Structural and biophysical studies of HIV-1 Rev protein

By
Ye Peng

APPROVED BY THE SUPERVISORY COMMITTEE

Robert O. Fox, Ph.D.

Miles W. Cloyd, Ph.D.

Edmund W. Czerwinski, Ph.D.

Vince J. Hilser, Ph.D.

Susan Marriott, Ph.D.

Dean, Graduate School

Structural and biophysical studies of HIV-1 Rev protein

by

Ye Peng, B.S.

Dissertation
Presented to the Faculty of the University of Texas Graduate School of
Biomedical Sciences at Galveston
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Approved by the Supervisory Committee

Robert O. Fox, Ph.D.
Vince J. Hilser, Ph.D.
Edmund W. Czerwinski, Ph.D.
Miles W. Cloyd, Ph.D.
Susan Marriott, Ph.D.

March, 2006
Galveston, Texas

Key Words: HIV-1, Rev, B23

To my wife and my daughter

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my mentor, Dr. Robert O. Fox, for his guidance, generosity, commitment to my education and research work. I would also like to thank my committee for their guidance throughout this process: Dr. Vince J. Hilser, Dr. Edmund W. Czerwinski, Dr. Miles W. Cloyd and Dr. Susan Marriott. Many thanks to my friends and co-workers in Dr. Fox's lab: Dr. Munia Mukherjee in helping me to do NMR experiment and spectral analysis, Dr. Xiuzhen Fan of great ideas on molecular biology and mass spectrometry , Deqian Liu in protein purification, Dr. Yonghong Zhao of great discussion. Special thanks to lab manager, Warna D. Kaluarachchi and Dr. Fox's secretary Yvette J Boyd.

I also would like to thank Dr. David W. Bolen, Dr. James C. Lee for kindly allowing me to use their equipments to finish my CD and analytical ultracentrifugation experiments. A special thanks goes to Dr. Luis Marcelo F. Holthauzen for teaching me useful techniques that have been a great help for my research, and Dr. Christopher Chin in helping me to do SV and SE experiments and analyze the data.

I extend special thanks to Biochemistry and Molecular Biology program director Dr. Lillian Chan and Program Coordinator Ms. Debora M. Botting for all their help in many ways.

Finally, I would especially like to thank my family, especially my wife, Xiaoming Hu, my daughter Katharine and my mother for their love, support and encouragement.

The structural and biophysical study of HIV-1 Rev protein and its interactions with human B23

Publication No. _____
Ye Peng, Ph.D.

The University of Texas Medical Branch at Galveston, January, 2006

Supervisor: Robert O. Fox

Rev is a key HIV-1 virus regulatory protein, which exports the unspliced and partially spliced viral mRNA out of the nucleus, thereby inducing the switch from the early phase to the late phase of the viral replication cycle. Because it is prone to form filaments at very low concentration, the X-ray crystallography or NMR approaches to solving Rev structure are very difficult. In my dissertation studies, I have designed several structural mutants of Rev. Under physiological conditions, I found RevC mutant behaves as a natively unfolded monomer in solution. The segmentally disordered character of RevC was supported by CD and HSQC studies. Another structural Rev mutant is the Rev loop deletion. I used a Gly-Ser-Gly-Ala linker to take the place of the Rev loop and most of the Arg-rich domain. This mutant can stay in solution as oligomers in sub-millimolar concentration for NMR studies. The HSQC spectra of the Rev loop deletion and the RevC can be well overlaid except the HSQC of Rev loop deletion which has less peaks under the same threshold. It brings a filament model with a thin wall formed by Rev N-terminal two helixes and a flexible Rev C-terminal freely moving inside. The flexible character of Rev C-terminal is also proved to be true for the Rev wild type in comparing the HSQC of Rev wild type and RevC. The HSQC of Rev wild type was obtained from 15N labeled Rev and 132B23 complex. B23 is a major cellular component which binds Rev. I designed an N-terminal core of B23, 132B23, and found that it binds with Rev and forms a decamer with Rev and 132B23 heterodimer.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

.

LIST OF ABBREVIATIONS

AIDS	Acquired Immune Deficiency Syndrome
ATP	Adenosine triphosphate
AUC	Analytical ultracentrifugation
BME	2-mercaptoethanol
CD	Circular Dichroism
CRM1	Chromosome Region Maintenance 1
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
eIF5A	eukaryotic translation initiation factor 5A
ESI-MS	ElectroSpray Ionization Mass Spectrometry
FID	Free Induction Decay
GDP	Guanosine 5'-Diphosphate
GTP	Guanosine 5'-Triphosphate
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethane sulphonic acid

HIV-1	Human Immunodeficiency Virus-1
HSQC	Heteronuclear single quantum coherence
IPTG	Isopropyl-beta-D-thiogalactopyranoside
kDa	Kilodalton
LB Media	Luria-Bertani Media
MALDI	matrix-assisted laser desorption/ionization
NMR	Nuclear magnetic resonance
NPC	Nuclear Pore Complex
PCR	Polymerase chain reaction
PDB	Protein data bank
PMSF	Phenyl-methylsulfonyl fluoride
RanGAP1	Ran GTPase-activating protein
RCC1	Ran GTPase Guanine Exchange Factor
Rev	Regulator of expression of virion proteins
RNA	Ribonucleic acid
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPR	Surface Plasmon Resonance
TCEP	Tris(2-carboxyethyl)phosphine

TEMPO	2,2,6,6-Tetramethylpiperidyl-1-Oxyl
TEV	Tobacco Etch Virus
TFA	Trifluoroacetic acid

CHAPTER 1: INTRODUCTION OF REV PROTEIN

Human Immunodeficiency Virus-1 (HIV-1) is the etiological agent of Acquired Immune Deficiency Syndrome (AIDS) (Weiss *et al.* 1986). Based on UNAIDS/WHO, in 2005, there are about 40.3 million people living with AIDS, and about 3.1 million deaths caused by it. Everyday, there are another 14,000 new infections.

But the HIV-1 vaccine development encounters a considerable barrier because of the extensive viral sequence variation of HIV-1 isolates (Marx 1989). The variation in HIV-1 is caused by the error-prone reverse transcriptase and by recombination between different viral strains. The difficulties in developing the vaccine to the viral envelope proteins drive other approaches to prevent the disease. Can the physiological stages of the viral life cycle be disturbed? HIV-1 infects cells in the immune system and the central nervous system. It begins its infection by binding to the CD4 receptor on the host cell, mainly the T helper lymphocyte (McClure *et al.* 1987). After the fusion of the virus with the host cell, the viral RNA is released and undergoes reverse transcription into DNA with the help of viral reverse transcriptase. Once this viral DNA is made, it enters the host cell nucleus and the viral integrate helps the DNA to integrate into the cell's genome. After the viral DNA is integrated into the host's genome, HIV-1 may stay in a latent state for many years. This ability of HIV-1 to persist in certain latently infected cells is the major barrier to eradication or cure of HIV-1. If the activation of the host cells happens, the latent HIV-1 starts the transcription of viral DNA into messenger RNA (mRNA), and then translated into viral proteins. One of the viral proteins, HIV-1 protease, is required to process other HIV-1 polyproteins into their functional forms. At the same time, the new viral RNA with the genetic material of the next generation of viruses forms. Finally, the viral RNA and viral proteins assemble at the cell membrane into a new virus, and bud out of the cell (Coffin 1996; Himathongkham and Luciw 1996). There are three virally encoded proteins, Tat, Nef and Rev, involved in the information transduction with the

cell (Robert-Guroff *et al.* 1990). These viral regulation proteins are responsible for the viral physiology, and thus they are the optimum targets for the therapeutic purpose to control the viral propagation. The Rev protein is the key component of viral life cycle. The biophysical characterization of Rev and the interactions with other cell factors should help us to further understand its unique role in the viral propagation and is the first step in a rational drug design process aimed at blocking the viral life cycle.

