO 02-129, 02-130	5' TATGAAACGCGATGAAACCGCGCTGCAG GAAGAAGAAGAACTGCAGCTGGGCTGGCG CTGAGCCAGAGCGAAGCGGAAGAAAAAGAA CGCCTGCGCCAGAAAAGCACCTATTAAG3'
Hrs-B 270S/A	O14964	WISP 02-35 WISO 02-119, 02-120	5' TATGCAGGAAGAAGAAGAACTGCAGCTG GCGCTGGCGCTGGCGCAGAGCGAAGCGGAA GAAAAATAAG3'
Hrs-B 273E/A	O14964	WISP 02-36 WISO 02-121, 02-122	5' TATGCAGGAAGAAGAAGAACTGCAGCTG GCGCTGGCGCTGAGCCAGAGCGCAGCGGAA GAAAAATAAG3'

<sup>a</sup>SwissProt database entry gives the full sequence of the relevant protein

8.0, 150 mM NaCl 5 mM  $\beta$ -mercaptoethanol, supplemented with Complete Protease Inhibitor Tablets (Roche)), lysed with lysozyme (20  $\mu$ g/mL) and sonication, and the supernatants were clarified by centrifugation at 40,000xg for 30 min. The resulting soluble extracts were held on ice and used immediately to minimize protein degradation.

UIM peptides used in NMR chemical shift perturbation and crystallization experiments were made by Fmoc solid-phase synthesis, purified using reverse phase high performance liquid chromatography, and confirmed by MALDI mass spectrometry.

Ubiquitin was expressed and purified as described (Beal et al., 1996).  $^{15}$ N labeled ubiquitin was expressed and purified using same protocol except that it is expressed in M9 minimal media with  $^{15}$ NH<sub>4</sub>Cl as the sole source of nitrogen.

#### Biosensor binding experiments

Biosensor binding experiments were performed at 10 and 18 °C on a BIACORE 3000 using a CM5 research-grade sensor chip derivatized with anti-GST antibodies (Yoo et al., 1997). GST-UIM fusion peptides or GST alone (negative control) were captured from soluble E.coli lysates at final densities of 1.1-2.3 kRU. Ubiquitin in running buffers (see figure captions) was flowed over the GST-UIM and GST surfaces at concentrations of 0, 4.1, 12.3, 37.0, 111, 333, and 1000  $\mu$ M. Binding responses were recorded and globally fit to simple 1:1 binding models using the CLAMP software (Myszka and Morton, 1998).

#### Chemical shift perturbation mapping experiments

Chemical shift perturbation experiments were performed at 18 °C on a Varian Inova 600 MHz spectrometer equipped with a triple-resonance  $^1$ H/ $^{13}$ C/ $^{15}$ N probe and z-

axis pulsed field gradient capability. Both peptide and Ub samples were dissolved in NMR buffer (90%  $^1\text{H}_2\text{O}/10\%$   $^2\text{H}_2\text{O}$  containing 20 mM sodium phosphate pH 6.0, 10 mM NaCl), and the unlabeled UIM peptides ( $\sim 10$  mM) were titrated into 0.8 mM  $^{15}\text{N}$ -labeled Ub at final molar ratios of 0:1, 0.25:1, 0.5:1, and 1:1 (UIM:Ub).  $^1\text{H}/^{15}\text{N}$ -HSQC spectra (Mori et al., 1995) were collected at the different titration points, processed using FELIX 97 (MSI), and analyzed using SPARKY (T.D. Goddard and D.G. Kneller, University of California, San Francisco). Normalized chemical shift changes ( $\delta$ ) were calculated using the equation:  $\delta = 25[(\delta_{\text{HN}})^2 + (\delta_{\text{N}}/5)^2]^{0.5}$  (Cheever et al., 2001). Amide chemical shift assignments for human ubiquitin were obtained from the VLI Research, Inc. website <http://www.vli-research.com/ubshifts.htm>.

#### Vps27p crystallization

Vps27p-2 UIM was crystallized in two slightly different conditions using the sitting drop vapor diffusion method at 23 °C. Crystal form I was grown by mixing 3  $\mu\text{l}$  of protein solution containing 0.8 mM selenomethionine substituted ubiquitin (50 mM Tris pH 7.5, 150 mM NaCl) and 1.6 mM Vps27p-2 with 2  $\mu\text{l}$  of reservoir solution (0.1M imidazole pH 7.6, 0.2 M zinc acetate, and 26% 1,4-butanediol). Crystal form II grew from the same protein solution as form I when 3  $\mu\text{l}$  of protein were mixed with 2  $\mu\text{l}$  of a reservoir solution comprised of 0.08 M sodium cacodylate pH 6.5, 0.16 M zinc acetate, 10.4% PEG-8000, and 20% glycerol. Both forms crystallized in space-group  $P6_222$  with essentially identical unit cell dimensions  $a = b = 34.9 \text{ \AA}$ ,  $c = 64.2 \text{ \AA}$ , indicating the two crystal forms are isomorphous. After solving the structure and finding that the crystals contained only Vps27p-2 peptide, we also confirmed experimentally that the same crystals could also be grown in the absence of ubiquitin (not shown).

### Structure determination

Data were collected from single crystals of each of crystal form. Crystals were suspended in a nylon loop, rapidly cooled by plunging into liquid nitrogen, and maintained at 100 K for data collection. Prior to cooling, the form I crystal was soaked in reservoir solution made up with 15% glycerol. Form I crystal data were collected at the Brookhaven synchrotron (National Synchrotron Light Source) at a wavelength of 0.979 Å whereas the form II crystal data were collected on our home rotating anode x-ray source ( $\lambda = 1.5418$  Å). All data were integrated and scaled in the HKL package (Otwinowski and Minor, 1987).

Data collected from the form I crystal showed a large anomalous signal, and the program SOLVE (Terwilliger, 2002) was used to determine phases and locations for three zinc atoms using the single-wavelength anomalous diffraction (SAD) method. Phases were refined and a polyaniline helix model was calculated in RESOLVE (Terwilliger, 2002). The resulting 1.7 Å maps showed clear continuous density that allowed for the modeling of a 20-residue  $\alpha$ -helix. Model building was performed in the program O (Jones et al., 1991), and successive rounds of structure refinement were completed in REFMAC5 (Murshudov et al., 1999), as provided in the CCP4 suite (Project, November 4, 1994), against the form II crystal data set (which was of higher overall quality). Figures of the Vps27p-2 UIM crystal structure were created in MOLSCRIPT (Esnouf, 1997) and PyMOL (DeLano Scientific).

### Analytical ultracentrifugation

Equilibrium ultracentrifugation experiments were performed on a Beckman Optima XL-A analytical ultracentrifuge. Protein samples of Vps27p-2 UIM (60-220  $\mu$ M)

in 50 mM Tris pH 7.5, 150 mM NaCl were centrifuged at 20 °C at a rotor speed of 52,000 rpm. Absorbance measurements were recorded at 230 nm at 0.001 cm intervals every four hours, until equilibrium was achieved, as judged by a stable protein distribution. Absorbance data were averaged and corrected for background absorbance against a buffer blank.

Equilibrium data from three different concentration distributions were simultaneously fit to a single homogenous monomer species model (McRorie and Voelker, 1993) using non-linear least squares techniques and the analysis program NONLIN (Johnson et al., 1981). For these calculations, the partial specific volume was derived from the Vps27p-2 UIM sequence and estimated to be 0.7143 ml g<sup>-1</sup> at 20 °C (Laue et al., 1992).

## **Results and Discussion**

### Ubiquitin binding by the UIM

We initially surveyed ubiquitin binding to six peptides corresponding to minimal UIM sequences from the proteins Hrs, Vps27p (two motifs), Stam1, and Eps15 (two motifs) (Hofmann and Falquet, 2001). Binding was tested at neutral pH (7.0) and low salt (10 mM NaCl) and was quantified using biosensor experiments in which pure recombinant ubiquitin was allowed to bind to immobilized GST-UIM fusion peptides (Table 3.2 and Figure 3.1). Ubiquitin bound to the GST-UIM surfaces with rapid, reversible kinetics and the interaction was specific as ubiquitin did not bind to a control GST surface (Figure 3.1A inset, and data not shown). However, ubiquitin binding was generally weak, varying between a  $K_d$  of ~400  $\mu$ M (for both Vps27p UIM peptides) to undetectable binding (for the second UIM from Eps15-2).

Table 3.2. Ubiquitin binding affinities of various UIM peptides

UIM Peptide	Peptide Sequence <sup>a</sup>	K <sub>d</sub> (μM) <sup>b</sup>
Hrs-A	258 <b>QEEEEELQLALALSQSEAEK</b>	1500(50) <sup>c</sup>
Hrs-B	251 KRDETAL <b>QEEEEELQLALALSQSEAEK</b> ERLRQKSTY	252 ± 18 <sup>d</sup>
Vps27p-1	258 <b>DEEELIRKAIELSLKESRNS</b>	397(7) <sup>c</sup>
Vps27p-2	301 <b>EEDPDLKAAIQESLREAEAA</b>	437(4) <sup>c</sup>
STAM1-A	171 <b>KEEEDLAKAIELSLKEQRQQ</b>	691(31) <sup>c</sup>
STAM1-B	166 TVANK <b>KEEEDLAKAIELSLKEQRQQ</b> STTLST	281(2) <sup>c</sup>
STAM2-A	165 <b>KEDEDIAKAIELSLQEQQQQ</b>	ND <sup>e</sup>
STAM2-B	159 MSSNKN <b>KEDEDIAKAIELSLQEQQQQ</b> HTETKSLYPS	179(2) <sup>c</sup>
Eps15-1	852 <b>SEEDMIEWAKRESEREQQQR</b>	6500(900) <sup>c</sup>
Eps15-2	878 <b>QEQQEDLELAIALSKSEISEA</b>	>10,000 <sup>f</sup>

<sup>a</sup>The position of each peptide within the relevant protein is given, and the "consensus" UIM sequences are shown in bold, after Hofmann and Falquet, 2001.

<sup>b</sup>K<sub>d</sub> values were determined in BIACORE biosensor measurements of soluble Ub binding to immobilized GST-UIM fusion peptide at 18 °C in a buffer of 25 mM Tris·HCl pH 7.0, 10 mM NaCl, 0.005% P20, 50 mg/mL BSA.

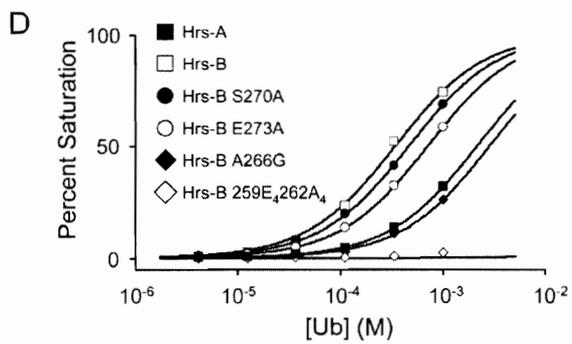
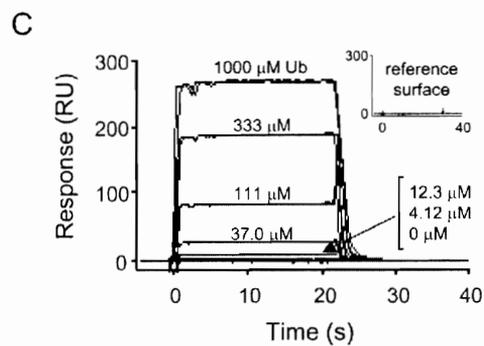
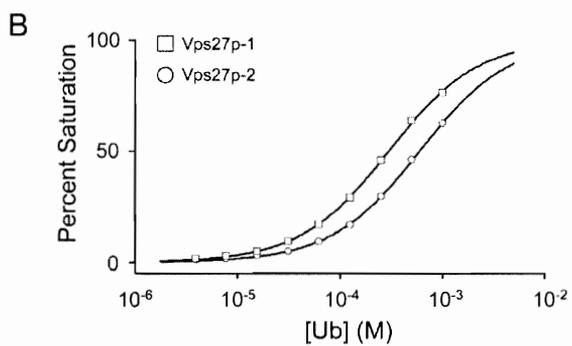
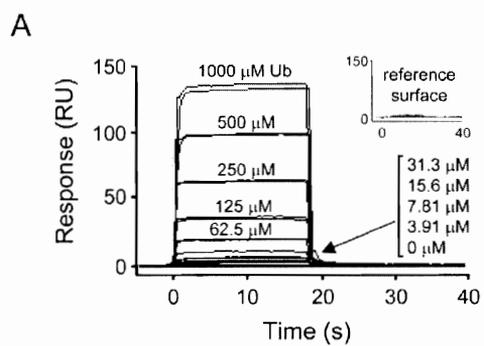
<sup>c</sup>Binding constant and error were estimated from a statistical fit of a single binding isotherm with duplicate measurements at each free ubiquitin concentration. Values in parentheses are statistical errors estimated from the fit of the data to a simple 1:1 binding model.

<sup>d</sup>Binding constant is the mean of two completely independent measurements, and the error is the range in the two measurements.

<sup>e</sup>Not measured

<sup>f</sup>Less than 1% saturation binding at 1 mM free ubiquitin protein concentration.

Figure 3.1. Quantification of ubiquitin/UIM peptide interactions. (A) Surface plasmon resonance biosensor analysis of the Ub/Vps27p-2 interaction. Free ubiquitin was injected in duplicate at the indicated concentrations over GST-Vps27p-2 captured on an anti-GST surface. The inset (negative control) shows the response obtained for 1000  $\mu$ M ubiquitin injected over recombinant GST alone captured on an anti-GST surface. (B) Isotherms for ubiquitin binding to GST-Vps27p-1 (open squares,  $K_d=247\pm 2$   $\mu$ M), and GST-Vps27p-2 (open circles,  $K_d=598\pm 7$   $\mu$ M). (C) Biosensor data for the Ub/Hrs-B interaction. Ubiquitin was injected in duplicate at the indicated concentrations GST-Hrs-B captured on an anti-GST surface. The inset depicts the same negative control as in A. (D) Isotherms for ubiquitin binding to wt and mutant GST-Hrs peptides. Peptide designations are indicated within the figure, and sequences and estimated binding affinities are given in Table 3. All of the binding experiments shown here were performed at 18 °C in a buffer of 20 mM sodium phosphate pH 6.0, 10 mM NaCl, 0.005% P20, 50 mg/mL BSA.



The effect of different solution conditions on the affinity of the Ub/UIM interactions was also surveyed. Ubiquitin binding to each UIM construct was quantified at three different pHs (6.0, 7.0, and 8.0), two different salt conditions (10 and 75 mM NaCl), and two different temperatures (10 and 18 °C). In general, Ub/UIM interactions were slightly tighter at lower temperatures, relatively insensitive to pH, and somewhat weaker at higher salt, although exceptions to each of these trends were noted (data not shown). As solution conditions had no remarkable effects on the interaction, subsequent mutational analyses and NMR studies were performed under low salt and pH conditions (pH 6.0, 10 mM NaCl), which are ideal for NMR spectroscopic studies. Representative binding data for the Vps27p-2/Ub interaction under these conditions are shown in Figure 3.1A. Binding isotherms and fits to simple 1:1 binding models for the Vps27p-2/Ub and Vps27p-1/Ub interactions are shown in Figure 3.1B.

As the measured ubiquitin binding affinities for the different minimal UIM peptides were surprisingly weak, we considered the possibility that adding flanking residues to either side of the minimal UIM sequences might be necessary to complete the motif and/or to negate “end” effects. This possibility was tested with using longer constructs corresponding to predicted helices that encompassed the minimal UIM sequences from human Hrs, Stam1, and Stam2 (denoted Hrs-B, Stam1-B, and Stam2-B, respectively). These three proteins were chosen for further study as we are particularly interested in the roles of ubiquitin in vacuolar protein sorting and HIV budding. Representative ubiquitin binding data for the shorter (Hrs-A) and longer (Hrs-B) constructs are shown in Figures 3.1C, D; and binding data for all of the longer constructs are summarized in Table 3.2. Notably, the addition of 5-9 extra residues on either side of

the minimal UIM motifs of Hrs and Stam-1 increased ubiquitin binding affinity 2 to 6-fold (the comparison was not made for Stam-2). Moreover, all three of the longer UIM constructs bound ubiquitin with moderate (and similar) affinities ( $K_d = 150\text{-}300 \mu\text{M}$ ). We therefore conclude that the residues that normally flank minimal UIM motifs can enhance ubiquitin binding *in vitro*.

Several other groups have recently reported that isolated UIM motifs can bind specifically to free ubiquitin (Polo et al., 2002; Raiborg et al., 2002; Shekhtman and Cowburn, 2002; Walters et al., 2002), and our experimental results provide further confirmation of this fact. Moreover, two groups have previously quantified the ubiquitin binding affinities of different UIM-containing Hrs constructs, and all of the results are in good accord. Specifically, Shekhtman and Coburn used NMR chemical shift titrations to estimate that ubiquitin bound with a dissociation constant of  $230 \pm 50 \mu\text{M}$  to an UIM peptide corresponding to Hrs residues 257-278, and Raiborg et al. performed biosensor experiments to show that ubiquitin bound to an immobilized Hrs fragment (residues 1-289) with a dissociation constant of  $300 \mu\text{M}$ . Our estimates of the Hrs-B/Ub dissociation constant under low salt conditions agree very well with these published estimates ( $250\text{-}290 \mu\text{M}$  depending upon the pH, see Tables 3.2 and 3.3). Furthermore, the fact the Hrs-B, Stam1-B and Stam2-B UIM peptides all bound ubiquitin with similar affinities in spite of significant sequence variation at the nonconserved UIM positions suggests that canonical UIM motifs can be expected to bind ubiquitin with dissociation constants in the  $150\text{-}300 \mu\text{M}$  range. Hrs, Stam-1, and Stam-2 normally form a complex *in vivo* (Asao et al., 1997; Raiborg et al., 2001), and our results indicate that this complex has the potential to display (at least) three active UIM sequences simultaneously.

### Role of conserved UIM residues in ubiquitin binding

The relative contributions of different UIM sequence elements to ubiquitin binding affinity were examined in mutational studies of the Hrs-B UIM peptide. Guided by the initial description of sequence conservation in the UIM (Hofmann and Falquet, 2001), we selected the four most conserved sequence elements for mutagenesis and ubiquitin binding studies. These were: 1) the N-proximal acidic patch (an <sup>259</sup>EEEE<sub>262</sub> to AAAA mutation in Hrs-B, denoted Hrs-B<sub>259</sub>EEEE/AAAA), 2) the nearly invariant alanine residue at UIM position 9 (Hrs-B<sub>266</sub>A/G), and 3) the nearly invariant serine at UIM position 13 (Hrs-B<sub>270</sub>S/A), and the conserved glutamate residue at UIM position 15 (Hrs-B<sub>273</sub>E/A).

As shown in Figure 3.1D and summarized in Table 3.3, all of the mutations in conserved UIM elements reduced the affinity of ubiquitin binding, although the magnitudes of the effects differed significantly. Specifically, mutation of the conserved N-proximal acidic element eliminated all detectable ubiquitin binding, mutation of the nearly invariant alanine at UIM position 9 severely reduced (but did not eliminate) ubiquitin binding (~10-fold reduction), whereas the Hrs-B<sub>273</sub>E/A and Hrs-B<sub>270</sub>S/A mutations reduced ubiquitin binding only 2.5- and 1.5-fold, respectively. The modest reduction in the magnitude of ubiquitin binding for the Hrs-B<sub>270</sub>S/A mutation is somewhat surprising given the very high degree of conservation of this serine (>98% in putative UIM sequences). However, a Ser to Ala substitution is rather conservative and so may be less disruptive than alternative substitutions at this position. Indeed, there is only one predicted naturally occurring UIM motif in which a serine is not found at position 13, and in that case there is also a Ser to Ala substitution (Hofmann and Falquet,

Table 3.3. Ubiquitin binding affinities of mutant Hrs UIM peptides

Hrs-B Mutation	Peptide Sequence <sup>a</sup>	K <sub>d</sub> (μM) <sup>b</sup>
Hrs-B (wt)	251 KRDETAL <b>QEEEEELQLALALSQSEAE</b> EKERLRQKSTY	290±32 <sup>c</sup>
S <sub>270</sub> A	251 KRDETAL <b>QEEEEELQLALAL</b> <u>A</u> QSEAEKERLRQKSTY	465(4) <sup>d</sup>
E <sub>273</sub> A	251 KRDETAL <b>QEEEEELQLALALSQ</b> <u>S</u> AAEAKERLRQKSTY	712(4) <sup>d</sup>
A <sub>266</sub> G	251 KRDETAL <b>QEEEEELQLGLALSQ</b> SEAEKERLRQKSTY	2850(260) <sup>d</sup>
<sub>259</sub> EEEE <sub>262</sub> AAAA	251 KRDETAL <b>QAAAA</b> LQLALALSQSEAEKERLRQKSTY	>10,000 <sup>e</sup>

<sup>a</sup>The position of the peptide within Hrs is given, and the “consensus” UIM sequence is shown in bold, after Hofmann and Falquet, 2001. Mutation sites are underlined.

<sup>b</sup>K<sub>d</sub> values were determined in BIACORE biosensor measurements of soluble Ub binding to immobilized GST-Hrs-B fusion peptide at 18 °C in a buffer of 20 mM sodium phosphate pH 6.0, 10 mM NaCl, 0.005% P20, 50mg/mL BSA.

<sup>c</sup>Binding constant is the mean of two completely independent measurements, and the error is the range in the two measurements.

<sup>d</sup>Binding constant and error were estimated from a statistical fit of a single binding isotherm with duplicate measurements at each free ubiquitin concentration. Values in parentheses are statistical errors estimated from the fit of the data to a simple 1:1 binding model.

<sup>e</sup>Less than 1% saturation binding at 1 mM free ubiquitin protein concentration.

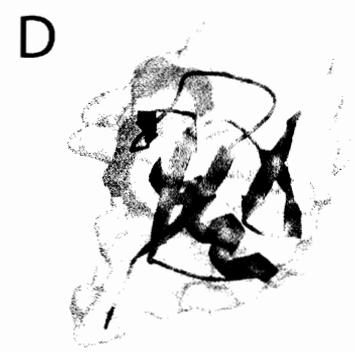
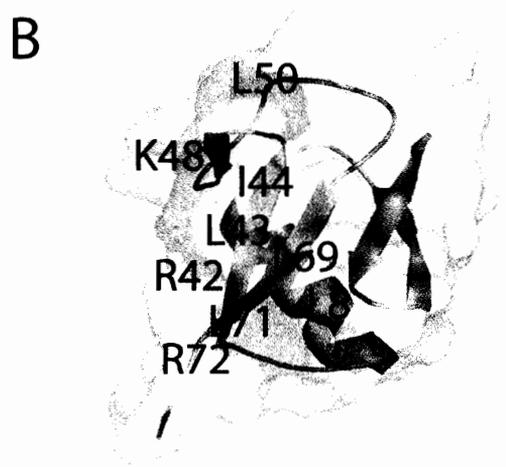
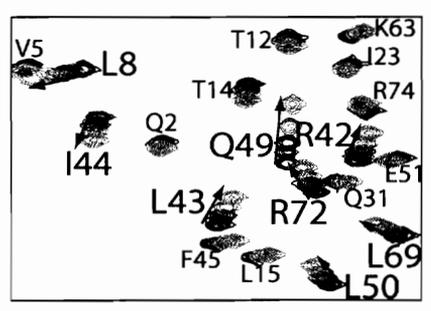
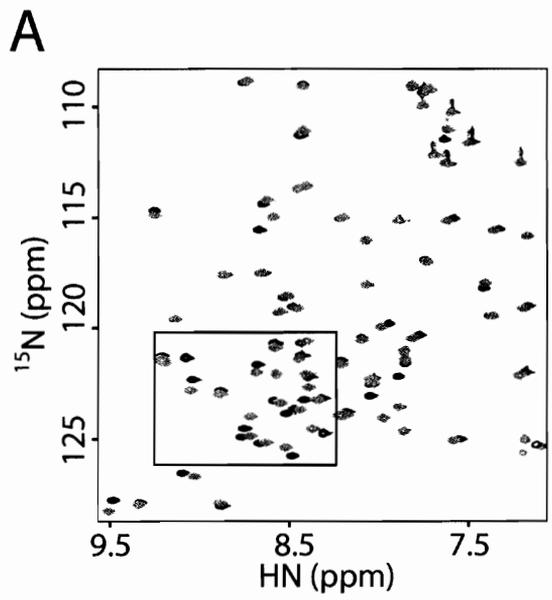
2001). In summary, our experiments strongly support the idea that the different sequence elements of the UIM motif are conserved, at least in part, because they contribute directly or indirectly to the affinity of ubiquitin binding.

Although we are the first to quantify the contributions of different UIM sequence elements to ubiquitin binding, others have tested the functional importance of several conserved UIM residues (Polo et al., 2002; Raiborg et al., 2002; Shih et al., 2002). Not surprisingly, these studies all support the idea that conserved UIM residues play functionally important roles. The experiments that are most directly relevant to ours demonstrated that mutation of the conserved serine residue at UIM position 13 (to either Asp or Glu) eliminated ubiquitin binding to Hrs and Vps27p UIM peptides in “GST-pulldown” experiments *in vitro*, and also blocked the functional sorting of ubiquitylated transferrin receptors by Hrs (Raiborg et al., 2002) and of carboxypeptidase S by Vps27p (Shih et al., 2002) *in vivo*. We are not aware of any other functional tests of single UIM point mutations, but our binding experiments make the strong prediction that mutation of either of the two conserved acidic patches or of the conserved Ala residue at UIM position 9 should also result in loss of UIM function.

#### The ubiquitin binding site of UIM peptides

$^1\text{H}/^{15}\text{N}$  NMR chemical shift perturbation experiments were used to map the ubiquitin binding sites for the Hrs-A, Hrs-B, Vps27-1, Vps27p-2, Stam1-A, and Stam1-B UIM peptides (Figure 3.2 and data not shown). Representative data showing the changes in the  $^1\text{H}/^{15}\text{N}$ -HSQC spectra of ubiquitin upon titration of 0-1 equivalent Hrs-A are provided in Figure 3.2A. 14/70 observable Ub backbone amide resonances shifted

Figure 3.2. NMR chemical shift mapping of the interaction surfaces of ubiquitin with different UIM motifs. (A) Overlaid  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectra of ubiquitin (0.8 mM) in the presence of 0 (blue), and 1 (gray) molar equivalents Hrs-A. The boxed insert in the left spectrum is expanded at right, and also includes intermediate titration points at: 0.125 (magenta), 0.25 (red), 0.5 (green) equivalents Hrs-A. (B-E) Residues with the greatest chemical shift changes upon addition of 1 equivalents of the UIM peptides: Hrs-A (B), Hrs-B (C), Vps27p-2 (D), and Stam1-A (E) are shown mapped onto a surface/ribbon representation of ubiquitin. Residues shifted by  $\delta \geq 2$  are colored using a gradient scheme from red ( $\delta=8$ , most shifted) to pink ( $\delta=2$ ). The binding surfaces of Vps27p-1 and Stam1-B were also mapped and were very similar (not shown). (F) For comparison, the residues that compose the “I44” functional surface of ubiquitin, as mapped by alanine scanning mutagenesis(40), are shown in blue.



significantly upon Hrs-A binding ( $\delta \geq 2$ ), and the positions of the shifted residues are shown mapped onto the structure of ubiquitin in Figure 3.2B.

As summarized in Figures 3.2B-E, all six of the UIM peptides that we tested behaved very similarly. All of the UIM peptides bound in fast exchange and shifted a very similar set of ubiquitin resonances. These shifted residues generally clustered about the C-terminal three-strands of the Ub  $\beta$ -sheet, and this is presumably the UIM binding surface. This surface corresponds very closely to the “I44 surface” of Ub (Figure 3.2F), which has previously been shown to function in endocytosis, proteasomal degradation, and HIV budding (Sloper-Mould et al., 2001; Strack et al., 2002). This concave surface of ubiquitin is quite hydrophobic, and includes the exposed side chains of L8, L43, I44, L50, L69, and L71, which could presumably interact with a complementary hydrophobic surface on the UIM (discussed below). This surface of ubiquitin also displays three basic side chains R42, K48, R72 and we speculate that these may interact with the two conserved acidic patches within the UIM motif. Interestingly, the K48 side chain is the site that is used to make poly(Ub) chains that target proteins for proteasomal degradation. The footprint of the different UIM motifs appear to encompass this site (see Figure 3.2 and Shekhtman and Cowburn, 2002). We therefore anticipate that covalent attachment of additional ubiquitin moieties to this site could have a significant effect on UIM binding. Indeed, there is some evidence that isolated UIM motifs can bind more tightly to poly(Ub) chains than to monomeric ubiquitin molecules (Beal et al., 1996; Polo et al., 2002; Young et al., 1998), although it is difficult to rule out possible avidity effects and/or multiple Ub/UIM contact sites in the binding experiments reported to date.

Our experiments are in excellent agreement with two recently published chemical shift mapping studies of the ubiquitin binding sites of another Hrs UIM peptide (corresponding to Hrs residues 257-278) (Shekhtman and Cowburn, 2002) and of the hS5a protein (which contains two UIM sequences) (Walters et al., 2002). In both cases, the mapped UIM binding sites on ubiquitin correspond very closely to the site that we mapped for the six UIM peptides that we studied. This agreement lends further credence to the idea that all UIM motifs adopt very similar structures and bind ubiquitin in very similar ways.

Our UIM mapping studies are also consistent with the observation that an I44A mutation in ubiquitin abolished the binding of UIM peptides from Ent1p and Vps27p as assayed by co-affinity purification, whereas an F44A mutation in ubiquitin had no effect on UIM binding (Shih et al., 2002). As shown in Figure 3.2B, the ubiquitin I44 residue is in the center of our mapped UIM binding sites, whereas the F4 residue is located on the opposite side of the molecule (not shown). The fact that multiple UIMs from proteins that function in the same biological pathway can bind on the same surface of ubiquitin suggests the possibility that they may compete for binding and that this could provide a mechanism for “passing” ubiquitylated substrates along a pathway. Furthermore, not only do UIM sequences bind to this surface of ubiquitin, but both of the other ubiquitin binding motifs whose ubiquitin interaction surfaces have been characterized (UBA, and NZF) also bind to this same surface (Bertolaet et al., 2001; Mueller and Feigon, 2002; Pornillos et al., 2002a). Thus, it seems likely that competition for binding to this functionally critical surface of ubiquitin must play an important role in the recognition and trafficking of ubiquitylated proteins.

### Structure of the Vps27p-2 UIM

In an attempt to characterize the UIM in structural detail, we surveyed crystallization conditions for the various UIM peptides, both free and in the presence of ubiquitin. Crystals of Vps27p-2 were grown under the conditions described in the Experimental Procedures. X-ray diffraction data were collected to high (1.7 Å) resolution, and phases were determined using the anomalous signals from the three zinc atoms in the crystal lattice. After solvent flattening, the electron density map showed clear, continuous density that was readily interpretable, and it was ultimately possible to model the entire Vps27p-2 peptide (Vps27 residues 301-320, corresponding to UIM consensus residues 1-20). X-ray crystallographic data collection and structure refinement statistics are summarized in Table 3.4.

Previous sequence analyses have suggested that the UIM adopts a helical conformation, which could be embedded in a larger domain(s) (Hofmann and Falquet, 2001; Young et al., 1998). The UIM likely exists as a helix embedded within a larger domain (Hofmann and Falquet, 2001). The crystal structure of Vps27p-2 demonstrates that the UIM is indeed an amphipathic  $\alpha$ -helix. As shown in Figure 3.3, Vps27p adopts an  $\alpha$ -helical structure along its length, except for the two N-terminal residues Glu-301 and Glu-302, which adopt extended conformations. The helix is markedly amphipathic, with a hydrophobic stripe along one side (Figure 3.3A).

Unexpectedly, the Vps27-2 UIM crystallized as a left-handed, antiparallel four-helix bundle, with the hydrophobic face of each amphipathic helix packing into the middle of the bundle and the hydrophobic residues projecting from the exterior. Intriguingly, three of the four most highly conserved residues in the UIM (Leu-306, Ala-

Table 3.4. X-ray crystallographic data statistics

<b>Crystal</b>	
Space Group	P6 <sub>2</sub> 22
Cell Parameters (Å)	<i>a</i> = 34.895 <i>b</i> = 34.895 <i>c</i> = 64.176
<b>Data collection and processing statistics</b>	
Data set used for phasing	
Wavelength (Å)	0.97904
Resolution (Å)	50-1.7
Completeness (%)	99.9 (100) <sup>a</sup>
R <sub>merge</sub> (%) <sup>b</sup>	4.5 (6.9) <sup>a</sup>
Redundancy	11.3
I/σI	47.6 (38.0) <sup>a</sup>
<b>Data set used for refinement</b>	
Wavelength (Å)	1.54180
Resolution (Å)	30-1.7
Number of observed reflections	10024
Number of unique reflections	2729
Completeness (%)	95.1 (86.7) <sup>a</sup>
R <sub>merge</sub> (%) <sup>b</sup>	3.7 (9.6) <sup>a</sup>
Redundancy	3.7
I/σI	23.2 (13.7) <sup>a</sup>
<b>Refinement Statistics</b>	
Reflections used in refinement	2662
Reflections used in R <sub>free</sub>	67 (2.5%)
Water molecules incorporated	26
R <sub>cryst</sub> (%) <sup>c</sup>	18.0
R <sub>free</sub> (%) <sup>c</sup>	20.6
Ramachandran plot <sup>d</sup>	100% core region
Average B-factor (Å <sup>2</sup> )	14.2
<b>Errors</b>	
R.m.s.d. <sup>e</sup> of bond length (Å)	0.014
R.m.s.d. <sup>e</sup> of bond angles (°)	1.574

<sup>a</sup>Number in parenthesis is for highest resolution bin

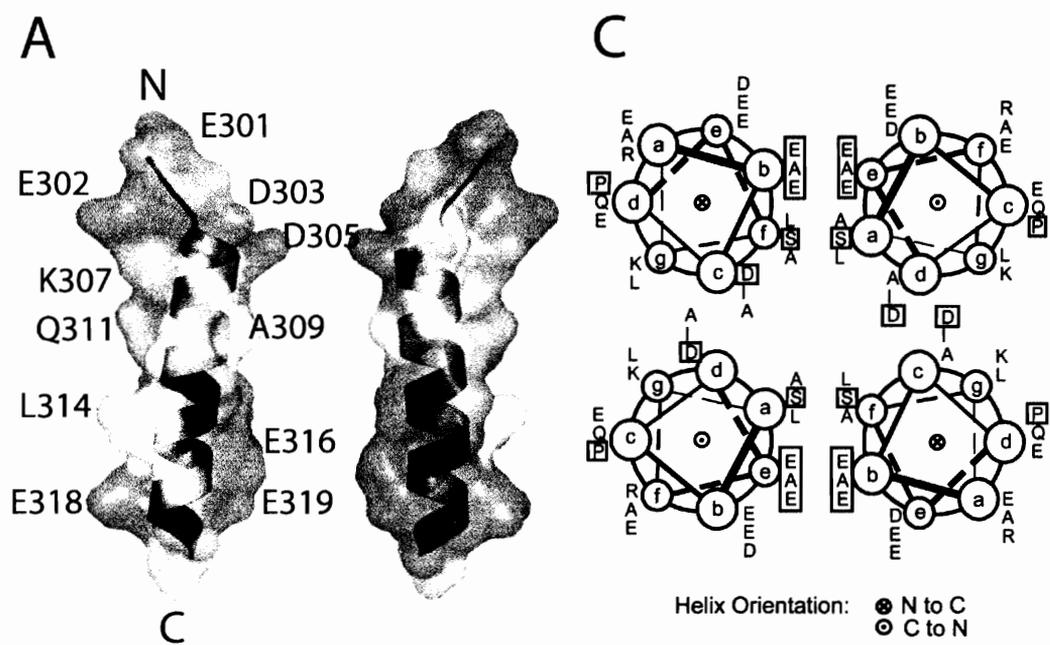
$$^b R_{\text{merge}} = \frac{\sum \sum_j |I - \langle I \rangle|}{\sum \langle I \rangle}$$

$$^c R = \frac{\sum \sum_j |F - \langle F \rangle|}{\sum \langle F \rangle}$$

<sup>d</sup>Stereochemistry was assessed with PROCHECK (77)

<sup>e</sup>R.m.s.d. root mean squared deviation

Figure 3.3. Structure of the Vps27p-2 UIM. (A) Space filling and ribbon representation of the two sides of the monomeric Vps27p UIM helix. Residues are colored according to the following scheme: acidic residues, red; basic residues, blue; uncharged hydrophilic residues, green; hydrophobic residues, grey. (B) Stereoview of the antiparallel four-helix bundle formed by Vps27p. The side chains of conserved residues are shown explicitly to emphasize how they tend to cluster on the interior of the helical bundle. (C) Helical wheel view of the residue positions in the four strands of the Vps27p bundle. Residues in highly conserved UIM positions are boxed.