1.1 The essential role of Rev protein in the HIV-1 life cycle:

HIV-1 belongs to the lentivirus subfamily of retrovirus. All retroviruses utilize the 5' long terminal repeat (LTR) as promoter to synthesize the full-length viral mRNA. Over 30 different viral mRNA species, derived from a single primary transcript as a precursor, are found in the HIV-1 infected cells. At least four different 5' splice sites and eight 3' splice sites are found (Felber *et al.* 1990; Guatelli *et al.* 1990; Purcell and Martin 1993). The various viral mRNAs could be separated into three main groups based on the splicing levels. These RNA molecules include unspliced (about 9kb) transcripts which encode Gag and Pol; many partially spliced species (about 4kb) encode Env, Vif, Vpr or Vpu; and many fully spliced species (about 2kb) encode Rev, Tat or Nef (Muranyi and Flugel 1991). The diversity of HIV-1 mRNAs is produced by the alternative splicing of the full-length transcript (Legrain and Rosbash 1989). Under normal circumstances, viral RNAs are retained in the nucleus, and they are fully spliced by the splicing factors or completely degraded. One of the fully spliced viral mRNAs can be translated into Rev protein. The unspliced and partially spliced viral RNAs are carried out of the nucleus by interacting with Rev protein (Chang and Sharp 1989). In the early phase of the infective cycle, extensive processing of the primary transcript generates a family of fully spliced mRNA of about 2kb that encode the viral trans-acting regulatory proteins, Tat, Nef and Rev. Tat is a transcriptional activator of the viral long terminal repeat (LTR) promoter (Jones and Peterlin 1994; Jones *et al.* 1997). Nef may reduce the surface expression of CD4, the

major receptor for HIV-1, and down regulate the MHC class I, preventing the infected cells from being lysed by cytotoxic T lymphocytes (Peter 1998).

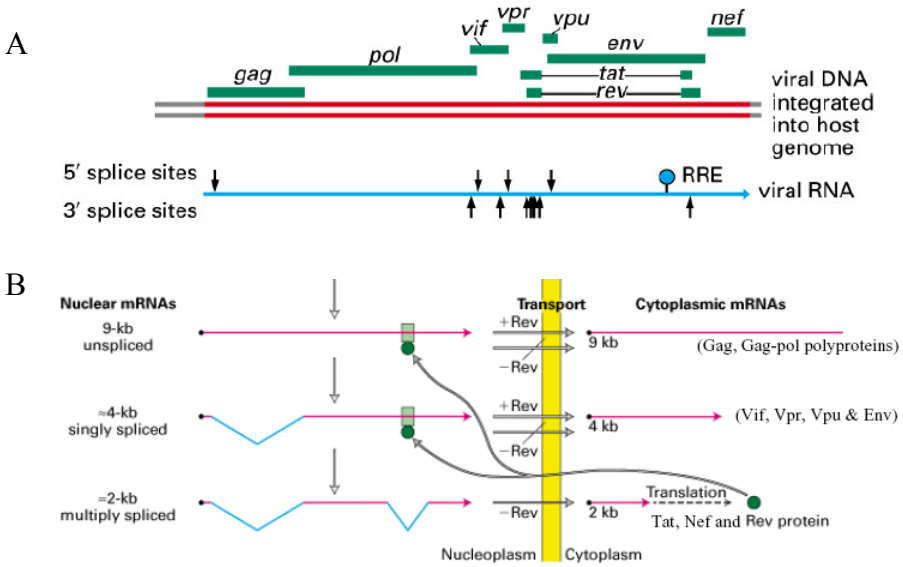


Figure 1.1 HIV-1 Rev functions as viral RNA export factor. (A) The HIV-1 proteins are translated from 30 alternatively spliced mRNAs. There are four 5' splice sites and nine 3' splice site in the primary 9kb viral mRNA. (B) Fully spliced mRNAs (2kb) only express regulating proteins (Tat, Nef and Rev). The unspliced and singly spliced viral mRNAs have to bind to Rev with RRE to translocate from nucleoplasm to cytoplasm (*Adapted from Flint S.J. principles of Virology Chapter 10 Figure10.15*).

Rev is a key HIV-1 regulatory protein produced early in infection. It enables the unspliced and partially spliced viral mRNA to export from the nucleus, thereby inducing a switch to the late phase of the viral replication cycle (Hanly *et al.* 1989;Malim and Cullen 1993). Rev binds to the RRE region located in the env gene via a specific interaction between its arginine-rich motif and the RRE phosphate groups (Felber *et al.* 1989). Because Rev has both a nuclear export signal and a nuclear localization signal, it functions as a nucleocytoplasmic shuttle protein (Figure 1.1), Rev conveys the viral singly spliced mRNA, which encode the Gag, Pol and Env polyproteins, and viral unspliced RNA, which is packaged as the genomic RNA, out through the nuclear pore

(Hope 1999). As Rev is the key protein in the HIV-1 infective cycle, it becomes an important target for anti-HIV-1 strategies.

1.2 The domains of HIV-1 Rev protein:

Rev (regulator of expression of virion proteins) is a small (13 kDa), basic (pI=9.2) protein with a high affinity for RNA (Cochrane *et al.* 1989). The extensive site-directed mutagenesis of the 116 amino acid HIV-1 Rev protein has defined two main functional domains and additional regions which are necessary for the function (Hope 1990). The amino-terminal domain is a basic, arginine-rich motif (aa 35-50), which is responsible for the nuclear and nucleolar localization of Rev as well as specific binding of Rev to RRE (Rev response element) (Kubota *et al.* 1989; Malim *et al.* 1989). The Circular Dichroism (CD) spectra of Rev wild type and Rev truncations show that Rev contains about 50% α-helical structure, mainly positioned within the N-terminal (Wingfield *et al.* 1991), and it is proved by epitope mapping experiments using a panel of monoclonal antibodies (Jensen *et al.* 1997). The proline-rich segment, PPPNPEG, is in the middle of the domain and is supposed to form a loop. Together with the structural mutants of Rev, the N-terminus domain was indicated to form a potential helix-loop-helix motif (Auer *et al.* 1994). By comparison with numerical simulations, helical φ and ψ angles were shown at residues Leu13 and Val16 in the predicted helix 1 segment and at residues Arg39, Arg 42, Arg43, and Arg44 in the predicted helix 2 segment. The hydrophobic residues of H1 Leu13, Val16, Ile19 or Ala15, Leu18, Phe21, respectively, may form a classical helix-helix contact to the H2 hydrophobic residues Ile55, Ile59, where the ridges and the grooves are made by residues separated in sequence by three (i + 3) / four (i + 4). This knob and hole pattern results in an angle of about 20° between the two helices. Helix 2 (H2) contains a short arginine-rich RNA binding domain which specifically binds to RRE-IIB (Blanco *et al.* 2001). The stable complex structure of the 23aa H2 peptide binding with 34b RRE-IIB was solved by NMR. The α-helical peptide binds the major groove of the RNA through base-specific, phosphate backbone and van der Waals

contacts. Four conserved amino acids (Arg35, Arg39, Asn40 and Arg44) are essential in base-specific interactions with the RNA major groove (Battiste *et al.* 1996; Gosser *et al.* 2001).

The other essential leucine-rich domain at the C-terminus (aa 73-84) functions as nuclear export signal of Rev (NES). This domain is recognized by a cellular transport receptor, human CRM1 (also known as exportin 1), which is one essential component of the receptor-mediated export pathway, and helps the Rev-RRE RNA ribonucleoprotein complex out of the nucleus (Marques *et al.* 2003; Hakata *et al.* 2001). This is also a crucial domain for Rev transactivation function *in vivo*. Mutations of essential amino acids within this domain produces a trans-dominant negative phenotype in culture that inhibits the nuclear export of coexpressed wild-type Rev in a dose-dependent manner, which may be caused by the mutant and wild-type Rev form a functionally inactive and export-deficient heteromers (Fischer *et al.* 1995). The C-terminal part of Rev displays only nonalpha helical secondary structural elements, in a circular dichroism study. Therefore the leucine rich activation domain in HIV-1 Rev may only form a β-strand or loop related structural contact with its cellular cofactor. Several Rev C-terminal peptides contain NES behavior similar with respect to solution conformation, concentration dependence and variations in ionic strength and pH. Temperature studies revealed an unusual induction of β-structure with rising temperatures in all activation domain peptides (Thumb *et al.* 1999).

1.3 The Rev nuclear transport cycle

The viral gene expression can be separated into two stages - early and late. The transition from early stage to late stage is achieved by Rev as a trans-activator in binding ~9kb and ~4kb mRNA and helping them to get out of nucleus. In HIV-1 early stage, one of the fully spliced mRNAs (~2kb) encodes Rev, and Rev is translated in the cytoplasm. The Rev nuclear transport cycle is initiated by nuclear import. Recent studies have implicated importin-β as the receptor that mediates the import of Rev (Gorlich 1997).

Interestingly, Rev binds directly to importin-β without the help of importin-α which is commonly needed in nuclear import (Henderson and Percipalle 1997). When we compare the sequences between the arginine-rich NLSs of Rev and importin-β-binding (IBB) domain of importin-α, we find that these domains are very similar. Therefore, Rev utilizes importin-β in the absence of a bridging factor like importin-α. After translocation into the nucleus, the binding of importin-β with RanGTP is thought to induce disassembly of the importin-β/Rev complex and releases Rev into the nucleoplasm (Palmeri and Malim 1999). The dissociation of Rev results in the protein available for binding to the RRE (Zapp and Green 1989). Acturally, the arginine-rich NLS of Rev also functions as the RRE binding domain. Rev first binds to its primary site in stems IIB and IID of viral RNA's RRE region (Daly *et al.* 1989). Additional copies of Rev could bind to additional sites within the RRE by a series of cooperative Rev-Rev and Rev-RNA interactions (Bartel *et al.* 1991; Olsen *et al.* 1990).

Similar to nuclear import, the nuclear export includes migration to the nuclear face of the NPC (Nuclear Pore Complex), docking, interactions between CRM1 (Chromosome Region Maintenance 1) and FG-repeat nucleoporins (including Rip/Rab), translocation, and disassembly of the complex in the cytoplasm. The CRM1 (also known as exportin 1) is the export receptor for Rev (Stade *et al.* 1997; Fukuda *et al.* 1997). Although monomeric leucine-rich NESs may serve as a ligand for CRM1, the binding of a single Rev to the RRE is incapable of activating RNP export in somatic cells (Venkatesh *et al.* 1990). Multimerization of Rev is required to export the unspliced and partially spliced viral RNAs (Thomas *et al.* 1998). Thus Rev function is nonlinear with respect to the intracellular concentration of Rev. In the nucleus, Rev forms a complex with the RRE, CRM1 and Ran-GTP. Once the appropriate Rev/RRE/CRM1/Ran-GTP complexes are assembled, they are presumably targeted for efficient export through the NPC. The export of the Rev/RRE/CRM1 complex requires the nucleoporins Nup214 and Nup98 (Zolotukhin and Felber 1999). In somatic cells, Rev facilitates the export of only nascent transcripts (Malim and Cullen 1993). Conversion of Ran-GTP in this complex to Ran-

GDP leads Ran to dissociate from the exported complex (Dahlberg and Lund 1998). After the dissociation in the cytoplasm, Rev returns to the nucleus for a subsequent round of export.

This cycle of factors shuttling continuously between the nucleus and the cytoplasm generates a system of which the small amounts of Rev presenting in an HIV-infected cell have the capacity to mediate the export of significant amounts of unspliced and partially spliced viral RNA (Figure 1.2).

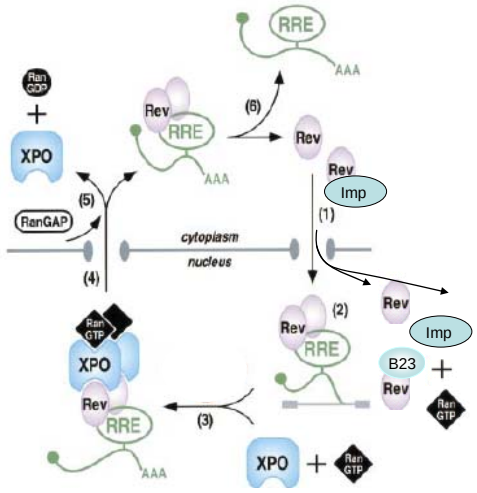


Figure 1.2 Rev cycle (1) Rev-Importin beta complex imports to the nucleus (2) Rev associates RRE (3) Rev-RRE-Exportin 1(CRM1)-RanGTP export complex forms (4) export complex exports from the nucleus (5) RanGTP hydrolyzes to RanGDP (6) Rev dissociates from RRE

1.4 The interactions between Rev and its cellular partners

As mentioned above, Rev basically contains two domains. In the nucleus, Rev multimerizes on the viral mRNA RRE region via its N-terminal RNA-binding domain.

This complex, together with Ran in its GTP binding form, further forms the export complex in binding to CRM1 through the Rev C-terminal nuclear export signal (NES), also known as the activation domain. Mutation of any of three leucines (Leu78, 81, 83) within this domain would not only abolish Rev function, but also block any of the coexpressed wild-type Rev's activity (Mermer *et al.* 1990). As an essential component of the export complex, Exportin1 directly binds to Leucine-rich NES domain. Exportin1 is an 112kD soluble protein, and it belongs to the importin β family, in which all proteins contain an N-terminal Ran-binding domain (Fornerod *et al.* 1997). The 25kD Ran is a GTPase, and exists in two forms, RanGTP and RanGDP. The reaction of RanGTP to RanGDP could be stimulated by the Ran GTPase-activating protein, RanGAP1. The exchange factor, RCC1, could trigger the reaction of RanGDP to RanGTP. Because RanGAP1 is mainly located in the cytoplasm, the majority of RCC1 is associated with chromatin, RanGDP exists mainly in the cytoplasm, and RanGTP is primarily found in the nucleus (Richards *et al.* 1997). Exportin1 forms a RNP complex with Rev-RNA and RanGTP, but not with RanGDP (Fornerod M, 1997). Compared with other NES-containing proteins, Rev shows a lower cooperativity in forming NES-proein/Exportin1/RanGTP complex. The relative low affinity of Rev binding with Exportin1 could explain why the steady-state localization of Rev is in the nucleus. When several Rev molecules bind to viral RNA and the local Rev concentration increases, the association rate of Rev with Export1 and RanGTP would also increase, the export rate could thus accelerate (Askjaer *et al.* 1999). After the RNP complex translocates through the nuclear pore complex (NPC), RanGAP1 promotes the hydrolysis of RanGTP, and releases Rev-RRE RNA.

Another Rev NES binding protein is eukaryotic translation initiation factor 5A (eIF5A), a small 154-aa protein with a molecular mass of 16.8 kDa and ~5.0 pI (Ruhl *et al.* 1993). eIF5A is the only protein known to contain a hypusine residue, which is formed by posttranslational modification of the lysine residue at position 50. Inhibitors that block this modification result in cell growth arrest and apoptosis (Park *et al.* 1993).

Some studies have found that if the expression of eIF-5A is blocked, the Rev nuclear export is inhibited (Andrus *et al.* 1998), although contrary results have also been reported (Fischer *et al.* 1994). Thus the direct effect of eIF-5A in the Rev nuclear export process is still unclear.

During Rev nuclear import process, importin-β is an essential partner. Importin-β is a large protein with MW about 97kD. It has three major binding activities: Ran-, IBB- (importin-β binding) and NPC (Cingolani *et al.* 2002). The Ran binding domain of importin-β is located at the N-terminus, and IBB binding domain is at the C-terminus. Unlike Exportin1, importin-β binds the substrate without binding RanGTP. After docking and translocating, importin-β releases NLS-substrate by associating RanGTP. Importin-β contains 19 tandem HEAT motifs. These HEAT repeats share a common core of 21 residues, and form a helical hairpin structure. Further, these tandemly repeated HEAT motifs are assembled into an elongated, right-handed superhelix. Within the HEAT motifs, the hydrophobic residues are conserved in positions 9, 10, 13 and 17 in helix A and at positions 24, 28, 31, 32, 35 and 36 in helix B. These residues are involved in packing interactions within and between HEAT repeats. Furthermore, the importin-β is flexible (Lee *et al.* 2000). From the structure of importin-β-IBB complex, we can see the IBB is embraced by 8 importin-β HEAT repeats. The interaction between importin-β and IBB possibly involves a conformation change of importin-β (Cingolani *et al.* 1999). The protease-resistant assay suggests importin-β adopts a more open conformation when not bound to the IBB. Rev could mimic IBB domain and directly bind to importin-β (Henderson and Percipalle 1997). But the affinity and exact binding site still need to be clearified.