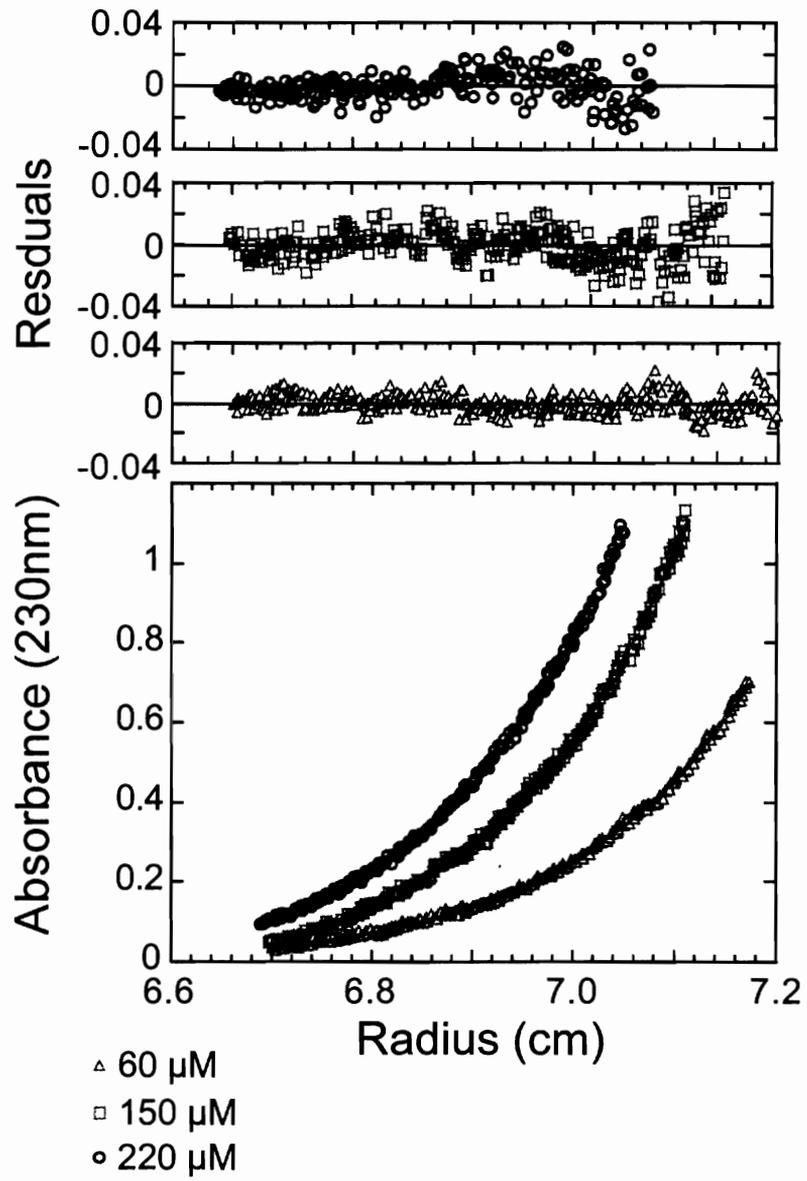
309, and Ser-313 in Vps27p-2) all lie on the same face of the helix and make homotypic interactions with their symmetry related mates in the center of each helix. Specifically, the Leu-306 side chains pack against related Leu-306 side chains across the bundle, the Ala-309 residues contact related Ala-309 residues on adjacent helices, and the Ser-313 side chains form water-mediated hydrogen bonds with related Ser-313 side chains on adjacent helices.

The three-dimensional packing arrangement in the crystal appears to be stabilized primarily by interbundle contacts mediated by the three unique zinc ions in the lattice. Zn-1 is coordinated by carboxylate oxygens from Glu-316, Glu-319, Asp-303 and Asp-305, Zn-2 is coordinated by carboxylate oxygens from two symmetry-related Glu-316 side chains and two symmetry-related Asp-303 side chains, as well as two water molecules, and Zn-3 is coordinated by oxygens from the C-terminal carboxylate, the side chain carboxylates of Glu-302 and Glu-318, and a nitrogen of an N-terminal amino group.

#### Oligomeric state of Vps27p-2 UIM in solution

The tetrameric structure of Vps27-2 in the crystal raised the possibility that the Vps27-2 peptide might also oligomerize in solution. Equilibrium analytical ultracentrifugation experiments were therefore performed to determine the oligomeric state of Vps27p-2 in solution (Figure 3.4). Centrifugation data from three different UIM concentrations (60-220  $\mu\text{M}$ ) were satisfactorily fit to a single species model, and the estimated molecular weight for the single species was  $2279 \pm 100$  Da, which agrees very well with the molecular weight of the Vps29p-2 monomer (2243 Da). These data demonstrate that Vps27p-2 UIM does not form higher order species under the solution

Figure 3.4. Equilibrium sedimentation analysis of the oligomeric state of Vps27p-2 in solution. Raw sedimentation data from three different initial Vps27p-2 concentrations (220  $\mu$ M, 150  $\mu$ M, 80  $\mu$ M) are shown below, together with the optimized global fits to a single species model. Fit residuals for the three different concentrations are shown in the upper panels. The estimated molecular weight from the single species model was  $2279 \pm 100$  Da, which agrees very well with the molecular weight of the Vps29p-2 monomer (2243 Da).



conditions tested. As this experiment sampled Vps27p-2 concentrations as high as ~1 mM, this implies that the Vps27p-2 tetramer, if it can exist in solution at all, cannot be a very stable structure.

#### Potential biological implications of the Vps27p tetramer

The most important question raised by our work is whether or not the Vps27p-2 UIM tetramer seen in the crystal is a biologically relevant assembly. It is not currently possible to answer this question definitively, and reasonable arguments can be made to support both sides of the question. On the negative side, the tetrameric packing of the Vps27p-2 UIM may simply reflect the fact that an amphipathic helix is being forced out of solution by the conditions of crystallization and that the hydrophobic face of this helix, which might normally be used to bind ubiquitin, will pack preferentially against itself (as occurs in the tetramer). Moreover, the fact that the Vps27p-2 peptide does not stably associate in solution supports the idea that the tetrameric Vps27p-2 structure is not an energetically favorable one, at least under the conditions tested thus far. It will, however, be important to repeat the analytical ultracentrifugation experiment in the presence of zinc, in order to test the possibility that zinc binding makes an important contribution to the stability of the Vps27p-2 tetramer.

In support of the possible biological relevance of UIM tetramerization, it is striking that many of the intramolecular packing interactions within the Vps27p-2 tetramer are formed by highly conserved UIM residues. Indeed, if zinc binding is considered to be an integral part of the structure, then every single one of the conserved UIM elements appears to make an important stabilizing contact in the Vps27p-2 crystal structure. Moreover, oligomerization of the UIM motif might help to explain why UIMs

are often found in clusters, in spite of the fact that they sort monoubiquitylated protein substrates. It has been noted previously that a number of proteins contain multiple UIM sequences, including epsin (3 UIM), epsin2 (2), EPS15 (2), EPS15R (2), Vps27p (2), hS5a (2), Ent1p (2), Ent2p (2), Vps27p (2) (and this number may be even greater if “cryptic” UIM sequences are considered) (Hofmann and Falquet, 2001). Intramolecular UIM association should be more favorable than intermolecular tetramerization, and the tandem UIM motifs found in many proteins could, of course, pack in an antiparallel orientation as is seen in the crystal. In addition, many UIM-containing proteins are known to associate with other proteins that also have UIM motifs. For example, Hrs associates with the Stam proteins (Asao et al., 1997), and Eps15 proteins associate with epsins (Rosenthal et al., 1999). Thus, there is the potential for “mixing and matching” of strands in the tetrameric structures, as is seen in other biological coiled-coil systems such as the SNARE complexes (Pelham, 2001) and in the networks formed by various transcription factors (Amati and Land, 1994; Shaulian and Karin, 2002).

If UIM tetramerization is indeed biologically relevant, then the next important question is whether tetramerization promotes or represses ubiquitin binding. Although neither possibility can be ruled out at this stage, it seems more attractive to think of UIM tetramerization as a mechanism for preventing unwanted ubiquitin binding, which would therefore allow UIM systems to be regulatable. In this model, conserved UIM residues would perform important functions in both the repressed state (in stabilizing the UIM tetramer) and activated state (in binding ubiquitin). We note that Eps15 is phosphorylated by the ligand-activated EGFR at a site just two residues upstream of its first UIM motif (Tyr850). This phosphorylation even is required for ligand-regulated endocytosis

(Confalonieri et al., 2000), and it is plausible to suppose that the phosphorylation event could destabilize a tetramer involving the first Eps15 UIM and thereby expose the UIM to bind ubiquitin. Although still hypothetical, this model points out the potential importance of the UIM tetramer seen in the Vps27p-2 crystal structure and emphasizes the need for additional research in this area.

## References

- Amati, B., and Land, H. (1994). Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. *Curr. Opin. Genet. Dev.* 4, 102-108.
- Asao, H., Sasaki, Y., Arita, T., Tanaka, N., Endo, K., Kasai, H., Takeshita, T., Endo, Y., Fujita, T., and Sugamura, K. (1997). Hrs is associated with STAM, a signal-transducing adaptor molecule. Its suppressive effect on cytokine-induced cell growth. *J. Biol. Chem.* 272, 32785-32791.
- Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M., and Pickart, C. (1996). Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc. Natl. Acad. Sci. U S A* 93, 861-866.
- Ben-Neriah, Y. (2002). Regulatory functions of ubiquitination in the immune system. *Nat. Immunol.* 3, 20-26.
- Bertolaet, B. L., Clarke, D. J., Wolff, M., Watson, M. H., Henze, M., Divita, G., and Reed, S. I. (2001). UBA domains of DNA damage-inducible proteins interact with ubiquitin. *Nat. Struct. Biol.* 8, 417-422.
- Bilodeau, P. S., Urbanowski, J. L., Winistorfer, S. C., and Piper, R. C. (2002). The Vps27p Hse1p complex binds ubiquitin and mediates endosomal protein sorting. *Nat. Cell Biol.* 4, 534-539.
- Buchberger, A. (2002). From UBA to UBX: new words in the ubiquitin vocabulary. *Trends Cell Biol.* 12, 216-221.
- Cheever, M. L., Sato, T. K., de Beer, T., Kutateladze, T. G., Emr, S. D., and Overduin, M. (2001). Phox domain interaction with PtdIns(3)P targets the Vam7 t-SNARE to vacuole membranes. *Nat. Cell Biol.* 3, 613-618.

Chen, H., Fre, S., Slepnev, V. I., Capua, M. R., Takei, K., Butler, M. H., Di Fiore, P. P., and De Camilli, P. (1998). Epsin is an EH-domain-binding protein implicated in clathrin-mediated endocytosis. *Nature* 394, 793-797.

Conaway, R. C., Brower, C. S., and Conaway, J. W. (2002). Emerging roles of ubiquitin in transcription regulation. *Science* 296, 1254-1258.

Confalonieri, S., Salcini, A. E., Puri, C., Tacchetti, C., and Di Fiore, P. P. (2000). Tyrosine phosphorylation of Eps15 is required for ligand-regulated, but not constitutive, endocytosis. *J. Cell Biol.* 150, 905-912.

Deveraux, Q., Ustrell, V., Pickart, C., and Rechsteiner, M. (1994). A 26 S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* 269, 7059-7061.

Dupre, S., Volland, C., and Haguenaer-Tsapis, R. (2001). Membrane transport: ubiquitylation in endosomal sorting. *Curr. Biol.* 11, R932-934.

Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model* 15, 132-134, 112-133.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* 67, 425-479.

Hicke, L. (1999). Gettin' down with ubiquitin: turning off cell-surface receptors, transporters and channels. *Trends Cell Biol.* 9, 107-112.

Hicke, L. (2001). Protein regulation by monoubiquitin. *Nat. Rev. Mol. Cell Biol.* 2, 195-201.

Higginson, D., Pornillos, O. W., Stray, K., and Sundquist, W. I. The protein recruiting functions of Hrs are sufficient to support Tsg101-dependent budding of virus-like particles of HIV-1 Gag. In preparation.

Hofmann, K., and Bucher, P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem. Sci.* 21, 172-173.

Hofmann, K., and Falquet, L. (2001). A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem. Sci.* 26, 347-350.

Jason, L. J., Moore, S. C., Lewis, J. D., Lindsey, G., and Ausio, J. (2002). Histone ubiquitination: a tagging tail unfolds? *Bioessays* 24, 166-174.

Johnson, M., Correia, J., Yphantis, D., and Halvorson, H. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* *36*, 575-588.

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta. Crystallogr. A* *47 (Pt 2)*, 110-119.

Katzmann, D. J., Babst, M., and Emr, S. D. (2001). Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell* *106*, 145-155.

Katzmann, D. J., Odorizzi, G., and Emr, S. D. (2002). Receptor downregulation and multivesicular-body sorting. *Nat. Rev. Mol. Cell Biol.* *3*, 893-905.

Laue, T., Shah, B., Ridgeway, T., and Pelletier, S. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In *Ultracentrifugation in Biochemistry and Polymer Science*, A. Rowe, and J. Horton, eds. (Cambridge, England, Royal Society of Chemistry), pp. 90-125.

Madura, K. (2002). The ubiquitin-associated (UBA) domain: on the path from prudence to prurience. *Cell Cycle* *1*, 235-244.

McRorie, D., and Voelker, P. (1993). Self-associating systems in the analytical ultracentrifuge (Fullerton, CA, Beckman Instruments Inc).

Meyer, H. H., Wang, Y., and Warren, G. (2002). Direct binding of ubiquitin conjugates by the mammalian p97 adaptor complexes, p47 and Ufd1-Npl4. *EMBO J.* *21*, 5645-5652.

Mori, S., Abeygunawardana, C., Johnson, M. O., and van Zijl, P. C. (1995). Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B* *108*, 94-98.

Mueller, T. D., and Feigon, J. (2002). Solution structures of UBA domains reveal a conserved hydrophobic surface for protein-protein interactions. *J. Mol. Biol.* *319*, 1243-1255.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S., and Dodson, E. J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta. Crystallogr. D. Biol. Crystallogr.* *55 (Pt 1)*, 247-255.

Myszka, D. G., and Morton, T. A. (1998). CLAMP: a biosensor kinetic data analysis program. *Trends Biochem. Sci.* *23*, 149-150.

- Otwinowski, Z., and Minor, W. (1987). Data collection software. *Methods in Enzymol.* *276*, 307-326.
- Pelham, H. R. (2001). SNAREs and the specificity of membrane fusion. *Trends Cell Biol.* *11*, 99-101.
- Pickart, C. M. (2001a). Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* *70*, 503-533.
- Pickart, C. M. (2001b). Ubiquitin enters the new millennium. *Mol. Cell* *8*, 499-504.
- Polo, S., Sigismund, S., Faretta, M., Guidi, M., Capua, M. R., Bossi, G., Chen, H., De Camilli, P., and Di Fiore, P. P. (2002). A single motif responsible for ubiquitin recognition and monoubiquitination in endocytic proteins. *Nature* *416*, 451-455.
- Ponting, C. P. (2000). Proteins of the endoplasmic-reticulum-associated degradation pathway: domain detection and function prediction. *Biochem. J.* *351 Pt 2*, 527-535.
- Pornillos, O., Alam, S., Rich, R. L., Myszka, D. G., Davis, D. R., and Sundquist, W. I. (2002a). Structure and functional interactions of the Tsg101 UEV domain. *EMBO J.* *21*, 2397-2406.
- Pornillos, O. P., Garrus, J. E., and Sundquist, W. I. (2002b). Mechanisms of enveloped RNA virus budding. *Trends Cell Biol.* *12*, 569-579.
- Project, C. C. (November 4, 1994). The CCP4 suite: programs for protein crystallography. *Acta. Crystallogr.* *D50*, 760-763.
- Raiborg, C., Bache, K. G., Gillooly, D. J., Madshus, I. H., Stang, E., and Stenmark, H. (2002). Hrs sorts ubiquitinated proteins into clathrin-coated microdomains of early endosomes. *Nat. Cell Biol.* *4*, 394-398.
- Raiborg, C., Bache, K. G., Mehlum, A., and Stenmark, H. (2001). Function of Hrs in endocytic trafficking and signalling. *Biochem. Soc. Trans.* *29*, 472-475.
- Reggiori, F., and Pelham, H. R. (2001). Sorting of proteins into multivesicular bodies: ubiquitin-dependent and -independent targeting. *EMBO J.* *20*, 5176-5186.
- Rosenthal, J. A., Chen, H., Slepnev, V. I., Pellegrini, L., Salcini, A. E., Di Fiore, P. P., and De Camilli, P. (1999). The epsins define a family of proteins that interact with components of the clathrin coat and contain a new protein module. *J. Biol. Chem.* *274*, 33959-33965.
- Shaulian, E., and Karin, M. (2002). AP-1 as a regulator of cell life and death. *Nat. Cell Biol.* *4*, E131-136.

Shekhtman, A., and Cowburn, D. (2002). A ubiquitin-interacting motif from Hrs binds to and occludes the ubiquitin surface necessary for polyubiquitination in monoubiquitinated proteins. *Biochem. Biophys. Res. Commun.* *296*, 1222-1227.

Shih, S. C., Katzmann, D. J., Schnell, J. D., Sutanto, M., Emr, S. D., and Hicke, L. (2002). Epsins and Vps27p/Hrs contain ubiquitin-binding domains that function in receptor endocytosis. *Nat. Cell Biol.* *4*, 389-393.

Shih, S. C., Sloper-Mould, K. E., and Hicke, L. (2000). Monoubiquitin carries a novel internalization signal that is appended to activated receptors. *EMBO J.* *19*, 187-198.

Sloper-Mould, K. E., Jemc, J. C., Pickart, C. M., and Hicke, L. (2001). Distinct functional surface regions on ubiquitin. *J. Biol. Chem.* *276*, 30483-30489.

Strack, B., Calistri, A., and Gottlinger, H. G. (2002). Late assembly domain function can exhibit context dependence and involves ubiquitin residues implicated in endocytosis. *J. Virol.* *76*, 5472-5479.

Terrell, J., Shih, S., Dunn, R., and Hicke, L. (1998). A function for monoubiquitination in the internalization of a G protein-coupled receptor. *Mol. Cell* *1*, 193-202.

Terwilliger, T. C. (2002). Automated structure solution, density modification and model building. *Acta Crystallogr. D. Biol. Crystallogr.* *58*, 1937-1940.

Thrower, J. S., Hoffman, L., Rechsteiner, M., and Pickart, C. M. (2000). Recognition of the polyubiquitin proteolytic signal. *EMBO J.* *19*, 94-102.

Urbanowski, J. L., and Piper, R. C. (2001). Ubiquitin sorts proteins into the intraluminal degradative compartment of the late-endosome/vacuole. *Traffic* *2*, 622-630.

van Delft, S., Schumacher, C., Hage, W., Verkleij, A. J., and van Bergen en Henegouwen, P. M. (1997). Association and colocalization of Eps15 with adaptor protein-2 and clathrin. *J. Cell Biol.* *136*, 811-821.

van Nocker, S., Sadis, S., Rubin, D. M., Glickman, M., Fu, H., Coux, O., Wefes, I., Finley, D., and Vierstra, R. D. (1996). The multiubiquitin-chain-binding protein Mub1 is a component of the 26S proteasome in *Saccharomyces cerevisiae* and plays a nonessential, substrate-specific role in protein turnover. *Mol. Cell Biol.* *16*, 6020-6028.