Another protein, shown directly interacting with Rev, is the B23 protein (Fankhauser C, 1991). B23 is an acidic protein with MW of 32kD. A nucleolar protein B23 shuttles between the nucleolus and cytoplasm, and could form a stable complex with the NLS domain of Rev (Szebeni and Olson 1999), and the complex may be dissociated by the

RRE-containing RNA. The two highly acidic segments of B23 may be involved in this interaction. B23 alone can facilitate the import process, rather than carry Rev protein into the nucleolus from the cytoplasm (Szebeni *et al.* 1997). So, the high affinity of nucleolar protein B23 for Rev could be benefited not in the import process, but in helping Rev from aggregating as chaperon.

1.5 Rev self-assembly

Rev undergoes self-association both *in vitro* (Malim and Cullen 1991) and *in vivo* (Madore *et al.* 1994). The oligomerization is thought to be associated with the RRE binding and Rev nuclear transport cycle. When the Rev concentration is above the critical concentration of 80 μg/ml, Rev reversibly assembles into unbranched filaments with an outer diameter of about 15 nm and an axial channel of 5–7 nm diameter. The filaments associate with each other to form aggregates (Watts *et al.* 1998). Based on equilibrium analytical centrifugation experiments, Rev could remain predominantly monomeric only at concentrations below approximately 10μM in the absence of RNA. The equilibrium analytical centrifugation data fit well to an unlimited isodesmic self-association model with an association constant for the additional monomer to aggregate as $K=10^6M^{-1}$ (Surendran *et al.* 2004). In the presence of RNA, Rev rapidly forms poorly ordered filaments about 8 nm in diameter and of a length approximately proportional to the size of the RNA (Heaphy *et al.* 1991). Because of these polymerizations, it is hard to solve the structure of Rev either by X-ray crystallography or by NMR spectroscopy.

An elegant genetic selection was used to identify Rev mutants with deficiencies in the Rev multimeric-assembly pathway (Jain and Belasco 2001). A reporter gene could only be expressed when two Rev proteins cooperatively bind to their RRE stem–loop IIB sites. Three classes of multimerization defects were found. Class three mutants exhibit defects in binding the RRE either as a monomer or as a multimer and are probably structurally defective. Class one Rev mutants bind to the RRE as monomers, but are defective in their ability to form dimers. Most interestingly, class two mutants are competent for

dimerization and RNA binding, but show greatly reduced multimerization properties. Thus, the Rev oligomerization could be separated into two regions. The refined molecular model suggests that there are two Rev-interaction surfaces, which are both within the helix-loop-helix region. One surface is required for Rev–Rev dimerization (note as tail-tail interaction, residues involved including Leu-18, Ile-55), the other surface is required for trimerization and higher-order assembly (note as head-head interaction, residues involved including Leu-12, Val-16, Leu-60).

The multimer formation of Rev is affected when the amino acids in the flanking region of the basic region (Bogerd and Greene 1993; Jensen *et al.* 1997), or in the NES domain (Madore *et al.* 1994) are mutated. But the map of Rev sequences involved in the multimerization process is far from clear because of many conflicting results coming from different assays.

CHAPTER 2: THE CONSTRUCTS OF REV WILD TYPE AND THE REV STRUCTUAL MUTANTS

2.1 Materials

The plasmid pREV1 (Jain and Belasco 1996), a T7 promoter-containing pSGA04 plasmid encoding the wild-type HIV-1 Rev protein with a hexa-histidine tag at the N-terminus, was used as source of constructs of both wild-type protein and as a template for generating mutants of Rev. The plasmid was a kind gift of Dr. Belasco.

The Sequence of Rev gene was confirmed by the sequencing result from the Sealy Center for Molecular Science:

```
GGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACACATG
                                                NdeI
GGCCATCATCATCATCATCATAGCAGCGGCCTGTTTAAACGACATATGGCTGGTCGCTCT
                                                  M  A  G  R  S
GGCGATTCTGATGAAGACCTTCTCAAAGCCGTTCGCCTGATCAAGTTTCTCTACCAGAGC
 G  D  S  D  E  D  L  L  K  A  V  R  L  I  K  F  L  Y  Q  S
AACCCTCCGCCAAACCCAGAGGGTACTCGCCAGGCGCGTCGCAACAGGCGTCGACGGTGG
 N  P  P  P  N  P  E  G  T  R  Q  A  R  R  N  R  R  R  R  W
CGCGAACGTCAGCGTCAAATCCACAGCATTTCCGAGCGCATTCTGAGCACTTACCTCGGC
 R  E  R  Q  R  Q  I  H  S  I  S  E  R  I  L  S  T  Y  L  G
CGTTCTGCTGAGCCCGTCCCACTTCAGTTGCCTCCCTTAGAGCGTTTAACTCTAGACTGC
 R  S  A  E  P  V  P  L  Q  L  P  P  L  E  R  L  T  L  D  C

AACGAGGATTGTGGAACTTCTGGGACGCAGGGTGTTGGAAGCCCTCAAATACTAGTGGAG
 N  E  D  C  G  T  S  G  T  Q  G  V  G  S  P  Q  I  L  V  E
TCCCCGACTGTTCTGGAGTCTGGAACCAAAGAGTAGTAAGCGGTCGACCTCAGGCATTTG
 S  P  T  V  L  E  S  G  T  K  E  -  -
AGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACAT
AAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATT
TCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCA
CCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAA
TTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGC
CAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGATCGATCAGGCCT
                                                NcoI
GCTAGCATGAC
```

The primers designed for Rev constructs were synthesized by Sigma Genosys (The Woodlands, TX). The PCR buffer and Taq DNA polymerase, dNTPs are from Fisher Scientific; the T4 DNA ligase is from Gibco BRL; the restrict enzymes for DNA cleavage are from NewEngland Biolabs; the QuikChange Site-Directed Mutagenesis Kit is from Stratagene. The PCR reactions were carried out by ERICOMP Twinblock easycycler series PCR machine. The DNA products were purified by QIAGEN miniprep, PCR and Gel purification kits. The Subcloning Efficiency DH5 Competent Cells were from Invitrogen.

2.2 Experiments and Methods

General PCR and ligation reactions:

PCR reaction component	volume
forward primer (~120 pmol/µl)	3 µl
reverse primer (~120 pmol/µl)	3 µl
25 mM dNTP	2.5 µl
10 X PCR buffer ([Mg2+] =15mM)	5 µl
DNA template (~ 0.032 µg/µl)	1 µl
deionized water	34.5 µl
Taq DNA polymerase (1u/µl)	1 µl
Total	50 µl

PCR program

Step 1 Initial denaturation at 94°C for 1 minute;

Step 2 Denaturation at 94°C for 30 seconds;

Annealing at 55°C for 30 seconds;

Extension at 72°C for 2 minutes;

Repeat step 2 for 30-35 cycles;

Step 3 Final extension at 72°C for 10 minutes;

Step 4 20°C hold

Ligation reaction

Digested vectors	4 μl
Sticky-end PCR products	2 μl
10X ligation buffer	1 μl
T4 DNA ligase	1 μl
deionized H2O	2 μl
total	10 μl

2.3 Generate TEV-cleavage site containing Rev plasmid

The original Rev plasmid contains an Amp resistant gene. This vector comes from pSG003 (which was constructed based on pET11a) (4933bp) (Ghosh and Lowenstein1996), an engineered vector for general expression of foreign protein in *E. coli*. The Leu-Phe-Lys-Arg segment in the N-terminus is the recognition site for Kex2

protease. But the cleavage can not go to completion and the enzyme is extremely expensive. Recently, Tobacco Etch Virus (TEV) NIa protease has become widely used in molecular biology. TEV protease is a 27 kDa catalytic domain of the Nuclear Inclusion a (NIa) protein encoded by the TEV (Waugh D, 2002). Because its sequence specificity is far more stringent than that of factor Xa, thrombin, or enterokinase, TEV protease has become a widely used reagent for cleaving fusion proteins. It is also relatively easy to express and purify in large quantities (Doudna J, 2001). Based on the advantage of TEV protease, a TEV cleavage site was introduced into the N-terminus of Rev.

Rev original DNA sequence is:

```
                                                  NdeI
GGCCATCATCATCATCATCATAGCAGCGGCCTGTTTAAACGACA^TATGGCTGGTCGCTCTGGCGA
 G   H   H   H   H   H   H   S   S   G   L   F   K   R   H    M   A   G   R   S   G
```

The original Rev DNA sequence was cut by NdeI restriction enzyme:

```
GGCCATCATCATCATCATCATAGCAGCGGCCTGTTTAAACGACA       TATGGCTGGTCGCTCTGGCGA
CCGGTAGTAGTAGTAGTAGTATCGTCGCCGGACAAATTTGCTGTAT       ACCGACCAGCGAGACCGCT
```

Two paired 68base oligos were synthesized as insert:

```
        MseI                                                    NdeI
5'-CATCACCTTAACGATTACGACATCCCCACTACTGAGAATCTTTATTTTCAGGGACATATGTCGACTGT
   GTAGTGGAATTGCTAATGCTGTAGGGGTGATGACTCTTAGAAATAAAAGTCCCTGTATACAGCTGACA-5'
```

MseI and NdeI were used to cut the oligos to generate insert piece:

Insert piece

```
5'-CATCACCT    TAACGATTACGACATCCCCACTACTGAGAATCTTTATTTTCAGGGACA    TATGTCGACTGT
   GTAGTGGAAT    TGCTAATGCTGTAGGGGTGATGACTCTTAGAAATAAAAGTCCCTGTAT    ACAGCTGACA-5'
```

T4 DNA ligase was used to creat the destination DNA sequence:

GGCCATCATCATCATCATCATAGCAGCGGCCTGTTTAAACGA
G H H H H H H S S G L F K R
CA^TAACGATTACGACATCCCCACTACTGAGAATCTTTATTTTCAGGGACA^TATG
H N D Y D I P T T E N L Y F Q G H M
GCTGGTCGCTCTGGC
A G R S G

CATCATCATCATCATCAT His tag for Ni binding
H H H H H H

CGATTACGACATCCCCACTACT spacer for TEV recognition
D Y D I P T T

GAGAATCTTTATTTTCAGGGA TEV recognition site
E N L Y F Q G

MseI and NdeI, were used to cut the 68bp oligos to create the TEV cleavage site insert. MseI generated the same sticky end as NdeI. The product was purified by PCR purification kit following the standard protocol provided by Qiagen. The original Rev vector was cut by NdeI to generate linear DNA with two sticky-ends. The digest solution was loaded in 0.8% agarose gel to separate the cleaved and uncleaved vector. The linear vector was purified by Gel purification kit. Then, the insert DNA and the linearized vector were mixed together with 3:1 molar ratio in the T4 DNA ligase buffer. After adding the ligase, the ligation reaction was carried out at 12ºC over night. After the ligation reaction was completed, the product was directly transformed into DH5α competent cells. First, 150 µl DH5α competent cells were thawed on ice; then, the cells were gently mixed with 10 µl ligation products in 1.5 ml microcentrifuge tubes and tubes were incubated on ice for 30 minutes. Then, cells were heat shocked for 40 seconds in a 42°C water bath without shaking. After heat shock, the tubes were placed on ice for 2 minutes. Pre-warmed SOC media (20g Tryptone, 5g yeast extract, 2ml of 5M NaCl, 2.5ml of 1M KCl, 10ml of 1M $MgCl_2$, 10ml of 1M $MgSO_4$ and 20ml of 1M glucose per liter), 840 ul, was added to the tubes and incubated at 37°C for 1 hour. Finally, 50 µl of cells were spread on pre-warmed Amp+ LB (10g tryptone, 5g yeast extract and 10g NaCl per liter) plates. The rest 950 µl remains of cells were spread on another pre-warmed

Amp+ LB plate to ensure that at least one plate will have well-spaced colonies. All plates were incubated overnight at 37°C. The positive clones were picked up for sequencing. Because the insert sticky ends are the same, the reversible insert also could occur. The DNA sequencing verified two correct clones out of six.

2.4 Rev truncated mutants

In order to further explore the biophysical details of Rev's different domains, two Rev truncated mutants were designed: Loop-deletion (LD); and C-terminal. After subcloning (RevLD and RevC), the constructs were transformed into DH5α competent cells.

The Rev loop deletion 5' and 3' primers sets were designed to replace Rev loop region with a Gly-Ser-Gly-Ala linker (Figure 2.1).