Varshavsky, A., Turner, G., Du, F., and Xie, Y. (2000). Felix Hoppe-Seyler Lecture 2000. The ubiquitin system and the N-end rule pathway. *Biol. Chem.* *381*, 779-789.

Walters, K. J., Kleijnen, M. F., Goh, A. M., Wagner, G., and Howley, P. M. (2002). Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a. *Biochemistry* *41*, 1767-1777.

Wiley, H. S., and Burke, P. M. (2001). Regulation of receptor tyrosine kinase signaling by endocytic trafficking. *Traffic* 2, 12-18.

Yoo, S., Myszka, D. G., Yeh, C., McMurray, M., Hill, C. P., and Sundquist, W. I. (1997). Molecular recognition in the HIV-1 capsid/cyclophilin A complex. *J. Mol. Biol.* 269, 780-795.

Young, P., Deveraux, Q., Beal, R. E., Pickart, C. M., and Rechsteiner, M. (1998). Characterization of two polyubiquitin binding sites in the 26 S protease subunit 5a. *J. Biol. Chem.* 273, 5461-5467.

## CHAPTER 4

### STRUCTURE AND UBIQUITIN INTERACTIONS OF THE CONSERVED NZF DOMAIN OF NPL4

Bin Wang<sup>\*†</sup>, Steven L. Alam<sup>\*†</sup>, Hemmo H. Meyer<sup>‡</sup>, Marielle Payne<sup>\*</sup>, Timothy L. Stemmler<sup>§</sup>, Darrell R. Davis<sup>\*¶</sup> and Wesley I. Sundquist<sup>\*¶</sup>

*Departments of \*Biochemistry and <sup>¶</sup>Medicinal Chemistry, University of Utah, Salt Lake City, UT 84132 USA.*

*<sup>§</sup>Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, 540 E. Canfield, Detroit, MI 48201 USA.*

*<sup>‡</sup>Department of Cell Biology, Yale University School of Medicine, 333 Cedar Street, SHM, C441, PO Box 208002, New Haven, CT, 06520-8002 USA.*

<sup>†</sup>These authors contributed equally to the structure determination.

<sup>¶</sup>Corresponding authors.

## Abstract

Ubiquitylated proteins are directed into a large number of different cellular pathways through interactions with proteins that contain conserved ubiquitin binding motifs. Here, we report the solution structure and ubiquitin binding properties of one such motif, the NZF (Npl4 Zinc Finger/RanBP2/Nup358 zinc finger) domain. Npl4 NZF forms a compact module comprised of four antiparallel  $\beta$ -strands linked by three ordered loops. A single zinc ion is coordinated by four conserved cysteines from the first and third loops, which form two rubredoxin knuckles. Npl4 NZF binds specifically, but weakly, to free ubiquitin using a conserved  $_{13}TF_{14}$  dipeptide to interact with the “I44” surface of ubiquitin. Our studies reveal the structure of this versatile class of protein binding domains and provide a means for identifying the subset of NZF domains likely to bind ubiquitin.

## Introduction

Protein ubiquitylation plays an important role in a large number of biological processes, including intracellular proteolysis, DNA repair, transcription, translation, signal transduction, cell cycle progression, organelle assembly, protein trafficking, and virus budding (Hershko and Ciechanover, 1998; Hicke, 2001; Pickart, 2001). Cells therefore dedicate an extensive array of machinery to the enzymology of ubiquitin transfer. Upon ubiquitylation, proteins must be directed to the correct intracellular locale, and pathways that utilize ubiquitin as a targeting signal therefore have “receptors” that specifically recognize ubiquitylated proteins. A number of such ubiquitin binding proteins are now known, and their study has led to the identification and characterization of several conserved ubiquitin binding motifs, including the UIM (Ubiquitin Interacting

Motif), the UBA (Ubiquitin-Associated domain), and the UEV (Ubiquitin E2 Variant/UBC-like domain) (Buchberger, 2002; Hofmann and Bucher, 1996; Hofmann and Falquet, 2001; Madura, 2002; Pickart, 2001). Each of these motifs forms an independent folding domain that can bind ubiquitin in vitro, although the molecular details of ubiquitin recognition are not yet understood in any case.

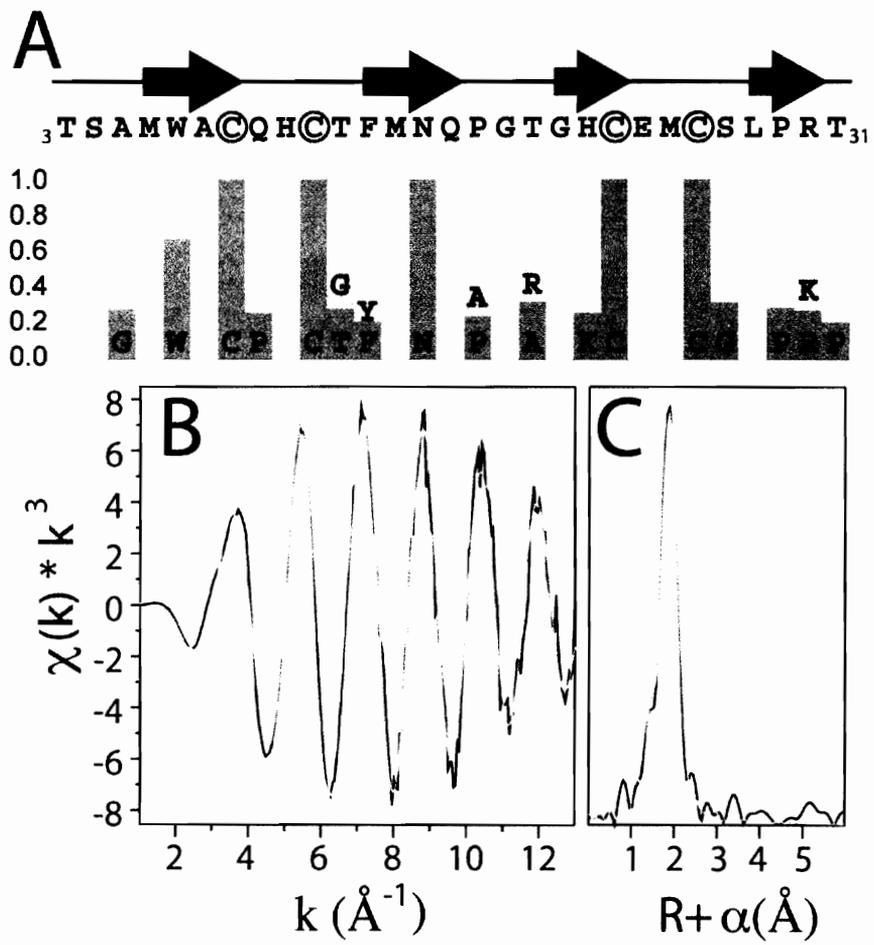
Given the generality of ubiquitin (Ub) as a targeting signal, it is likely that additional ubiquitin recognition motifs remain to be identified and characterized. One example is the recently discovered NZF domain (Npl4 Zinc Finger/RanBP2 zinc finger, see Figure 4.1A), which was first identified in Npl4, a protein that binds ubiquitin in vitro and mediates several different ubiquitin-dependent functions of the AAA-ATPase p97/Cdc48p. A hallmark of NZF domains is the presence of four regularly spaced cysteine residues, which suggests that the motif may be a metal binding module (Babst et al., 2002; Delaney et al., 1991; McMorroo et al., 1994; Meyer et al., 2002; Nakielny et al., 1999). NZF motifs appear to be quite prevalent, and protein database searches revealed 243 unique peptide sequences with four cysteines and an asparagine with the same spacing as those of Npl4 NZF (Figure 4.1A). This ensemble of sequences also exhibited significant conservation at 12/24 of the remaining residues, which presumably reflects the conserved structure and function(s) of this common domain.

In addition to Npl4 itself, several other NZF-containing proteins are known to be involved in ubiquitin-dependent processes. One example is Vps36p, a protein required for Vacuolar Protein Sorting of ubiquitylated proteins in yeast (Babst et al., 2002). Isolated NZF domains from both Npl4 and Vps36p bind mono- and poly-Ub chains in vitro, consistent with the idea that these NZF domains serve to recognize ubiquitylated

Figure 4.1. Primary sequence, conservation, and zinc binding by the Npl4 NZF domain.

(A) The primary sequence, numbering scheme, and secondary structure of the Npl4 NZF domain are shown above with the four cysteines that coordinate zinc circled in red. Shown below is the sequence conservation in putative NZF/Ran BP2-like domains from 243 sequences with four cysteines and one asparagine in the same spacing as Npl4 NZF. Residues present in more than 18% of NZF domains are shown explicitly, with bar heights representing their relative frequencies (see Experimental Procedures).

(B) Untransformed EXAFS spectra (black) and fit to a  $\text{Zn}(\text{Cys})_4$  model (green) of Npl4 NZF. (C) Fourier transformed EXAFS spectra (black) and fit (green) of Npl4 NZF.



protein substrates *in vivo* (Meyer et al., 2002). However, not all NZF-like domains bind ubiquitin, suggesting that those that do may represent a specialized subset of a much larger domain superfamily that shares a common three dimensional fold (Meyer et al., 2002).

The prevalence of the NZF domain and its apparent role in ubiquitin recognition in several important biological pathways led us to study the molecular basis for NZF/Ub interactions. Toward this end, we have analyzed the sequence conservation across putative NZF domains, characterized the metal and ubiquitin binding properties of the Npl4 and Vps36 NZF domains, determined the solution structure of the NZF domain from Npl4, and mapped the interaction surfaces in the Npl4/Ub complex.

## **Experimental Procedures**

### Data base searches for putative NZF domains

The SwissProt, PIR, PRF and KEGG GENE data bases were searched with pattern: x(6)-C-x(2)-C-x(3)-N-x(6)-C-x(2)-C-x(5), which yielded 243 unique matches in 166 different proteins (duplicates were removed manually). The apparent correlation between N16 and W7 (72%, see Figure 4.1A) was tested by a search with the pattern: x(4)-W-x-C-x(2)-C-x(10)-C-x(2)-C-x(5), which produced 161 matches (in 114 proteins), 84% of which had asparagine at the variable position corresponding to Asn-16 in Npl4 NZF. Searches with the even more stringent RanBP2-like pattern (Falquet et al., 2002) (and see <http://www.expasy.ch/cgi-bin/nicedoc.pl?PDOC50199>): x(4)-W-x-C-x(2)-C-x(3)-N-x(6)-C-x(2)-C-x(5), with both the W and N positions fixed, yielded stronger conservation at the variable positions, but was judged overly stringent as it excluded the known NZF domain of Vps36p (Meyer et al., 2002). Searches of the PDBSTR database

yielded no matches, indicating that the motif has not previously been characterized structurally.

#### Protein expression and purification

DNA encoding NZF domains from rat Npl4 (a.a. 580-608) and yeast Vps36p (a.a. 177-205) were cloned into pGEX-4T expression vectors. These constructs encoded GST-NZF fusion proteins with thrombin cleavages site between the GST and NZF domains (Haj et al., 2002).

Protein expression in DH5 $\alpha$  E. coli carrying the expression plasmids was induced with 0.5 mM IPTG (OD<sub>600</sub>=0.4). After 4 h at 23 °C, the cells were harvested by centrifugation and stored at -70 °C prior to protein purification. All steps in Npl4 NZF protein purification were performed at 4 °C, except where noted. Pellets from 6L of cells were resuspended in 40 mL buffer A (10 mM sodium phosphate pH 7.4, 150 mM NaCl, 5 mM  $\beta$ -mercaptoethanol (BME), 10  $\mu$ M ZnCl<sub>2</sub>), sonicated to lyse the cells, and centrifuged 1 h at 39,000 $\times$ g to clear insoluble debris. The soluble GST-Npl4 NZF fusion protein was purified by affinity chromatography on an FF 16/10 glutathione sepharose column (Amersham Biosciences). The bound protein was loaded and washed with  $\sim$ 10 column volumes buffer A and then eluted with 20 mM reduced glutathione in 50 mM Tris:HCl pH 8.0, 5 mM BME, 10  $\mu$ M ZnCl<sub>2</sub>.

Purified fractions ( $\sim$ 24 mL) were pooled and dialyzed 16 h against 2L thrombin cleavage buffer (20 mM Tris:HCl pH 8.0, 150 mM NaCl, 5 mM BME, 2.5 mM CaCl<sub>2</sub>, 10  $\mu$ M ZnCl<sub>2</sub>), quantified by optical absorption ( $\epsilon_{280}$ =47,330 M<sup>-1</sup>cm<sup>-1</sup>), incubated at 23°C for 16 hours with 1U of thrombin protease (Novagen) per mg GST-Npl4 NZF, and then

concentrated to ~3 mL by ultrafiltration (Amicon 3). Cleaved Npl4 NZF was separated from GST and uncleaved fusion protein by size exclusion chromatography (Superdex-75, APBiotech) in 100 mM NaCl, 20 mM sodium phosphate pH 8.0, 5 mM BME, 10  $\mu$ M ZnCl<sub>2</sub>. This procedure typically yielded ~4 mg of pure Npl4 NZF. <sup>15</sup>N- and <sup>15</sup>N/<sup>13</sup>C-labeled NZF peptides were expressed and purified as described above except that *E. coli* were grown in M9 minimal media with <sup>15</sup>NH<sub>4</sub>Cl or <sup>15</sup>NH<sub>4</sub>Cl/<sup>13</sup>C<sub>6</sub>-glucose as the sole nitrogen and carbon sources, respectively. Prior to atomic absorption and EXAFS spectroscopic analyses, unbound zinc was removed by gel filtration of pure Npl4 NZF (Superdex-75) in a degassed buffer lacking zinc (20mM sodium phosphate pH 5.5, 150 mM NaCl, 5 mM BME).

Thrombin cleavage left two non-native residues at the N-terminus of Npl4 NZF (NH<sub>2</sub>-Gly-Ser<sub>2</sub>). These residues are included in our numbering scheme, but are not shown in any of the figures. Purified Npl4 NZF was characterized by SDS-PAGE, N-terminal sequencing (NH<sub>2</sub>-Gly-Ser-Thr-Ser-Ala<sub>5</sub>), and MALDI mass spectrometry (MW<sub>calc</sub> = 3374.7 Da, MW<sub>obs</sub> = 3374.7 Da). The stoichiometry of bound zinc was analyzed by determining the protein concentration using optical spectroscopy ( $\epsilon_{280}$ =5840 M<sup>-1</sup>cm<sup>-1</sup>) and the zinc concentration by flame atomic absorption spectrometry. Unlabeled and <sup>15</sup>N-labeled ubiquitin were expressed and purified as described (Beal et al., 1996).

#### X-ray absorption spectroscopy

Pure Npl4 NZF was dialyzed against 2L XAS buffer (either 20mM sodium phosphate pH 5.5, 50 mM NaCl, 5mM BME or 20mM Tris pH 8.0, 50 mM NaCl, 5 mM BME), concentrated by ultrafiltration (Amicon 3), adjusted to a final concentration of

30% (w/v) glycerol, and snap frozen in a Lucite cell covered in Kapton tape at a final protein concentration of 1mM.

Data were collected at beamline 7-3 at the Stanford Synchrotron Radiation Laboratory (SSRL) using a Si(220) double crystal monochromator detuned 50% for harmonic rejection. Fluorescence excitation spectra were recorded using a 30-element Ge solid-state array detector. Samples were maintained at 10 K using a continuous-flow liquid helium cryostat (Oxford). XAS spectra were measured at 5 eV steps in the pre-edge, 0.35 eV steps in the edge (9640-9690 eV), and 0.05 Å<sup>-1</sup> increments in the extended X-ray absorption fine structure (EXAFS) region, integrating from 1-30 s in a  $k^3$  weighted manner over 45 minutes. EXAFS data represent the average of 9 scans. X-ray energies were calibrated with simultaneous measurement of a Zn foil absorption spectrum, assigning the first inflection point to 9659 eV. Similar spectra were obtained at both pH 5.5 and 8.0.

XAS data were analyzed using the Macintosh OS X version of the EXAFSPAK program, which integrates Feff v7.0 to generate theoretical models (see <http://www-ssrl.slac.stanford.edu/exafspak.html>). Data reduction employed second-order polynomial baseline flattening in the pre-edge and three-region cubic spline flattening throughout the EXAFS region. Data were converted to  $k$ -space using  $E_0=9680$  eV. Meaningful EXAFS data were limited to 10,325 eV due to monochromator imperfections. EXAFS data of crystallographically characterized model complexes were fit using amplitude and phase functions calculated in Feff v7.0 for a Zn-S interaction. This resulted in a scale factor of 1 and an  $E_0$  of -15.25, which were used for data fitting of the NZF protein. Coordination

numbers were fixed at half-integer values while R and  $\sigma^2$  were allowed to float during the final fits.

### NMR spectroscopy

Samples for structure determinations were ~1.0 mM NpI4 NZF in NMR buffer (20 mM sodium phosphate pH 5.5, 50 mM NaCl, 5 mM BME, 10  $\mu$ M ZnCl<sub>2</sub>, in 90%<sup>1</sup>H<sub>2</sub>O/10%<sup>2</sup>D<sub>2</sub>O). Samples were degassed and flame-sealed under argon in NMR sample tubes to reduce cysteine oxidation.

NMR spectra were recorded at 18 °C on a Varian Inova 600 MHz spectrometer equipped with a triple-resonance <sup>1</sup>H/<sup>13</sup>C/<sup>15</sup>N probe and z-axis pulsed field gradients. Backbone and sidechain assignments were made using a standard suite of triple resonance experiments (Clore and Gronenborn, 1994; Pornillos et al., 2002), except that 2D versions of the HNCACB and HN(CA)CO experiments were collected and analyzed. Sidechain assignments were completed using H(CCO)NH-TOCSY, (H)C(CCO)NH-TOCSY, <sup>15</sup>N-edited TOCSY, and <sup>13</sup>C-edited NOESY experiments, with aromatic resonances assigned using a combination of <sup>1</sup>H/<sup>13</sup>C HSQC and <sup>13</sup>C-edited NOESY experiments centered on the aromatic carbon resonances (125 ppm). Stereospecific assignments for 28 of 36  $\beta$ -methylene protons were obtained using a combination of HNHB, HN(CO)HB (Bax et al., 1994), <sup>15</sup>N-edited TOCSY, and NOESY data. Resonance assignments were complete except for amide proton and nitrogens for the first three amino acids, methionine methyls, and His H $\epsilon$ 1 aromatic protons.

3D <sup>15</sup>N-edited NOESY-HSQC (Mori et al., 1995; Zhang et al., 1994), and 3D <sup>13</sup>C-edited NOESY-HSQC (Muhandiram et al., 1993; Pascal et al., 1994) (140 msec mixing times) were used to generate distance restraints. Three-bond coupling constants (<sup>3</sup>J<sub>HN-HA</sub>)

were obtained from a 3D HNHA experiment (Kuboniwa et al., 1994). Hydrogen-bonded amides were identified using a long-range HNCO experiment with the N-CO coupling period tuned to 133.3 ms ( $\sim 1/4 \ ^1J_{\text{NCO}}$  for the one bond coupling (Meissner and Sorensen, 2000)). Amide temperature coefficients ( $\Delta\delta/\Delta T$ ) (Ballardin et al., 1978; Graham et al., 1992; Hsu et al., 2002) were calculated from the observed shifts in  $^{15}\text{N}/^1\text{H}$ -HSQC spectra recorded at 5, 10, 15, 20, 25 and 30 °C. Surface exposed amides for the free and ubiquitin-bound forms of NZF were characterized in SEAhsqc spectra (Pellecchia et al., 2001) tuned to a water-amide exchange time of 100 ms. All spectra were processed using FELIX 97 (MSI).

Npl4 NZF structures were refined using torsion angle dynamics (TAD) in DYANA (Guntert et al., 1997) and then regularized in CNS (Brunger et al., 1998) through a gentle simulated annealing with 362 NOE inter-proton restraints, four hydrogen-bonding restraints, and 13  $J_{\text{HNHA}}$  scalar-coupling restraints (Garrett et al., 1994). Initial rounds of refinement used only NOE data, which defined the general fold of the domain and revealed the stereochemistry about the coordinated zinc. Final refinements added restraints for Zn-S $\gamma$  (2.33-2.37Å), (Zn-C $\beta$  3.25-3.61Å), and S $\gamma$ -S $\gamma$  distances (3.78-4.15Å) to ensure approximate tetrahedral Zn coordination, as well as hydrogen-bonding restraints, and  $^3J_{\text{HNHA}}$  coupling constants. The 20 lowest energy structures overlay the mean coordinates with a root mean square deviation of  $0.11\pm 0.05$  and  $0.72\pm 0.05$ Å for backbone and heavy atoms respectively. NOE assignments and structure calculations were independently validated using the automated assignment/structure calculation package (CANDID within CYANA) (Herrmann et al., 2002), which produced essentially the same structure as was determined manually.

Structures were analyzed using PROCHECK-NMR (Laskowski et al., 1996), MOLMOL (Koradi et al., 1996), and INSIGHT II (MSI). Structure figures were created with MOLSCRIPT (Kraulis, 1991), BOBSCRIPT (Esnouf, 1997), GRASP (Nicholls et al., 1991), and PYMOL (DeLano Scientific).

#### Chemical shift perturbation mapping experiments

Chemical shift perturbation experiments were performed at 18 °C in NMR buffer. To identify the ubiquitin binding sites on the Npl4 NZF domain, unlabeled ubiquitin was titrated into 1.5 mM <sup>15</sup>N-labeled Npl4 NZF at final concentrations of 0 - 1.5 mM. To identify the Npl4 NZF binding site on ubiquitin, unlabeled Npl4 NZF was titrated into 0.15 mM <sup>15</sup>N-labeled ubiquitin at final concentrations of 0 - 0.15 mM. <sup>1</sup>H/<sup>15</sup>N HSQC spectra (Mori et al., 1995) were collected at different titration points, and normalized chemical shift changes ( $\delta$ ) were calculated using the equation:  $\delta = 25[(\delta_{\text{HN}})^2 + (\delta_{\text{N}}/5)^2]^{0.5}$  (Cheever et al., 2001). Amide chemical shift assignments for human ubiquitin were obtained from the VLI Research, Inc. website <http://www.vli-research.com/ubshifts.htm>.

#### Ubiquitin binding

Binding affinities of purified ubiquitin for GST-Npl4 NZF and GST-Vps36 NZF captured on anti-GST antibody biosensor surfaces were quantified as described previously (Garrus et al., 2001). All measurements were performed at 20 °C in 25 mM Tris·HCl pH 7.0, 5 mM BME, 2  $\mu$ M ZnCl<sub>2</sub>.

## Results

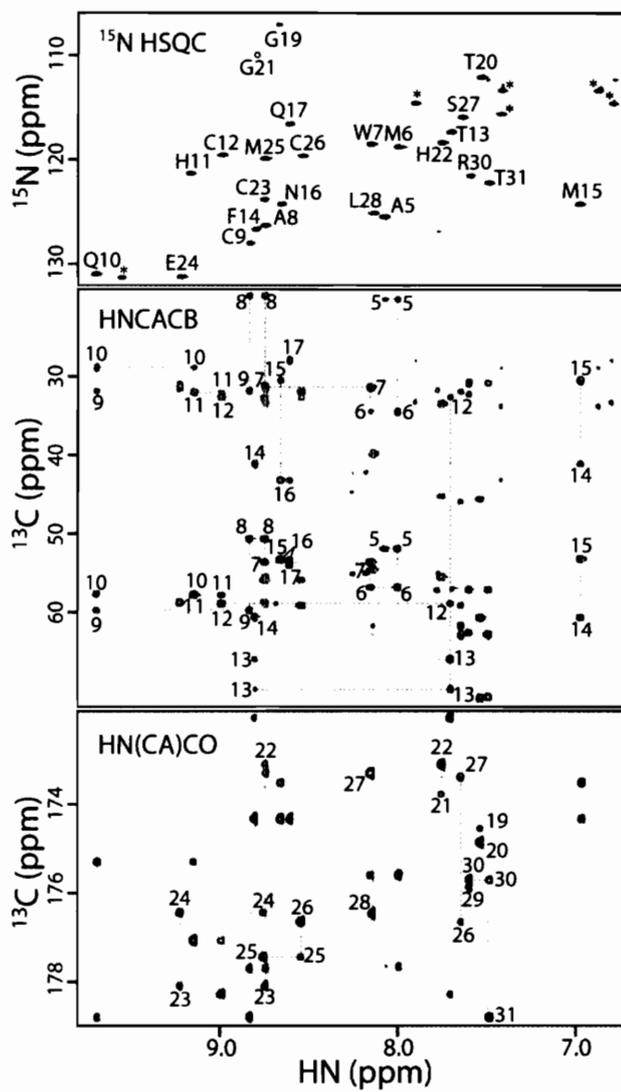
### The NZF is a zinc binding module

The Npl4 and Vps36p NZF constructs used in our studies span the conserved NZF region, and the sequence and numbering scheme for the rat Npl4 NZF polypeptide (Npl4 NZF) are given in Figure 4.1A. Npl4 NZF was expressed in *E. coli*, and the soluble protein purified to homogeneity under nondenaturing conditions (see Experimental Procedures). The presence of a single, covalently bound zinc ion was demonstrated using atomic absorption spectroscopy (0.93 mole Zn/mole Npl4 NZF) and EXAFS spectroscopy (Figures 4.1B and C). The best fit model to the transformed EXAFS spectrum revealed that the zinc coordination sphere contained four sulfur ligands ( $n=4.56$ ) and an average Zn-S bond of 2.33 Å. The Debye-Waller factor (0.00454 Å) was consistent with a slightly distorted tetrahedral geometry about the zinc center. We therefore conclude that Npl4 NZF coordinates a single, tetrahedral  $Zn^{2+}$  ion using the four conserved cysteines.

### NMR studies of Npl4 NZF

Nearly complete resonance and stereo specific assignments for Npl4 NZF were obtained using standard  $^1H/^{13}C/^{15}N$  heteronuclear NMR experiments (see Figure 4.2 and Experimental Procedures). As expected for a  $Zn(Cys)_4$  metal center, the four cysteine C $\beta$  resonances exhibited chemical shifts (31.5-32.5 ppm) that were intermediate between free cysteine (28.3 ppm) and oxidized cysteine (41.2 ppm). The structure of Npl4 NZF was initially calculated using a total of 362 assigned NOE restraints, which were sufficient to define both the protein structure and the zinc coordination geometry. Subsequent refinements included restraints for observed hydrogen bonds (8 restraints),  $J_{HNHA}$  scalar-

Figure 4.2. NMR assignments for the Npl4 NZF domain. Upper panel:  $^1\text{H}/^{15}\text{N}$  HSQC spectrum and chemical shift assignments for Npl4 NZF. Asterisks denote resonances from side chain amides. Center panel: 2D HNCACB spectrum showing intraresidue and sequential (i-1) connectivities. Red peaks are positive ( $\text{C}\alpha$ ), green peaks are negative ( $\text{C}\beta$ ). Lower panel: 2D HN(CA)CO spectrum showing intraresidue and sequential (i-1) connectivities. Vertical lines in the lower two panels provide correlations to the appropriate amide resonance in the upper panel. For clarity, residues 1-18 are labeled in the central panel and residues 19-31 are labeled in the lower panel. Note that correlations for residues 1-4 and glycine residues 19 and 21 were too weak to observe at these contour levels.



coupling (13 restraints), and tetrahedral zinc coordination (4 experimental, 10 idealized restraints). The final ensemble of 20 low-energy structures was of high quality (Table 4.1 and Figure 4.3), with root mean square deviations from the mean structure of  $0.11 \pm 0.05$  (backbone) and  $0.72 \pm 0.15 \text{ \AA}$  (all heavy atoms).

### Structure of Npl4 NZF

Npl4 NZF contains four short  $\beta$ -strands (S1-S4) that form two orthogonal hairpins (S1/S2 and S3/S4). The S1/S2 hairpin exhibits a canonical  $\beta$ -hairpin hydrogen bonding pattern, and the three expected H-bonds were observed directly in HNCO-LR spectra (Meissner and Sorensen, 2000). In contrast, the S3/S4 “hairpin” consists of two twisted, anti-parallel strands that lack canonical cross-strand hydrogen bonding. The four strands are connected by a compact central loop (L2) that contains overlapping type-I (Q17-T20) and reverse-gamma (G19-G21) turns, and two short metal binding loops (L1 and L3) that form “rubredoxin knuckles” (Figures 4.3 and 4.4) (Adman et al., 1975; Blake and Summers, 1994; Sarisky and Mayo, 2001).

The rubredoxin knuckle was initially recognized in the iron-binding centers of rubredoxins and has subsequently been observed at the zinc coordination sites in a number of proteins (Blake and Summers, 1994; Hammarstrom et al., 1996; Klug and Schwabe, 1995; Martinez-Yamout et al., 2000). Rubredoxin knuckles are six residue loops that connect two  $\beta$ -strands and display two metal binding Cys ligands (from loop positions 1 and 4). The canonical knuckle configuration orients the backbone amide nitrogens at positions 3 and 4 to hydrogen bond with the first cysteine ligand and orients the backbone amide at position 6 to hydrogen bond with the second cysteine ligand. This unusually tight metal binding loop is favored by a glycine residue at position 5, which

Table 4.1. Structure statistics for Npl4 NZF domain

	<TAD> <sup>a</sup>	<CNS> <sup>a</sup>
NOE distance restraints <sup>b</sup> (Å)	362	306
Intraresidue	101	87
Sequential ( $ i-j =1$ )	111	97
Medium range ( $2 \leq  i-j  \leq 5$ )	54	44
Long range ( $ i-j  > 5$ )	96	78
Zinc coordination restraints	14	14
Hydrogen bond distance restraints <sup>c</sup> (Å)	0	8
Hydrogen bonds		4
Three-bond $J_{\text{HNHA}}$ scalar coupling restraints (Hz)	0	13
Stereospecific assignments	28	28
DYANA Target function (Å <sup>4</sup> )	0.12±0.01	n/a
CNS energy	~800±200 <sup>d</sup>	66.5±0.5
Residual distance restraint violations		
Number of violations $\geq 0.1$ Å	0±0	0±0
Sum of violations (Å or kcal/mol) <sup>e</sup>	1.7±0.1	18.7±0.9
Maximum violation (Å)	0.14	
Residual scalar coupling restraint violations		
Number of violations $\geq 1^\circ$	na	0±0
Sum of violations ( $^\circ$ or kcal/mol) <sup>e</sup>	na	2.0±0.17
Maximum violation ( $^\circ$ )	na	
Van der Waals violations		
Number $\geq 0.1$ Å	0±0	0±0
Sum of violations (Å or kcal/mol) <sup>e</sup>	0.7±0.1	16.8±0.8
Maximum violation (Å)	0.09	
Ramachandran statistics (residue 5-31) <sup>f</sup>		
Favored	55.8%	60.5%
Allowed	35.7%	26.4%
Generously allowed	8.3%	13.2%
Disallowed	0.3%	0.0%
RMS deviations to the average coordinates <sup>g</sup> (Å)		
Residues 5-31		
Backbone	0.24±0.08	0.11±0.05
Heavy atoms	0.82±0.12	0.72±0.15

<sup>a</sup><TAD> is the ensemble of 40 lowest-penalty structures calculated using the program DYANA (Guntert et al., 1997). <CNS> is the same ensemble after 1000 steps (15 psec each) of simulated annealing at 25 K, 1000 slow-cooling steps to 0 K, and 10000 steps of restrained Powell minimization in cartesian space (anneal.inp protocol) (Brunger et al., 1998).

<sup>b</sup>Only meaningful and non-redundant restraints as determined by the DYANA CALIBA function.

<sup>c</sup>Two upper-limit distance restraints were used to define each hydrogen bond.

<sup>d</sup>Energies for structures input into CNS (from DYANA) were estimated within the generate\_easy.inp program after initial regularization without restraints.

<sup>e</sup>Violations from DYANA have units of Å, while violation energies from CNS are in kcal/mol.

<sup>f</sup>Determined using PROCHECK-NMR (Laskowski et al., 1996).

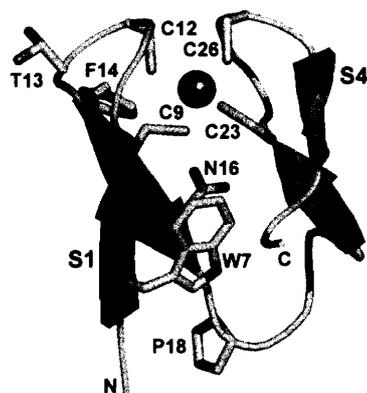
<sup>g</sup>Superposition and overall rmsds were calculated using the program MOLMOL (Koradi et al., 1996)

Figure 4.3. Structure of the Npl4 NZF domain. (A) Stereoview superposition of the 20 final structures of Npl4 NZF (backbone trace). Loops are shown in green, strands in red,  $Zn^{2+}$  in purple, and the four Zn S $\gamma$  ligands in gold. A subset of the conserved, well-ordered side chains are also shown (in green). (B) Ribbon diagram of the Npl4 NZF structure. The orientation and color coding are the same as in panel A. (C) View of the extensive long-range hydrogen bonding network in Npl4 NZF (dashed lines). In addition to the hydrogen bonds normally found in  $\beta$ -sheets and canonical turns, the following hydrogen bonds were predicted in Insight II (in 20/20 final structures, except where noted): 7N $\epsilon$ 1- T20C' (11/20), 9S $\gamma$ -10N, 9S $\gamma$ -11N, 12S $\gamma$ -13N (15/20), 16O $\delta$ -20O $\gamma$ , 16N $\delta$ 2-23S $\gamma$ , 26S $\gamma$ -27N, 26S $\gamma$ -28N.

A



B



C

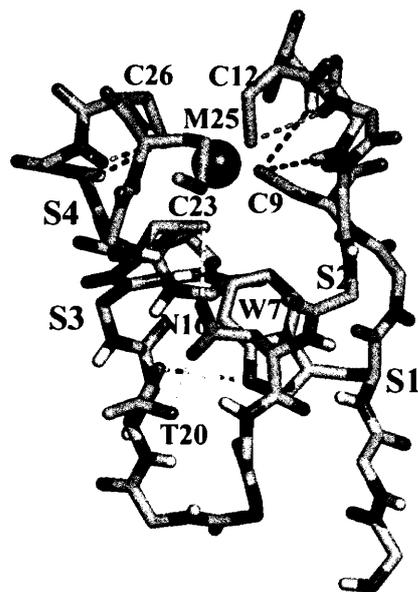
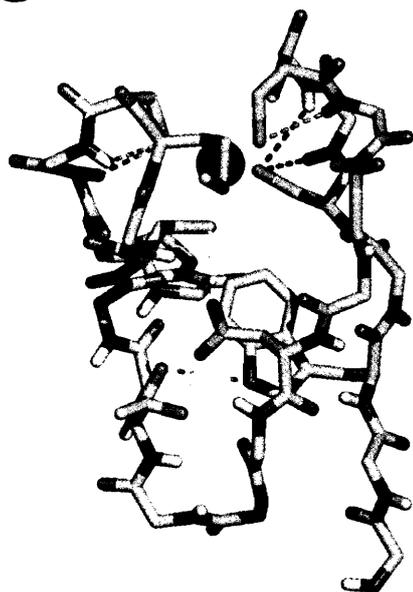
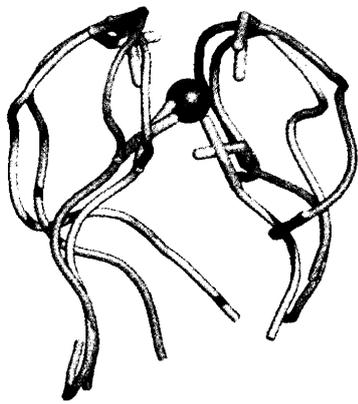
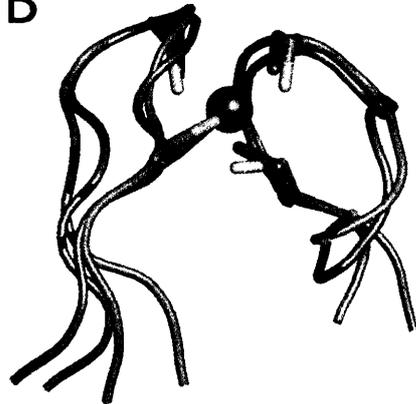


Figure 4.4. Structural similarity between the Npl4 NZF Zn site and the rubredoxin knuckles of other metalloproteins. (A) Superposition of Npl4 NZF (green) residues 6-15 (left side) and 22-29 (right side) onto the zinc binding site of the human transcriptional elongation factor TFIIIs (residues 9-18 and 39-46, salmon, 1TFI). This superposition gives an rmsd of 1.2 Å (backbone) (B) Superposition of Npl4 NZF (green) residues 7-15 and 22-30 onto the iron binding site from the rubredoxin of *Clostridium pasteurianum* (residues 4-12 and 38-46, purple, 1FHH). The superposition gives an rmsd of 0.83 Å (backbone).

A



B



adopts a positive backbone phi torsion angle. Variations in the details of this hydrogen bonding scheme are seen in some rubredoxins, for example in rubrerythrin (pdb entry 1dvb) (Sieker et al., 2000).