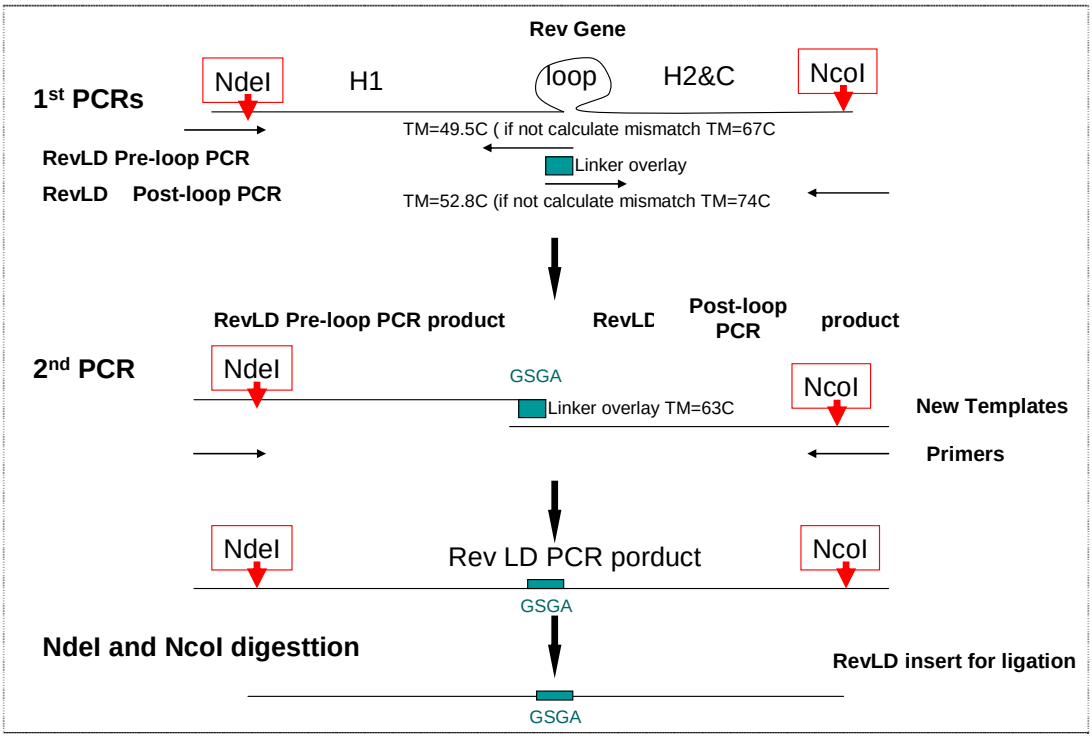


Figure 2.1 Rev loop deletion mutant design

The detailed manipulation of Rev gene was shown below with red based indicating the deleted loop region, green based showing the inserted Gly-Ser-Gly-Ala linker.

```
Rev sequence
GGCCATCATCATCATCATAGCAGCGGCCTGTTTAAACGACATATGGCTGGTCGCTCT
                                                M  A  G  R  S

GGCGATTCTGATGAAGACCTTCTCAAAGCCGTTCGCCTGATCAAGTTTCTCTACCAGAGC
linker oligo                              ATCAAGTTTCTCTACCAG---
 G  D  S  D  E  D  L  L  K  A  V  R  L  I  K  F  L  Y  Q  S

AACCCTCCGCCAAACCCAGAGGGTACTCGCCAGGCGCGTCGCAACAGGCGTCGACGGTGG
-------------------------------------------GGCTCGGGCGCGTGG
                                                G  S  G  A
 N  P  P  P  N  P  E  G  T  R  Q  A  R  R  N  R  R  R  R  W

CGCGAACGTCAGCGTCAAATCCACAGCATTTCCGAGCGCATTCTGAGCACTTACCTCGGC
CGCGAACGTCAGCGTC
 R  E  R  Q  R  Q  I  H  S  I  S  E  R  I  L  S  T  Y  L  G

CGTTCTGCTGAGCCCGTCCCACTTCAGTTGCCTCCCTTAGAGCGTTTAACTCTAGACTGC
 R  S  A  E  P  V  P  L  Q  L  P  P  L  E  R  L  T  L  D  C

AACGAGGATTGTGGAACTTCTGGGACGCAGGGTGTTGGAAGCCCTCAAATACTAGTGGAG
 N  E  D  C  G  T  S  G  T  Q  G  V  G  S  P  Q  I  L  V  E

TCCCCGACTGTTCTGGAGTCTGGAACCAAAGAGTAGTAAGCGGTCGACCTCAGGCATTTG
 S  P  T  V  L  E  S  G  T  K  E  -  -
```

Primer set for RevLD pre-loop region:

NdeI

5'-TTTCAGGGACATATGGCTGGTCGCTCTGGCG

5'-GACGCTGACGTTCGCGCCACGCGCCCGAGCCCTGGTAGAGAAACTTG

Primer set for RevLD post-loop region:

5'-ATCAAGTTTCTCTACCAGGGCTCGGGCGCGTGGCGCGAACGTCAGCGTC

NcoI

5'-TGATCGATCCATGGGCAAATATTATACGC

```
5′-CATCACCTTAACGATTACGACATCCCCACTACTGAGAATCTTTATTTTCAGGGACATATGTCGACTGT
   GTAGTGGAATTGCTAATGCTGTAGGGGTGATGACTCTTAGAAATAAAAGTCCCTGTATACAGCTGACA-5′
```

The Rev-C TEV was subcloned from Rev TEV wild type (green part is designed RevC region):

```
Rev wild type DNA sequence:
GGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACACATG
                                                    NdeI
GGCCATCATCATCATCATCATAGCAGCGGCCTGTTTAAACGACATATGGCTGGTCGCTCT
                                                      M  A  G  R  S
GGCGATTCTGATGAAGACCTTCTCAAAGCCGTTCGCCTGATCAAGTTTCTCTACCAGAGC
 G  D  S  D  E  D  L  L  K  A  V  R  L  I  K  F  L  Y  Q  S
AACCCTCCGCCAAACCCAGAGGGTACTCGCCAGGCGCGTCGCAACAGGCGTCGACGGTGG
 N  P  P  P  N  P  E  G  T  R  Q  A  R  R  N  R  R  R  R  W
CGCGAACGTCAGCGTCAAATCCACAGCATTTCCGAGCGCATTCTGAGCACTTACCTCGGC
 R  E  R  Q  R  Q  I  H  S  I  S  E  R  I  L  S  T  Y  L  G
CGTTCTGCTGAGCCCGTCCCACTTCAGTTGCCTCCCTTAGAGCGTTTAACTCTAGACTGC
 R  S  A  E  P  V  P  L  Q  L  P  P  L  E  R  L  T  L  D  C
AACGAGGATTGTGGAACTTCTGGGACGCAGGGTGTTGGAAGCCCTCAAATACTAGTGGAG
 N  E  D  C  G  T  S  G  T  Q  G  V  G  S  P  Q  I  L  V  E
TCCCCGACTGTTCTGGAGTCTGGAACCAAAGAGTAGTAAGCGGTCGACCTCAGGCATTTG
 S  P  T  V  L  E  S  G  T  K  E  ●  ●
AGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACAT
AAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATT
TCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCA
CCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAA
TTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGC
CAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGATCGATCAGGCCT
                                                    NcoI
GCTAGCATGAC
```

Primer set for RevC construct, introduce NdeI site before RevC gene:

```
              NdeI
5′-GTAGCACACATATGGAGCGCATTCTGAGCACTTAC

              NcoI
5′-TGATCGATCCATGGGCAAATATTATACGC
ATCGATCA
```

RevC DNA sequence was produced by PCR reaction and purified by Qiagen PCR purification kit. The DNA sequence of PCR product was shown as:

```
GTAGCACACATATGGAGCGCATTCTGAGCACTTACCTCGGC
               E  R  I  L  S  T  Y  L  G
CGTTCTGCTGAGCCCGTCCCACTTCAGTTGCCTCCCTTAGAGCGTTTAACTCTAGACTGC
 R  S  A  E  P  V  P  L  Q  L  P  P  L  E  R  L  T  L  D  C
AACGAGGATTGTGGAACTTCTGGGACGCAGGGTGTTGGAAGCCCTCAAATACTAGTGGAG
 N  E  D  C  G  T  S  G  T  Q  G  V  G  S  P  Q  I  L  V  E
TCCCCGACTGTTCTGGAGTCTGGAACCAAAGAGTAGTAAGCGGTCGACCTCAGGCATTTG
 S  P  T  V  L  E  S  G  T  K  E  ●  ●
AGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACAT
AAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATT
TCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCA
CCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAA
TTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGC
CAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGATCGATCA
```

The RevC PCR product was digested by NdeI and NcoI and purified by Qiagen PCR purification kit. The DNA sequence of insert was shown as:

```
          TATGGAGCGCATTCTGAGCACTTACCTCGGC
               E  R  I  L  S  T  Y  L  G
CGTTCTGCTGAGCCCGTCCCACTTCAGTTGCCTCCCTTAGAGCGTTTAACTCTAGACTGC
 R  S  A  E  P  V  P  L  Q  L  P  P  L  E  R  L  T  L  D  C
AACGAGGATTGTGGAACTTCTGGGACGCAGGGTGTTGGAAGCCCTCAAATACTAGTGGAG
 N  E  D  C  G  T  S  G  T  Q  G  V  G  S  P  Q  I  L  V  E
TCCCCGACTGTTCTGGAGTCTGGAACCAAAGAGTAGTAAGCGGTCGACCTCAGGCATTTG
 S  P  T  V  L  E  S  G  T  K  E  ●  ●
AGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACAT
AAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATT
TCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCA
CCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAA
TTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGC
CAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCC
```

Rev original sequence was cleaved by NdeI and NcoI, then 0.8% agarose gel was used to separate the cleaved vector, the Quagen Gel purification kit was used to purify the cleaved vector.

```
                                          NdeI     NcoI
GGCCATCATCATCATCATCATAGCAGCGGCCTGTTTAAACGACA     CATGGATCGATCAGGCCTGCTAGCATGAC
CCGGTAGTAGTAGTAGTAGTATCGTCGCCGGACAAATTTGCTGTAT       CTAGCTAGTCCGGACGATCGTACTG
```

The ligation was applied with RevC PCR product: vector as 3:1 in molar ratio in 10ul solution at 12°C for over night.