The knuckles in Npl4 NZF are slightly different from canonical rubredoxin knuckles because both NZF knuckles lack glycines at position 5 (Figures 4.1A, 4.3C, and 4.4). This opens up the loops slightly and alters the predicted hydrogen bonding pattern (see Figure 4.3C and caption). Moreover, in the second knuckle, the side chain amide nitrogen of N16 donates a hydrogen bond to the first cysteine (Cys 23, loop position 1). Nevertheless, the overall trajectories in the two Npl4 knuckles are similar to one another and to other rubredoxin knuckles (Figure 4.4).

As summarized in Table 4.2, nearly all of the residues that are conserved across NZF domains perform identifiable structural roles. The NZF tertiary structure is stabilized by a small hydrophobic “core,” which lies just below the zinc and consists primarily of the conserved W7 residue (Figures 4.1A and 4.3). The W7 H $\epsilon$ 1 proton is hydrogen bonded to the Thr20 carbonyl oxygen of strand 3, and the indole ring makes a series of hydrophobic contacts with residues from both strands S1 and S4. One face of the W7 ring is shielded from solvent by the guanidinium group of R30, and the other is shielded by the N16 side chain. The conserved N16 side chain also serves to bridge strands 2 and 3, as the amide nitrogen and oxygen hydrogen bond across to the T20 hydroxyl and C23 sulfur, respectively (Figure 4.3C).

#### Ubiquitin binding by the NZF domain

The Npl4 NZF motif can bind both mono- and poly-Ub, as analyzed by affinity co-purification (Meyer et al., 2002). To quantify the interaction between Npl4 NZF and

Table 4.2. Possible structural roles for conserved NZF residues

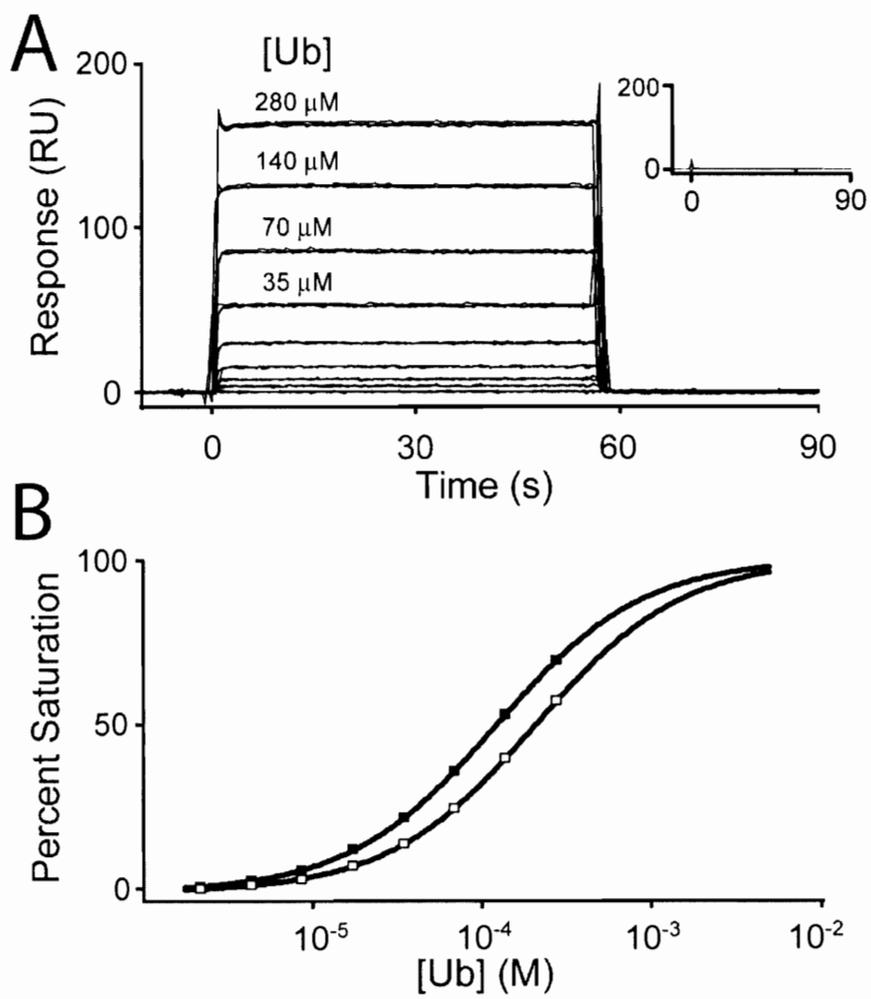
Residue <sup>a</sup>	Structural Role(s)
G(A)5	N-terminus of strand 1
W7	Hydrophobic core
C9	Zinc coordination
P(Q)10	Rubredoxin knuckle metal binding loop
C12	Zinc coordination
T/G13	Positive phi backbone torsion angle at position 5 of the rubredoxin knuckle (G), ubiquitin binding (T)
Y/F/L14	Ubiquitin binding
N16	Interstrand hydrogen bonding
A/P18	i+1 position of type-1 turn
R/A(T)20	Unclear from Npl4 NZF structure
K(H)22	Surface-exposed side chain
C23	Zinc coordination
C26	Zinc coordination
G(S)27	Positive phi backbone torsion angle at position 5 of the rubredoxin knuckle
P29	C-terminus of strand 4
K/R30	Surface-exposed side chain
P(T)31	Domain terminus

<sup>a</sup>Conserved residues in NZF domain (see Figure 1). The Npl4 NZF residue is given in parentheses when it does not correspond to the consensus sequence.

free ubiquitin, we performed biosensor binding experiments in which pure recombinant ubiquitin was allowed to bind to immobilized GST-NZF (Figure 4.5A). Ubiquitin bound to the Npl4 NZF surface with rapid, reversible kinetics, and the interaction was specific as ubiquitin did not bind to a control GST surface (Figure 4.5A, inset). A fit of the NZF/Ub binding data to a simple 1:1 model yielded a dissociation constant ( $K_D$ ) of  $122 \pm 2 \mu\text{M}$ . We speculate that this relatively weak binding affinity may reflect the fact that Npl4 NZF normally recognizes ubiquitylated proteins (rather than free ubiquitin) and may therefore gain additional binding energy through contacts with the conjugated C-terminal region of ubiquitin and/or through additional contacts with the Npl4 protein complex.

Analogous biosensor binding experiments were performed to quantify the interaction of ubiquitin with the immobilized NZF domain of yeast Vps36p (Figure 4.5B). Both the affinity ( $K_D = 199 \pm 17 \mu\text{M}$ ) and specificity (not shown) of the Vps36p NZF/Ub interaction were similar to those of the Npl4 NZF/Ub interaction. The observation that Vps36p NZF binds ubiquitin is not surprising from a biological perspective as Vps36p is required for the sorting of ubiquitylated proteins into the yeast vacuole (Babst et al., 2002). However, the observation is of interest from a structural perspective because Vps36p NZF has an asparagine residue in place of the conserved W7 residue that forms the hydrophobic core of Npl4 NZF. Our experiments demonstrate that the Vps36p NZF can nevertheless form a ubiquitin binding module, despite the absence of this hydrophobic core residue.

Figure 4.5. Quantitation of ubiquitin binding by NZF domains. (A) Surface plasmon resonance biosensor analysis of the Ub/Npl4 NZF interaction. Ubiquitin was injected in triplicate at concentrations of 0, 2.29, 4.38, 8.75, 17.5, 35.0, 70.0, 140, and 280  $\mu\text{M}$  over GST-Npl4 NZF captured on an anti-GST surface. The inset depicts the response obtained for 280  $\mu\text{M}$  ubiquitin injected over recombinant GST captured on an anti-GST surface (negative control). (B) Isotherms for Ub binding to the NZF domains from Npl4 and Vps36. Fitting the data to simple 1:1 binding models yielded  $K_D = 122 \pm 2 \mu\text{M}$  for Npl4 NZF (closed squares) and  $K_D = 199 \pm 17 \mu\text{M}$  for Vps36 NZF (open squares).

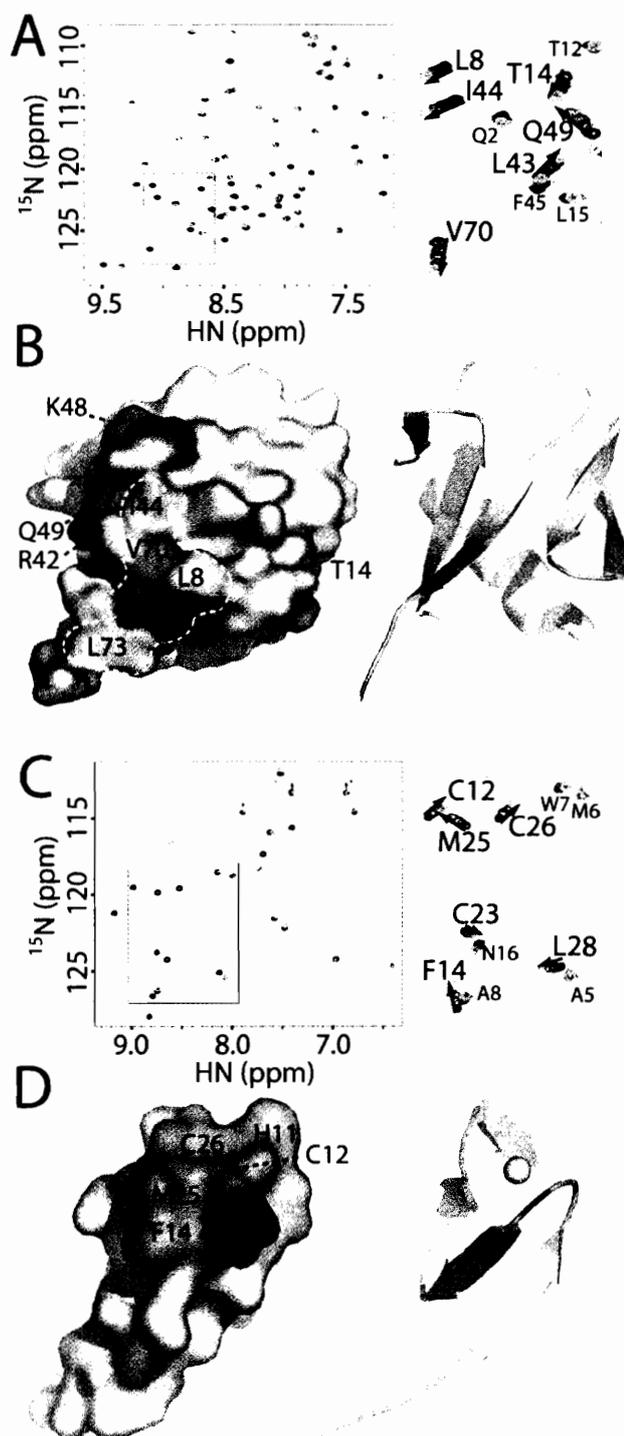


### Interaction surfaces in the NZF/Ub complex

Backbone amide NMR chemical shift perturbation experiments were used to map the interaction surfaces of both proteins within the Npl4 NZF/Ub complex (Figure 4.6). The Npl4 NZF/Ub complex was in fast exchange, as is typical for complexes with dissociation constants in the high micromolar range. Chemical shift changes observed in titration experiments therefore reflected the population-weighted average of shifts for the free and bound species.

A series of  $^1\text{H}/^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$ -labeled ubiquitin were collected during the stepwise titration of 0-1 equivalents of unlabeled Npl4 NZF (Figure 4.6A). Comparison of the spectra revealed that 15/70 observable ubiquitin backbone amide resonances shifted significantly upon complex formation ( $\delta \geq 2$ , see Figure 4.6A). The shifted ubiquitin residues generally clustered about the C-terminal three-strands of the  $\beta$ -sheet (with the exception of T14), and this site presumably corresponds to the NZF binding surface (Figure 4.6B). A prominent feature of the binding site is an exposed, hydrophobic surface formed by Ub residues L8, I44, V70, L71, and L73 (highlighted in Figure 4.6B). This corresponds to the “I44 surface” of ubiquitin, which has been shown by genetic analyses to function in endocytosis, proteasomal degradation, and HIV budding (Sloper-Mould et al., 2001; Strack et al., 2002). Ubiquitin residue K48 also shifted significantly upon NZF binding, which is notable because polyubiquitin chains linked via the K48 side chain target proteins for proteasome degradation. Our studies do not reveal, however, whether NZF binding is favored or disfavored by conjugation at this position.

Figure 4.6. NMR chemical shift mapping of the interaction surfaces of ubiquitin (A, B) and Npl4 NZF (C, D). (A) Overlaid  $^1\text{H}/^{15}\text{N}$  HSQC spectra of ubiquitin (0.15 mM) in the presence of 0 (blue), and 1.0 (gray) molar equivalents Npl4 NZF. The boxed insert in the left spectrum is expanded at right, and also includes intermediate titration points at: 0.125 (magenta), 0.25 (red), 0.5 (green) equivalents Npl4 NZF. (B) Residues with the greatest chemical shift changes upon addition of 1 equivalent Npl4 NZF are shown mapped onto surface and ribbon representations of ubiquitin. Residues shifted by  $\delta \geq 2$  are labeled at left and are colored using a gradient scheme from red ( $\delta=6$ , most shifted) to pink ( $\delta=2$ ). The hydrophobic patch on the “I44” surface of ubiquitin is outlined in white (see text). (C) Overlaid  $^1\text{H}/^{15}\text{N}$  HSQC spectra of Npl4 NZF (1.5 mM) in the presence of 0-1 molar equivalents of ubiquitin (shown with the same color coding used in panel A). (D) Residues with the greatest chemical shift changes upon addition of 1 equivalent ubiquitin are shown mapped onto surface and ribbon representations of Npl4 NZF. Residues with shifts of  $\delta \geq 2$  are labeled at left and are colored using a gradient scheme from red ( $\delta=8$ ) to pink ( $\delta=2$ ).



In the reciprocal experiment, a series of  $^1\text{H}/^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$ -labeled NZF were collected during the stepwise titration of 0-1 equivalents of unlabeled ubiquitin (Figure 4.6C). 10/25 observable NZF backbone amide resonances shifted significantly upon ubiquitin binding ( $\delta \geq 2$ ). All 10 shifted residues were clustered about the rubredoxin knuckles above the zinc coordination site and include the Zn ligands C12, C23, and C26, as well as T13, F14, and M25 (Figure 4.6D). Like its complementary counterpart on ubiquitin, the binding surface on NZF is primarily hydrophobic, but also includes residue T13. Thus, it appears that the interface between NZF and Ub is likely to resemble many other protein-protein interfaces that exhibit buried hydrophobic cores as well as key complementary hydrophilic interactions that provide orientation and specificity (Jones and Thornton, 1996).

## Discussion

Our studies demonstrate that the Npl4 NZF domain forms a tightly folded zinc binding module and provide a structural rationale for nearly every conserved residue in this domain family (Table 4.2). NZF now joins UBA (Dieckmann et al., 1998; Mueller and Feigon, 2002), UEV (Moraes et al., 2001; Pornillos et al., 2002; VanDemark et al., 2001), and UIM (Fisher et al., 2002) as a structurally characterized ubiquitin binding domain. Interestingly, all four of these structurally distinct motifs contact ubiquitin on the hydrophobic patch that surrounds the exposed surface of the C-terminal three strands of the  $\beta$ -sheet (the "I44 surface") (Bertolaet et al., 2001; Fisher et al., 2002; Mueller and Feigon, 2002; Pornillos et al., 2002; Shekhtman and Cowburn, 2002; Walters et al., 2002), highlighting the importance of this surface for ubiquitin function.

### Similarities between Npl4 NZF and other metalloproteins

The Npl4 NZF bears an unexpected resemblance to the “beta ribbon” zinc fingers found in a series of transcription-related proteins from archaea and eucarya, including TFIIIs and RPB9 (Chen et al., 2000; Qian et al., 1993a; Qian et al., 1993b; Wang et al., 1998). In all of these proteins, a single  $Zn^{2+}$  ion is coordinated by four Cys residues presented by the knuckles that link two  $\beta$ -hairpins. As shown in Figure 4.4A, zinc binding loops from Npl4 NZF and TFIIIs are similar, and overlay with heavy atom rmsd of less than 1.2 Å. The overall topologies of the TFIIIs (pdb code TFI), RPB9 (1QYP) and Npl4 NZF domains are also similar, although Npl4 NZF is more compact than the other two proteins and seems to represent a “stripped down” version of the fold. More generally, the zinc center of Npl4 NZF is also structurally related to the iron binding sites of the rubredoxin protein family, particularly to ruberythrin (Figure 4.4B). In spite of these structural similarities, however, there are no obvious functional connections between Npl4 NZF and the other proteins, beyond the possibility that all may mediate protein-protein interactions. Instead, the similarities appear to reflect the fact that the NZF fold provides a particularly favorable geometry for metal binding.

### Ubiquitin binding and conservation of the NZF domain

We have identified 243 sequences from 166 different proteins that exactly match the identity and spacing of the four cysteines and single asparagine of Npl4 NZF (Experimental and Figure 4.1A). Others had previously noted the presence of a highly related “zinc finger” sequence motif found in the nuclear pore protein, Ran BP2 (and related proteins) (Fauser et al., 2001; McMorrow et al., 1994; Sukegawa and Blobel, 1993; Wilken et al., 1995; Wu et al., 1995; Yokoyama et al., 1995). Searches based upon

the RanBP2/Nup358 motif have required the presence of residues equivalent to the Npl4 NZF cysteines, as well as W7 and N16, but allowed the spacing between the first two cysteines to vary between 2 and 4 residues (Falquet et al., 2002). These search criteria identify a comparable number of proteins to ours because they include family members that do not match our stringent spacing criteria but exclude NZF proteins like Vps36p (which lack the W7 residue). Thus, it appears that more than 200 known proteins contain NZF motifs, with the precise estimate depending upon the search criteria. Indeed, it will likely be difficult to come up with criteria that will unambiguously define the NZF motif given that there may be a continuum of protein structures that link the NZF and zinc ribbon motifs.

The prevalence of the NZF motif makes it important to understand the features that dictate different NZF functions. As shown above, the Npl4 and Vps36p NZF motifs both bind specifically to ubiquitin, as do the NZF motifs from two other proteins, TAB2 and RBCK2 (Meyer et al., 2002). However, the NZF domains from other proteins (e.g., mdm2 and RanBP2/Nup358), do not detectably bind ubiquitin in vitro (Meyer et al., 2002). Interestingly, all four known NZF motifs with ubiquitin binding activities display the highly exposed dipeptide sequence  ${}_{13}\text{TF}_{14}$ , whereas the two known NZF motifs that do not bind ubiquitin have very different residues at these positions (LV in RanBP2, and NE in mdm2). Moreover, our chemical shift mapping studies demonstrate that the  ${}_{13}\text{TF}_{14}$  dipeptide forms the primary binding site for Ub on Npl4 NZF (with T13 exhibiting the greatest chemical shift change of any NZF residue upon Ub binding). Note that on the basis of structural considerations alone, glycine should be heavily over-represented at residue 13, as this position in the rubredoxin knuckle adopts a positive phi backbone

torsion angle (Blake and Summers, 1994). These observations suggest that  ${}_{13}\text{TF}_{14}$  forms the major binding epitope in NZF domains that recognize ubiquitin.

To examine this possibility further, we tested for correlations between the identities of residues at positions 13 and 14 of the 243 NZF domains identified by our initial search criteria. Strikingly, we found that when a threonine is present at position 13, a large hydrophobic residue ( $\Phi$ ) almost always follows at position 14 (97% correlation, 63/65; F=32, Y=16, L=15). In contrast, when any residue except threonine is present at position 13, a large hydrophobic residue follows only 16% of the time (29/178 total). This analysis, together with our structural and binding studies, supports the idea that the NZF motif forms a scaffold that presents the surface-exposed  ${}_{13}\text{T}\Phi_{14}$  dipeptide (and surrounding residues) to recognize the hydrophobic “I44” surface of ubiquitin.

Our sequence analyses also revealed that the LV dipeptide co-varies strongly at NZF positions 13 and 14 (LV, 67% co-variation). We speculate that this dipeptide may also form a recognition epitope for another NZF domain. Interestingly, several proteins, including RanBP2/Nup358 and Nup153, contain multicopy tracts of NZF domains that bind the GDP form of the Ran GTPase (Nakielny et al., 1999; Yaseen and Blobel, 1999). These tracts are composed primarily of multiple copies of  ${}_{13}\text{LV}_{14}$ -containing NZF domains, suggesting that RanGDP may be the preferred protein binding partner for this class of NZF domains. The RanBP2 zinc fingers also bind exportin-1 (Singh et al., 1999), which is apparently yet another protein ligand that can be recognized by the NZF domain.

### Biological implications

The NZF motif was first recognized in the mammalian protein, Npl4, which is a subunit of the heterodimeric UN complex (together with Ufd1) (Meyer et al., 2002). UN

is one of at least two alternative adapters that target the AAA-ATPase p97/Cdc48p to specific protein substrates. The UN complex is required for at least three p97/Cdc48p-mediated reactions, including retro-translocation of ubiquitylated proteins from the ER into the cytosol (ERAD) and the topologically related ubiquitin-dependent processing of transcription factors Spt23 and Mga2 in the ER membrane and subsequent mobilization by Cdc48p (RUP) (Bays and Hampton, 2002; Tsai et al., 2002). p97-UN is also required for an as yet undefined reaction that leads to the formation of a closed nuclear envelope after mitosis (Hetzer et al., 2001).

Although the exact mechanistic role of UN in ERAD and RUP remains to be determined, UN can bind ubiquitin-conjugates by means of the Npl4 NZF domain and can simultaneously recruit the p97 ATPase (Meyer et al., 2002). One possibility is therefore that UN functions by recognizing ubiquitylated protein substrates in the ER membrane and recruiting p97, which then uses the energy of ATP hydrolysis either to pull its substrates out of the membrane or to separate them from other proteins.

Interestingly, the other known p97 adapter, p47, which functions in the reassembly of the Golgi apparatus (Kondo et al., 1997) and expansion of the nuclear envelope after mitosis (Hetzer et al., 2001), can also link the ATPase to ubiquitin conjugates (Meyer et al., 2002). In that case, however, ubiquitin recognition is mediated by a UBA domain in p47 and the ubiquitylated substrate remains to be identified.

Vps36p provides another example in which an NZF motif could help to recruit ubiquitylated proteins into a biological process. We have shown that the Vps36p NZF can bind ubiquitin (Meyer et al., 2002, and this work) and others have shown that Vps36p is a member of the Escrt-II complex, which is required for sorting of ubiquitylated proteins

into multivesicular bodies (MVB) (Babst et al., 2002). It is therefore plausible to suppose that Vps36p uses its NZF motif to recognize ubiquitylated protein substrates as they are sorted through the MVB pathway. In further analogy to Npl4, there is even the potential that Vps36p could also act as an adaptor for Vps4p, a AAA ATPase that functions late in this pathway (Babst et al., 1998; Bishop and Woodman, 2001; Scheuring et al., 2001).

Our studies also suggest that many other less well characterized proteins that contain NZF motifs of the “TF” class will also exhibit ubiquitin binding activities. Strengthening this suggestion is the fact that a number of these proteins contain additional domains linked to various aspects of ubiquitin biochemistry. Examples include proteins with E3 RING finger motifs that bind ubiquitin conjugating (E2) enzymes (e.g., Ubc7ip3, NCBI entry Q9BYM8) (Martinez-Noel et al., 1999; Meyer et al., 2002), proteins that bind E3 ubiquitin transferases (e.g., RYBP, NP\_036366) (Garcia et al., 1999), proteins with their own ubiquitin-like domains (e.g., Sharpin, NP\_112415), and proteins that contain other known ubiquitin binding domains (e.g., MGC45404, NP\_69000). There are also a number of proteins with NZF motifs of the TF class without any known links to ubiquitin, and we suggest that investigations in this direction are likely to be fruitful.

In summary, the NZF motif is a zinc binding module that can be incorporated into multifunctional proteins and used to bind ubiquitin, RanGDP, exportin-1, and possibly other proteins. Our studies reveal the NZF structure, define the NZF/Ub interface, and provide criteria for distinguishing functional variants of this versatile motif.

## Acknowledgements

We thank Rebecca Rich and David Myszka (University of Utah Protein Interactions Core Facility) for BIAcore biosensor analyses, Bob Schackmann (University of Utah DNA and Peptide Core Facility) for N-terminal protein sequencing, Dennis Winge (Utah) for atomic absorption spectroscopic analyses, the University of Utah Mass Spectrometry Core Facility for protein mass spectrometry, the SLAC staff, Eric Ross for computer support, Cecile Pickart for the ubiquitin expression clone and protocol, and Chris Hill, Dennis Winge, Mike Summers, and Katie Ullman for critical reading of the manuscript. WIS and DRD acknowledge NIH grant support, and HH Meyer was supported by grants from the NIH and the Human frontiers science project to Graham Warren. The Utah Biomolecular NMR Facility is supported by grants from NIH and NSF. Structure coordinates and chemical shifts have been deposited (xx, CNS ensemble).

## References

- Adman, E., Watenpaugh, K. D., and Jensen, L. H. (1975). NH---S hydrogen bonds in *Peptococcus aerogenes* ferredoxin, *Clostridium pasteurianum* rubredoxin, and Chromatium high potential iron protein. *Proc. Natl. Acad. Sci. U S A* 72, 4854-4858.
- Babst, M., Katzmann, D., Snyder, W., Wendland, B., and Emr, S. (2002). Endosome-Associated Complex, ESCRT-II, Recruits Transport Machinery for Protein Sorting at the Multivesicular Body. *Dev. Cell* 3, 283.
- Babst, M., Wendland, B., Estepa, E. J., and Emr, S. D. (1998). The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. *EMBO J.* 17, 2982-2993.
- Ballardin, A., Fischman, A. J., Gibbons, W. A., Roy, J., Schwartz, I. L., Smith, C. W., Walter, R., and Wyssbrod, H. R. (1978). Conformational studies on [Pro3, Gly4]-oxytocin in dimethyl sulfoxide by 1H nuclear magnetic resonance spectroscopy: evidence for a type II beta turn in the cyclic moiety. *Biochemistry* 17, 4443-4454.
- Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R., and Zhu, G. (1994). Measurement of homo- and heteronuclear J couplings from quantitative J correlation. *Methods Enzymol.* 239, 79-105.

Bays, N. W., and Hampton, R. Y. (2002). Cdc48-Ufd1-Npl4: stuck in the middle with Ub. *Curr. Biol.* *12*, R366-371.

Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M., and Pickart, C. (1996). Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc. Natl. Acad. Sci. U S A* *93*, 861-866.

Bertolaet, B. L., Clarke, D. J., Wolff, M., Watson, M. H., Henze, M., Divita, G., and Reed, S. I. (2001). UBA domains of DNA damage-inducible proteins interact with ubiquitin. *Nat. Struct. Biol.* *8*, 417-422.

Bishop, N., and Woodman, P. (2001). TSG101/mammalian VPS23 and mammalian VPS28 interact directly and are recruited to VPS4-induced endosomes. *J. Biol. Chem.* *276*, 11735-11742.

Blake, P. R., and Summers, M. F. (1994). Probing the unusually similar metal coordination sites of retroviral zinc fingers and iron-sulfur proteins by nuclear magnetic resonance. *Advan. in Biophys. Chem.* *4*, 1-30.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* *54 (Pt 5)*, 905-921.

Buchberger, A. (2002). From UBA to UBX: new words in the ubiquitin vocabulary. *Trends Cell Biol.* *12*, 216-221.

Cheever, M. L., Sato, T. K., de Beer, T., Kutateladze, T. G., Emr, S. D., and Overduin, M. (2001). Phox domain interaction with PtdIns(3)P targets the Vam7 t-SNARE to vacuole membranes. *Nat. Cell Biol.* *3*, 613-618.

Chen, H. T., Legault, P., Glushka, J., Omichinski, J. G., and Scott, R. A. (2000). Structure of a (Cys3His) zinc ribbon, a ubiquitous motif in archaeal and eucaryal transcription. *Protein Sci.* *9*, 1743-1752.

Clore, G. M., and Gronenborn, A. M. (1994). Multidimensional heteronuclear nuclear magnetic resonance of proteins. *Methods Enzymol.* *239*, 349-363.

Delaney, S. J., Hayward, D. C., Barleben, F., Fischbach, K. F., and Miklos, G. L. (1991). Molecular cloning and analysis of small optic lobes, a structural brain gene of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U S A* *88*, 7214-7218.

Dieckmann, T., Withers-Ward, E. S., Jarosinski, M. A., Liu, C. F., Chen, I. S., and Feigon, J. (1998). Structure of a human DNA repair protein UBA domain that interacts with HIV-1 Vpr. *Nat. Struct. Biol.* *5*, 1042-1047.

Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model* *15*, 132-134, 112-133.

- Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K., and Bairoch, A. (2002). The PROSITE database, its status in 2002. *Nucleic Acids Res.* *30*, 235-238.
- Fauser, S., Aslanukov, A., Roepman, R., and Ferreira, P. A. (2001). Genomic organization, expression, and localization of murine Ran-binding protein 2 (RanBP2) gene. *Mamm. Genome* *12*, 406-415.
- Fisher, R., Wang, B., Alam, S., Sundquist, W. I., and Hill, C. P. (2002). Structure and ubiquitin binding of the conserved Ubiquitin Interacting Motif (UIM). In preparation.
- Garcia, E., Marcos-Gutierrez, C., del Mar Lorente, M., Moreno, J. C., and Vidal, M. (1999). RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. *EMBO J.* *18*, 3404-3418.
- Garrett, D. S., Kuszewski, J., Hancock, T. J., Lodi, P. J., Vuister, G. W., Gronenborn, A. M., and Clore, G. M. (1994). The impact of direct refinement against three-bond HN-C alpha H coupling constants on protein structure determination by NMR. *J. Magn. Reson. B* *104*, 99-103.
- Garrus, J. E., von Schwedler, U. K., Pornillos, O. W., Morham, S. G., Zavitz, K. H., Wang, H. E., Wettstein, D. A., Stray, K. M., Cote, M., Rich, R. L., *et al.* (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* *107*, 55-65.
- Graham, W. H., Carter, E. S., 2nd, and Hicks, R. P. (1992). Conformational analysis of Met-enkephalin in both aqueous solution and in the presence of sodium dodecyl sulfate micelles using multidimensional NMR and molecular modeling. *Biopolymers* *32*, 1755-1764.
- Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* *273*, 283-298.
- Haj, F. G., Verveer, P. J., Squire, A., Neel, B. G., and Bastiaens, P. I. (2002). Imaging sites of receptor dephosphorylation by PTP1B on the surface of the endoplasmic reticulum. *Science* *295*, 1708-1711.
- Hammarstrom, A., Berndt, K. D., Sillard, R., Adermann, K., and Otting, G. (1996). Solution structure of a naturally-occurring zinc-peptide complex demonstrates that the N-terminal zinc-binding module of the Lasp-1 LIM domain is an independent folding unit. *Biochemistry* *35*, 12723-12732.
- Herrmann, T., Guntert, P., and Wuthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* *319*, 209-227.
- Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* *67*, 425-479.

- Hetzer, M., Meyer, H. H., Walther, T. C., Bilbao-Cortes, D., Warren, G., and Mattaj, J. W. (2001). Distinct AAA-ATPase p97 complexes function in discrete steps of nuclear assembly. *Nat. Cell Biol.* 3, 1086-1091.
- Hicke, L. (2001). Protein regulation by monoubiquitin. *Nat. Rev. Mol. Cell Biol.* 2, 195-201.
- Hofmann, K., and Bucher, P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem. Sci.* 21, 172-173.
- Hofmann, K., and Falquet, L. (2001). A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem. Sci.* 26, 347-350.
- Hsu, S. T., Breukink, E., de Kruijff, B., Kaptein, R., Bonvin, A. M., and van Nuland, N. A. (2002). Mapping the targeted membrane pore formation mechanism by solution NMR: the nisin Z and lipid II interaction in SDS micelles. *Biochemistry* 41, 7670-7676.
- Jones, S., and Thornton, J. M. (1996). Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U S A* 93, 13-20.
- Klug, A., and Schwabe, J. W. (1995). Protein motifs 5. Zinc fingers. *FASEB J.* 9, 597-604.
- Kondo, H., Rabouille, C., Newman, R., Levine, T. P., Pappin, D., Freemont, P., and Warren, G. (1997). p47 is a cofactor for p97-mediated membrane fusion. *Nature* 388, 75-78.
- Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* 14, 51-55.
- Kraulis, P. J. (1991). MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Applied Cryst.* 24, 946-950.
- Kuboniwa, H., Grzesiek, S., Delaglio, F., and Bax, A. (1994). Measurement of HN-H alpha J couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J. Biomol. NMR* 4, 871-878.
- Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477-486.
- Madura, K. (2002). The ubiquitin-associated (UBA) domain: on the path from prudence to prurience. *Cell Cycle* 1, 235-244.

- Martinez-Noel, G., Niedenthal, R., Tamura, T., and Harbers, K. (1999). A family of structurally related RING finger proteins interacts specifically with the ubiquitin-conjugating enzyme UbcM4. *FEBS Lett.* *454*, 257-261.
- Martinez-Yamout, M., Legge, G. B., Zhang, O., Wright, P. E., and Dyson, H. J. (2000). Solution structure of the cysteine-rich domain of the Escherichia coli chaperone protein DnaJ. *J. Mol. Biol.* *300*, 805-818.
- McMorrow, I., Bastos, R., Horton, H., and Burke, B. (1994). Sequence analysis of a cDNA encoding a human nuclear pore complex protein, hnup153. *Biochem. Biophys. Acta.* *1217*, 219-223.
- Meissner, A., and Sorensen, O. W. (2000). New techniques for the measurement of C'N and C'H(N) J coupling constants across hydrogen bonds in proteins. *J. Magn. Reson.* *143*, 387-390.
- Meyer, H. H., Wang, Y., and Warren, G. (2002). Direct binding of ubiquitin conjugates by the mammalian p97 adaptor complexes, p47 and Ufd1-Npl4. *EMBO J.* *21*, 5645-5652.
- Moraes, T. F., Edwards, R. A., McKenna, S., Pastushok, L., Xiao, W., Glover, J. N., and Ellison, M. J. (2001). Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2-hUbc13. *Nat. Struct. Biol.* *8*, 669-673.
- Mori, S., Abeygunawardana, C., Johnson, M. O., and van Zijl, P. C. (1995). Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Magn. Reson. B* *108*, 94-98.
- Mueller, T. D., and Feigon, J. (2002). Solution structures of UBA domains reveal a conserved hydrophobic surface for protein-protein interactions. *J. Mol. Biol.* *319*, 1243-1255.
- Muhandiram, D. R., Xy, G. Y., and Kay, L. E. (1993). An enhanced-sensitivity pure absorption gradient 4D <sup>15</sup>N, <sup>13</sup>C-edited NOESY experiment. *J. Biomol. NMR* *3*, 463-470.
- Nakielny, S., Shaikh, S., Burke, B., and Dreyfuss, G. (1999). Nup153 is an M9-containing mobile nucleoporin with a novel Ran-binding domain. *EMBO J.* *18*, 1982-1995.
- Nicholls, A., Sharp, K. A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* *11*, 281-296.
- Pascal, S. M. M., Yamazaki, D. R., Forman-Kay, J. D., and Kay, L. E. (1994). Simultaneous acquisition of <sup>15</sup>N- and <sup>13</sup>C edited NOE spectra of proteins dissolved in H<sub>2</sub>O. *J. Mag. Res.* *191*, 197-201.

- Pellecchia, M., Meininger, D., Shen, A. L., Jack, R., Kasper, C. B., and Sem, D. S. (2001). SEA-TROSY (solvent exposed amides with TROSY): a method to resolve the problem of spectral overlap in very large proteins. *J. Am. Chem. Soc.* *123*, 4633-4634.
- Pickart, C. M. (2001). Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* *70*, 503-533.
- Pornillos, O., Alam, S., Rich, R. L., Myszkka, D. G., Davis, D. R., and Sundquist, W. I. (2002). Structure and functional interactions of the Tsg101 UEV domain. *EMBO J.* *21*, 2397-2406.
- Qian, X., Gozani, S. N., Yoon, H., Jeon, C. J., Agarwal, K., and Weiss, M. A. (1993a). Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. *Biochemistry* *32*, 9944-9959.
- Qian, X., Jeon, C., Yoon, H., Agarwal, K., and Weiss, M. A. (1993b). Structure of a new nucleic-acid-binding motif in eukaryotic transcriptional elongation factor TFIIS. *Nature* *365*, 277-279.
- Sarisky, C. A., and Mayo, S. L. (2001). The beta-beta-alpha fold: explorations in sequence space. *J. Mol. Biol.* *307*, 1411-1418.
- Scheuring, S., Rohricht, R. A., Schoning-Burkhardt, B., Beyer, A., Muller, S., Abts, H. F., and Kohrer, K. (2001). Mammalian cells express two VPS4 proteins both of which are involved in intracellular protein trafficking. *J. Mol. Biol.* *312*, 469-480.
- Shekhtman, A., and Cowburn, D. (2002). A ubiquitin-interacting motif from Hrs binds to and occludes the ubiquitin surface necessary for polyubiquitination in monoubiquitinated proteins. *Biochem. Biophys. Res. Commun.* *296*, 1222-1227.
- Sieker, L. C., Holmes, M., Le Trong, I., Turley, S., Liu, M. Y., LeGall, J., and Stenkamp, R. E. (2000). The 1.9 Å crystal structure of the "as isolated" rubrerythrin from *Desulfovibrio vulgaris*: some surprising results. *J. Biol. Inorg. Chem.* *5*, 505-513.
- Singh, B. B., Patel, H. H., Roepman, R., Schick, D., and Ferreira, P. A. (1999). The zinc finger cluster domain of RanBP2 is a specific docking site for the nuclear export factor, exportin-1. *J. Biol. Chem.* *274*, 37370-37378.
- Sloper-Mould, K. E., Jemc, J. C., Pickart, C. M., and Hicke, L. (2001). Distinct functional surface regions on ubiquitin. *J. Biol. Chem.* *276*, 30483-30489.
- Strack, B., Calistri, A., and Gottlinger, H. G. (2002). Late assembly domain function can exhibit context dependence and involves ubiquitin residues implicated in endocytosis. *J. Virol.* *76*, 5472-5479.
- Sukegawa, J., and Blobel, G. (1993). A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm. *Cell* *72*, 29-38.