RevC construct:

```
                                              NdeI
ATCATCATCATCATCATAGCAGCGGCCTGTTTAAACGACATATGGAGCGCATTCTGAGCACTTACCTCGGC
                                                  E  R  I  L  S  T  Y  L  G
CGTTCTGCTGAGCCCGTCCCACTTCAGTTGCCTCCCTTAGAGCGTTTAACTCTAGACTGC
 R  S  A  E  P  V  P  L  Q  L  P  P  L  E  R  L  T  L  D  C
AACGAGGATTGTGGAACTTCTGGGACGCAGGGTGTTGGAAGCCCTCAAATACTAGTGGAG
 N  E  D  C  G  T  S  G  T  Q  G  V  G  S  P  Q  I  L  V  E
TCCCCGACTGTTCTGGAGTCTGGAACCAAAGAGTAGTAAGCGGTCGACCTCAGGCATTTG
 S  P  T  V  L  E  S  G  T  K  E  ● ●
AGAAGCACACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACAT
AAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCGAATT
TCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCA
CCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAA
TTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGC
CAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGATCGATCAGGCCTGCTAGCATGAC
                                              NcoI
```

2.5 Results and Discussion:

After the transformation, the positive clones were amplified and sent to the Sealy Center for Molecular Science for sequencing. The correct clones were stored both at -20°C as plasmid forms and at -80°C in glycerol as cell forms.

Besides the correct clones, there are many false-positive clones, which are mainly original Rev plasmid. This situation turns into a disaster for making the RevLD TEV construct. Because the lengths of PCR products of RevLD and Rev wild type are quite similar, it is hard to tell the difference on the agarose gel. The sequencing is the only way to screen the positive clones. Because the first trail of all ten positive clones turned out to be false-positive, a new vector (YP-TEV) (Figure 2.2) was designed to eliminate the possible false-positive clones. After checking the vector restriction enzyme list of original Rev plasmid, an oligo containing multicloning sites was inserted between NdeI and NcoI to replace the Rev gene:

NdeI MfeI EcoRI XhoI KpnI NcoI

CA^TATGGCAATTGGAGAATTCTCTCGAGGGGTACCCCATGG multi-cloning site

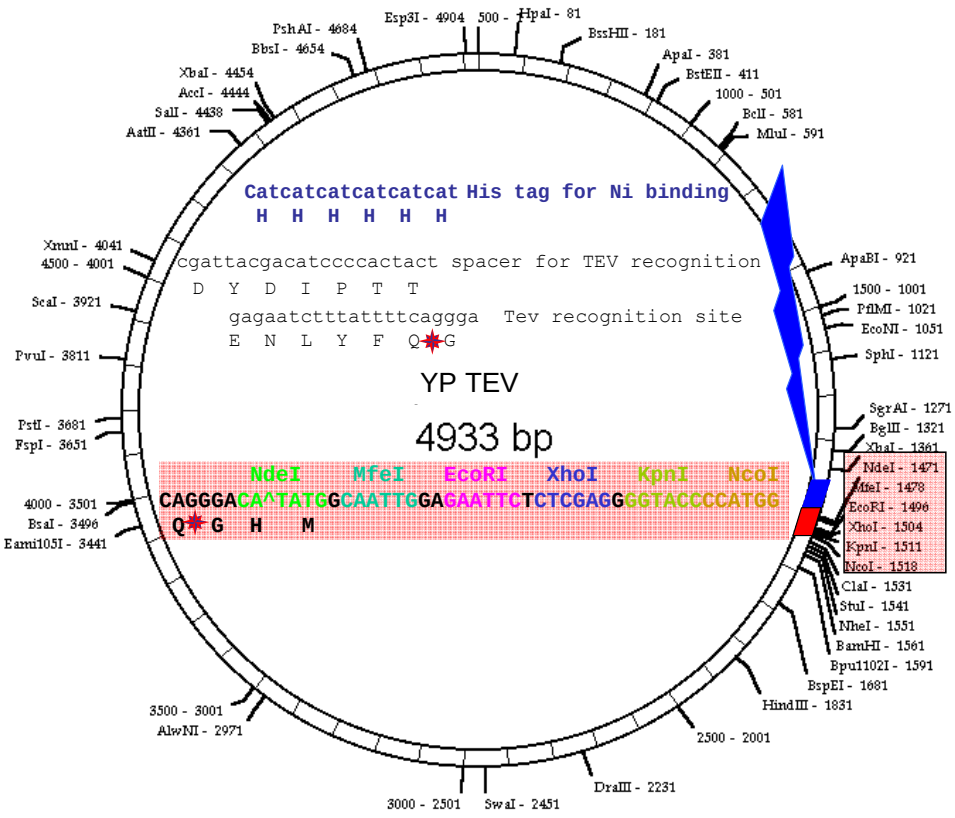


Figure 2.2 YP TEV plasmid restriction enzyme map. The blue segment demonstrates the His tag- spacer for TEV recognition-TEV recognition region. The red segment demonstrates the new multi-cloning sites, with the pink box showing the details.

This new vector has four unique restriction enzyme cleavage sites (MfeI, EcoRI, XhoI and KpnI) between NdeI and NcoI. After the ligation using this new vector, the ligase was heat inactivated and clones were selected for two or more restriction enzymes among those four to cut the ligation products. If there is any vector coming from self-ligation, it

will be cleaved after this treatment whereas the plasmid with insert will not be affected. This new vector led me to get the RevLD TEV construct successfully.

CHAPTER 3: THE BIOPHYSICAL AND STRUCTURAL STUDIES OF REV C-TERMINAL

Analysis of Rev truncation mutants (Figure 3.1) revealed that Rev C can remain monomeric in solution at high concentrations.

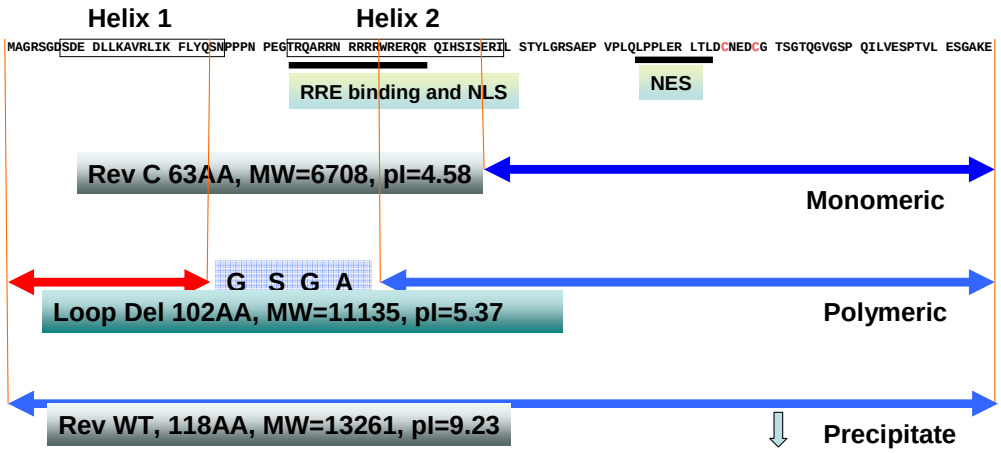


Figure 3.1 Rev constructs overview. The primary Rev sequence was shown as a reference with the prediction of two helices (hollow box) and two function domains (black bar). The primary characters of the three Rev constructs, RevC, Rev loop deletion and Rev wild type were indicated in the filled boxes. The assembly abilities of the three constructs were also illustrated.

3.1 The expression and purification of RevC TEV mutant

3.1.1 The expression of Rev C TEV

RevC TEV plasmid was transformed into *E. coli* BL21DE3 Code plus cells. The cells were plated overnight onto LB plates with 50mg/L Ampicillin at 37°C. One clone was picked up and amplified in 50ml LB medium with 50mg/L Ampicillin for overnight

growth at 37°C. The cells were finally pooled into 2L 2XYT with 50mg/L Ampicillin at 37°C. When the absorbance OD_{600} of the cell culture reached between 0.4 and 0.6, isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to the cell culture to 1mM final concentration for induction. After 3-hour induction, the cells were spin down at 2,000 × g for 30minutes and the pellets stored at -80°C.