- Tsai, B., Ye, Y., and Rapoport, T. A. (2002). Retro-translocation of proteins from the endoplasmic reticulum into the cytosol. *Nat. Rev. Mol. Cell Biol.* 3, 246-255.
- VanDemark, A. P., Hofmann, R. M., Tsui, C., Pickart, C. M., and Wolberger, C. (2001). Molecular insights into polyubiquitin chain assembly: crystal structure of the Mms2/Ubc13 heterodimer. *Cell* 105, 711-720.
- Walters, K. J., Kleijnen, M. F., Goh, A. M., Wagner, G., and Howley, P. M. (2002). Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a. *Biochemistry* 41, 1767-1777.
- Wang, B., Jones, D. N., Kaine, B. P., and Weiss, M. A. (1998). High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases. *Structure* 6, 555-569.
- Wilken, N., Senecal, J. L., Scheer, U., and Dabauvalle, M. C. (1995). Localization of the Ran-GTP binding protein RanBP2 at the cytoplasmic side of the nuclear pore complex. *Eur J. Cell Biol.* 68, 211-219.
- Wu, J., Matunis, M. J., Kraemer, D., Blobel, G., and Coutavas, E. (1995). Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *J. Biol. Chem.* 270, 14209-14213.
- Yaseen, N. R., and Blobel, G. (1999). Two distinct classes of Ran-binding sites on the nucleoporin Nup-358. *Proc. Natl. Acad. Sci. U S A* 96, 5516-5521.
- Yokoyama, N., Hayashi, N., Seki, T., Pante, N., Ohba, T., Nishii, K., Kuma, K., Hayashida, T., Miyata, T., Aebi, U., and et al. (1995). A giant nucleopore protein that binds Ran/TC4. *Nature* 376, 184-188.
- Zhang, O., Kay, L. E., Olivier, J. P., and Forman-Kay, J. D. (1994). Backbone <sup>1</sup>H and <sup>15</sup>N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR* 4, 845-858.

## CHAPTER 5

### SUMMARY

Gag is the major structural protein of the HIV-1 virus. Late in the infectious cycle, Gag is synthesized in the cytoplasm as a polyprotein precursor and is then transported to the plasma membrane where the immature virion assembles and buds from the cell. The plasma membrane targeting function of Gag is essential for viral assembly, and the N-terminal MA domain of Gag is largely responsible for Gag membrane targeting and membrane binding (Yuan et al., 1993; Zhou et al., 1994). In addition to its structural roles, Gag also recruits other viral and cellular components into the virion. Recent reports have shown that in late stages of viral assembly, Gag also recruits endosomal proteins of the vacuolar protein sorting pathway to complete viral budding. Ubiquitin has also been shown to play an important role in this process (Garrus et al., 2001; Patnaik et al., 2000; Schubert et al., 2000). Ubiquitylated proteins are usually recognized by downstream proteins through their ubiquitin binding domains, and two putative ubiquitin binding domains UIM and NZF have been identified in various proteins involved in vacuolar protein sorting (and possibly also in HIV-1 budding) (Hofmann and Falquet, 2001; Meyer et al., 2002; Pornillos et al., 2002b). Studies of the interaction between ubiquitin and these ubiquitin binding domains are therefore expected to provide insight into a basic biological process and may also shed light on HIV-1 budding. To learn more about the functions of MA and ubiquitin in HIV-1 viral assembly and budding, I have focused my dissertation research in two areas: 1) the interaction between HIV-1 MA protein and the lipid membrane bilayer and 2) the interaction between ubiquitin and the UIM and NZF ubiquitin binding domains.

MA is myristoylated on its N-terminus and also contains a highly basic region near its N-terminus. Genetic studies have shown that both elements are important for Gag

membrane targeting and membrane binding (Bryant and Ratner, 1990; Gottlinger et al., 1989; Pal et al., 1990; Yuan et al., 1993). I established a biochemical system to study the MA/membrane interaction using purified recombinant proteins. Surface plasmon resonance experiments showed that both a hydrophobic interaction between the myristoyl group and the lipid membrane and an electrostatic interaction between the MA basic region and the negatively charged lipid membrane surface contribute measurably to MA membrane binding. My experiments also demonstrated that Gag-Gag interactions mediated by the CA domain add cooperativity to MA membrane binding and thereby result in higher membrane binding affinities.

Due to the relatively low hydrophobicity of the myristoyl group, the membrane binding of myristoylated proteins is typically reversible and is often controlled by a myristoyl switch mechanism (Kim et al., 1994; Magee and Courtneidge, 1985; Olson et al., 1985; Peitzsch and McLaughlin, 1993). Myristoyl switch proteins have two different conformations: a soluble conformation or “myristoyl in” conformation in which the myristoyl group is sequestered within the protein, and a membrane bound conformation or “myristoyl out” conformation in which the myristoyl group is extruded from the protein and inserts into the lipid membrane bilayer. Studies on HIV-1 MA protein suggested that the membrane binding of MA is also regulated by such a myristoyl switch mechanism (Freed et al., 1994; Paillart and Gottlinger, 1999; Spearman et al., 1997). I developed methods for expressing and purifying recombinant myristoylated HIV-1 MA protein and used NMR spectroscopy to study the structural changes within MA that occur upon myristoylation. Significant changes due to N-myristoylation were observed and were mainly localized to the N-terminal of the protein, especially helices I and II. These

observations suggested that the myristoyl group may pack against this region of MA when it forms the “myristoyl in” conformation. These results are consistent with genetic studies, which indicate that the myristoyl group is sequestered by the  $\alpha$ -helical core of MA when the protein is in solution. The high resolution structure of the myristoylated HIV-1 MA protein remains to be determined in order to understand the molecular details of this myristoyl switch, and this represents the most outstanding question in this field.

I also used NMR spectroscopy to map the binding site of PtdIns(3,4,5)P<sub>3</sub> on MA, as recent reports have shown that PtdInsP can modulate HIV-1 Gag assembly through interactions with MA (Campbell et al., 2001). The MA binding site of PtdIns(3,4,5)P<sub>3</sub> is comprised of the same basic residues that make up the protein’s membrane binding surface. We therefore speculate that MA may use these basic residues to bind the phosphate head groups of PtdIns(3,4,5)P<sub>3</sub> or other PtdInsP and that this interaction may help target Gag protein to the plasma membrane where viral assembly is initiated. However, the specificity and biological function of this intriguing interaction remains to be established.

Ubiquitin, a highly conserved small protein that is found in all eukaryotic cells, plays a central role in a wide variety of cellular processes, including protein degradation, DNA repair, transcription, signal transduction, cell cycle control, and endosomal trafficking (Hershko and Ciechanover, 1998). Recently, it has also been shown that ubiquitin is involved in the process of HIV-1 budding through its function(s) in the vacuolar protein sorting pathway (Garrus et al., 2001; Patnaik et al., 2000; Pornillos et al., 2002b; Schubert et al., 2000). Ubiquitin is covalently attached to the lysine side chains of substrate proteins through a complex process that is catalyzed by a three-enzyme cascade.

The ubiquitylated target proteins are then recognized by downstream effector proteins and targeted to different cellular locations and pathways (Ciechanover, 1994; Hershko and Ciechanover, 1998). Many of these downstream effector proteins, including many of the proteins in the endosomal pathway, have modular domain structures that include conserved ubiquitin binding domains that recognize ubiquitylated substrates as well as additional domains that mediate other cellular functions. A number of such ubiquitin-binding domains have been discovered, including UBA, UEV, UIM, and NZF (Hofmann and Bucher, 1996; Hofmann and Falquet, 2001; Madura, 2002; Meyer et al., 2002). However, the interactions between ubiquitin and these ubiquitin binding domains are not yet understood in molecular detail.

The UIM was initially discovered as a sequence of the S5a/Rpn10 subunit of the proteasome that could bind ubiquitin (Young et al., 1998). UIM sequences are found in a wide variety of proteins, including proteins of the endosomal pathway, such as Hrs, Vps27p, and STAM. NZF is a zinc-containing domain initially identified in Npl4, a protein involved in ER-associated degradation (Meyer et al., 2002). NZF domains are also found in other proteins, including Vps36p, which functions in late endosomal protein sorting and is required for sorting ubiquitylated proteins into the multivesicular body (MVB). I initially studied the interaction between ubiquitin and these two ubiquitin binding domains using surface plasmon resonance biosensor experiments. My results showed that both the UIM and NZF bind ubiquitin directly, making weak, but specific, interactions. UIM and NZF domains from various proteins have been tested and shown to have similar binding affinities, indicating that ubiquitin binding is a general property of UIM and NZF domains. The relatively weak affinities observed *in vitro* may reflect the

fact that the *in vivo* targets of these ubiquitin binding motifs are normally ubiquitin-protein conjugates rather than free ubiquitin.

The ubiquitin binding sites of UIM and NZF have also been mapped using NMR spectroscopy. Both UIM and NZF bind ubiquitin on a hydrophobic surface that surrounds residue I44. Interestingly, this “I44 surface” is also the binding site for the two other known ubiquitin binding motifs, UBA and UEV (Mueller and Feigon, 2002; Pornillos et al., 2002a). Although these four ubiquitin binding motifs share no structural similarities, they nevertheless recognize ubiquitin through the same surface of ubiquitin, highlighting the importance of this surface for ubiquitin function. Not surprisingly, this ubiquitin surface was also recently characterized as being critical for ubiquitin function (Sloper-Mould et al., 2001).

The solution structure of the Npl4 NZF domain was determined using NMR spectroscopy and is described in Chapter 4. The Npl4 NZF domain has a compact structure composed of four short antiparallel  $\beta$ -strands (S1 – S4) and three well ordered loops (L1-L3) that connect the strands. S1, L1, and S2 form the first  $\beta$ -hairpin, which is orthogonal to the second  $\beta$ -hairpin (formed by S3, L3, and S4). The four zinc coordinating cysteine residues are located on loops L1 and L3, which form rubredoxin knuckles. The NZF domain structure is stabilized by a small hydrophobic “core” and an extensive hydrogen bond network. Nearly all of the conserved residues in NZF sequences have structural roles. These include W7, which forms the hydrophobic “core,” and N16, which forms a hydrogen bonding network that bridges strands S2 and S3. The NZF zinc center shares structural similarities with the “ $\beta$ -ribbon” zinc centers seen in a series of transcription-related proteins, such as TFIIIs and RPB9 (Chen et al., 2000; Qian et al.,

1993a; Qian et al., 1993b; Wang et al., 1998), although there are no obvious functional connections between NZF-containing proteins and these proteins. The similarities may instead reflect the fact that this common fold is a favorable structure for zinc binding.

NMR chemical shift perturbation mapping experiments showed that the ubiquitin binding site on NZF is composed of residues from the loops (L1 and L3) above the zinc coordination site. Like the NZF binding site on ubiquitin, the ubiquitin binding site on NZF is primarily hydrophobic, showing that the NZF ubiquitin interaction is mainly mediated by hydrophobic interactions. Sequence conservation analyses indicated that residue 13 (threonine) and 14 (a large hydrophobic residue,  $\Phi$ ) are highly conserved in NZF motifs that interact with ubiquitin. Residues 13 and 14 also showed the most significant chemical shift changes upon ubiquitin binding, indicating that the T $\Phi$  dipeptide sequence forms the major recognition site for ubiquitin.

The next important question in this area is to understand precisely how the NZF domain binds to ubiquitin. My work on the Npl4 NZF domain has helped to make this a tractable problem, and my collaborator, Steven L. Alam, is now well on his way to determining the solution structure of the Npl4 NZF/ubiquitin complex.

## References

- Bryant, M., and Ratner, L. (1990). Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc. Natl. Acad. Sci. U S A* 87, 523-527.
- Campbell, S., Fisher, R. J., Towler, E. M., Fox, S., Issaq, H. J., Wolfe, T., Phillips, L. R., and Rein, A. (2001). Modulation of HIV-like particle assembly in vitro by inositol phosphates. *Proc. Natl. Acad. Sci. U S A* 98, 10875-10879.
- Chen, H. T., Legault, P., Glushka, J., Omichinski, J. G., and Scott, R. A. (2000). Structure of a (Cys3His) zinc ribbon, a ubiquitous motif in archaeal and eucaryal transcription. *Protein Sci.* 9, 1743-1752.

- Ciechanover, A. (1994). The ubiquitin-proteasome proteolytic pathway. *Cell* *79*, 13-21.
- Freed, E. O., Orenstein, J. M., Buckler-White, A. J., and Martin, M. A. (1994). Single amino acid changes in the human immunodeficiency virus type 1 matrix protein block virus particle production. *J. Virol.* *68*, 5311-5320.
- Garrus, J. E., von Schwedler, U. K., Pornillos, O. W., Morham, S. G., Zavitz, K. H., Wang, H. E., Wettstein, D. A., Stray, K. M., Cote, M., Rich, R. L., *et al.* (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* *107*, 55-65.
- Gottlinger, H. G., Sodroski, J. G., and Haseltine, W. A. (1989). Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U S A* *86*, 5781-5785.
- Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. *Annu. Rev. Biochem.* *67*, 425-479.
- Hofmann, K., and Bucher, P. (1996). The UBA domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biochem. Sci.* *21*, 172-173.
- Hofmann, K., and Falquet, L. (2001). A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem. Sci.* *26*, 347-350.
- Kim, J., Blackshear, P. J., Johnson, J. D., and McLaughlin, S. (1994). Phosphorylation reverses the membrane association of peptides that correspond to the basic domains of MARCKS and neuromodulin. *Biophys. J.* *67*, 227-237.
- Madura, K. (2002). The ubiquitin-associated (UBA) domain: on the path from prudence to prurience. *Cell Cycle* *1*, 235-244.
- Magee, A. I., and Courtneidge, S. A. (1985). Two classes of fatty acid acylated proteins exist in eukaryotic cells. *EMBO J.* *4*, 1137-1144.
- Meyer, H. H., Wang, Y., and Warren, G. (2002). Direct binding of ubiquitin conjugates by the mammalian p97 adaptor complexes, p47 and Ufd1-Npl4. *EMBO J.* *21*, 5645-5652.
- Mueller, T. D., and Feigon, J. (2002). Solution structures of UBA domains reveal a conserved hydrophobic surface for protein-protein interactions. *J. Mol. Biol.* *319*, 1243-1255.
- Olson, E. N., Towler, D. A., and Glaser, L. (1985). Specificity of fatty acid acylation of cellular proteins. *J. Biol. Chem.* *260*, 3784-3790.

- Paillart, J. C., and Gottlinger, H. G. (1999). Opposing effects of human immunodeficiency virus type 1 matrix mutations support a myristyl switch model of gag membrane targeting. *J. Virol.* *73*, 2604-2612.
- Pal, R., Reitz, M. S., Jr., Tschachler, E., Gallo, R. C., Sarngadharan, M. G., and Veronese, F. D. (1990). Myristoylation of gag proteins of HIV-1 plays an important role in virus assembly. *AIDS Res. Hum. Retroviruses* *6*, 721-730.
- Patnaik, A., Chau, V., and Wills, J. W. (2000). Ubiquitin is part of the retrovirus budding machinery. *Proc. Natl. Acad. Sci. U S A* *97*, 13069-13074.
- Peitzsch, R. M., and McLaughlin, S. (1993). Binding of acylated peptides and fatty acids to phospholipid vesicles: pertinence to myristoylated proteins. *Biochemistry* *32*, 10436-10443.
- Pornillos, O., Alam, S. L., Rich, R. L., Myszka, D. G., Davis, D. R., and Sundquist, W. I. (2002a). Structure and functional interactions of the Tsg101 UEV domain. *EMBO J.* *21*, 2397-2406.
- Pornillos, O. P., Garrus, J. E., and Sundquist, W. I. (2002b). Mechanisms of enveloped RNA virus budding. *Trends Cell Biol.* *12*, 569-579.
- Qian, X., Gozani, S. N., Yoon, H., Jeon, C. J., Agarwal, K., and Weiss, M. A. (1993a). Novel zinc finger motif in the basal transcriptional machinery: three-dimensional NMR studies of the nucleic acid binding domain of transcriptional elongation factor TFIIS. *Biochemistry* *32*, 9944-9959.
- Qian, X., Jeon, C., Yoon, H., Agarwal, K., and Weiss, M. A. (1993b). Structure of a new nucleic-acid-binding motif in eukaryotic transcriptional elongation factor TFIIS. *Nature* *365*, 277-279.
- Schubert, U., Ott, D. E., Chertova, E. N., Welker, R., Tessmer, U., Princiotta, M. F., Bannink, J. R., Krausslich, H. G., and Yewdell, J. W. (2000). Proteasome inhibition interferes with gag polyprotein processing, release, and maturation of HIV-1 and HIV-2. *Proc. Natl. Acad. Sci. U S A* *97*, 13057-13062.
- Sloper-Mould, K. E., Jemc, J. C., Pickart, C. M., and Hicke, L. (2001). Distinct functional surface regions on ubiquitin. *J. Biol. Chem.* *276*, 30483-30489.
- Spearman, P., Horton, R., Ratner, L., and Kuli-Zade, I. (1997). Membrane binding of human immunodeficiency virus type 1 matrix protein in vivo supports a conformational myristyl switch mechanism. *J. Virol.* *71*, 6582-6592.
- Wang, B., Jones, D. N., Kaine, B. P., and Weiss, M. A. (1998). High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases. *Structure* *6*, 555-569.

Young, P., Deveraux, Q., Beal, R. E., Pickart, C. M., and Rechsteiner, M. (1998). Characterization of two polyubiquitin binding sites in the 26 S protease subunit 5a. *J. Biol. Chem.* *273*, 5461-5467.

Yuan, X., Yu, X., Lee, T. H., and Essex, M. (1993). Mutations in the N-terminal region of human immunodeficiency virus type 1 matrix protein block intracellular transport of the Gag precursor. *J. Virol.* *67*, 6387-6394.

Zhou, W., Parent, L. J., Wills, J. W., and Resh, M. D. (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. *J. Virol.* *68*, 2556-2569.