3.1.2 The purification of RevC TEV with Ni-NTA column

Due to the presence of a hexa-histidine tag at the N-terminus of the expressed protein, RevC TEV protein was purified using the Ni-NTA column. The cell pellets were thawed on ice and suspended in lyses buffer (10 mM imidazole, 100 mM NaCl, 20 mM Tris, and 5mM 2-mercaptoethanol at pH 7.5). Cells were lysed by sonicator on ice after 10 bursts (30 seconds / burst with 30 seconds interval between bursts), followed by centrifugation at 38,000 × g for 30 minutes. The supernatant liquid containing RevC TEV protein was loaded onto the nickel column (pre-equilibrated with the lyses buffer) containing 30ml of Ni-NTA superflow resin (Qiagen). The column was washed with over 5 column volume (150ml) of lyses/wash buffer until the absorbance OD_{280} reached the baseline. RevC TEV fraction was collected using an imidazole gradient (10mM to 750mM at 4ml/min for 40 minutes) and dialyzed against TEV cleavage buffer (5 mM 2-mercaptoethanol, 100 mM NaCl, 20 mM Tris pH 8.0).

3.1.3 The fusion tag cleavage by TEV protease

In order to remove the fusion tag, RevC TEV was first concentrated by centrifugation using a Centriprep YM-3 (Millipore). The concentration of the protein was measured by UV spectrometer at OD_{280} with the extinction coefficient of 3960 M^{-1} cm^{-1}. The cleavage reaction was performed by adding TEV protease in 1:10 molar ratio with 1mg/ml RevC TEV at 16°C overnight.

Contaminants, such as the fusion His-tag, TEV protease and any uncleaved RevC TEV proteins were finally removed by using a second Nickel affinity column. The

fractions containing RevC were concentrated and loaded onto a Superdex G-75 gel filtration column for further purification and molecular weight determination.

3.1.4 The gel filtration purification of RevC

Gel filtration is a separation technique in the purification of biological macromolecules. Gel filtration separates molecules by their native molecular weight due to their differential penetration of the gel matrix. Very large molecules, which never enter the gel matrix, move through the chromatographic bed fastest. Smaller molecules which can enter the gel pores, move more slowly through the column, since they spend a portion of their time in the stationary phase (gel). Molecules are therefore, eluted in order of decreasing molecular size.

After equilibrating the Superdex G-75 column (Pharmacia) with the elution buffer, (100mM NaCl, 10mM Tris, 0.5mM TCEP and pH 7.2 at 2.5ml/minute) 0.6ml (7mg/ml) of RevC was loaded onto the column and eluted using a flow rate of 2.5ml/minute. The absorbance at 280 nm wave length was recorded with automatic one channel recorder (Pharmacia). The fractions in each protein peak were separately collected and concentrated by Centriprep YM3. The elution fractions were checked for their purity by SDS-PAGE (Figure 3.2). The fractions that corresponded to the molecular weight of RevC were pooled together.

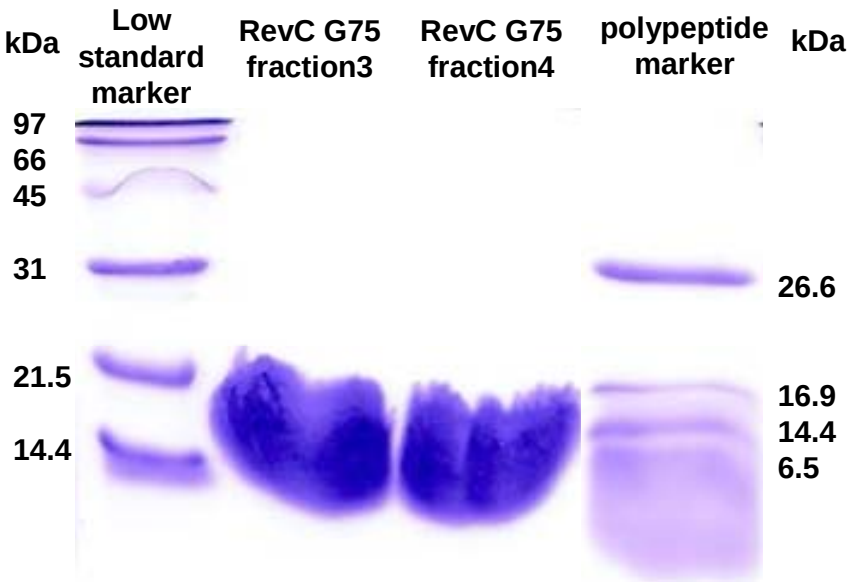


Figure 3.2 RevC SDS-PAGE shows the purity of RevC after Superdex G-75 gel filtration column purification. The fractions (3, 4) of RevC have no observed contamination from SDS-PAGE gel.

3.2 Estimate molecular weight of RevC by gel filtration column

The molecular weight of a protein can be accurately determined by calibrating the Superdex G-75 column using standard molecular weight markers. After applying a mixture of gel filtration molecular weight marker proteins (Bio-Rad 151-1901), the fractions were collected and the elution volumes (Ve) of each marker protein from the recorded peaks were determined. Gads-SH3 was added as one of the molecular weight markers because its size (6878 Da) is similar to that of RevC, which is 6708.5 Da. The void volume (Vo) was determined by elution volume of thyroglobulin (670k Da). The gel bed volume (Vt) was determined using vitamin B_{12} (1.35k Da). For each protein marker, the value of Kav was calculated by Kav=(Ve-Vo)/ (Vt-Vo), (Table 3.1). The standard

graph was drawn by plotting Kav against log molecular weight (Figure 3.3) and the approximate molecular weight for RevC protein was estimated.

Table 3.1 Superdex G75 calibration

superdex 75			Kav=(Ve-Vo)/(Vt-Vo)	
sample	MW (Da)	LogMW	Kav	Ve (ml)
Vt				297
vitamin B12 Vt	1350	3.130333768	0.938072159	285.5
Gads SH3	6877.6	3.837436914	0.644588045	231
myoglobin	17000	4.230448921	0.437264405	192.5
ovalbumin	44000	4.643452676	0.219170705	152
thyroglobulin Vo	670000	5.826074803	0	111.3

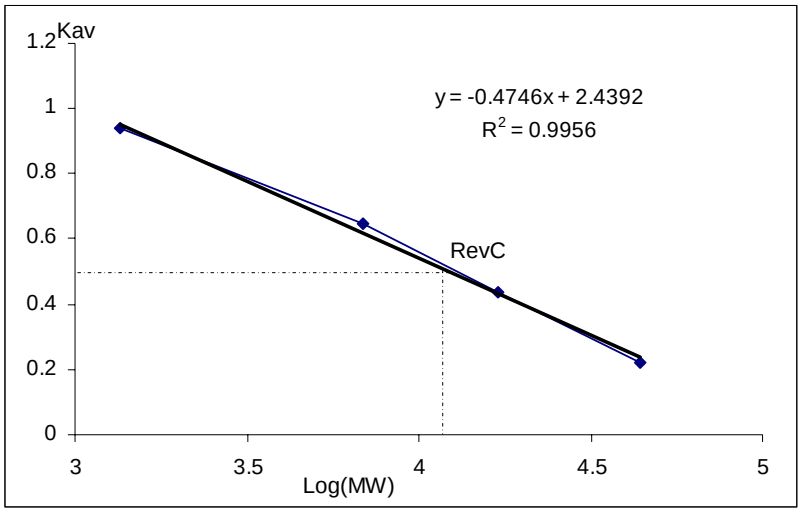


Figure 3.3 RevC molecular weight estimation from standard curve of the Superdex G-75 column. RevC was eluted at 202ml, which was used to calculate the Kav of RevC. From the Kav of RevC, the Log value of estimated RevC molecular weight was obtained from the standard curve.

The estimated molecular weight of RevC is 11.8 kDa, and is larger than expected (6708.5 Da). This phenomenon could be explained as self-association, adsorption of gel

matrix or not well packed as folded protein. Because the estimated size of RevC is between monomer and dimer, the protein may have an equilibration between the two forms. However, the elution profile is symmetrical which excludes the possibility of equilibration between two forms and adsorption of gel matrix, (Figure 3.4). So, the larger size of RevC is due to the natively unfolded feature of the protein monomer. RevC remains monomeric even at high concentration. This result indicates that the multimerization reactions do not involve the Rev C-terminus.

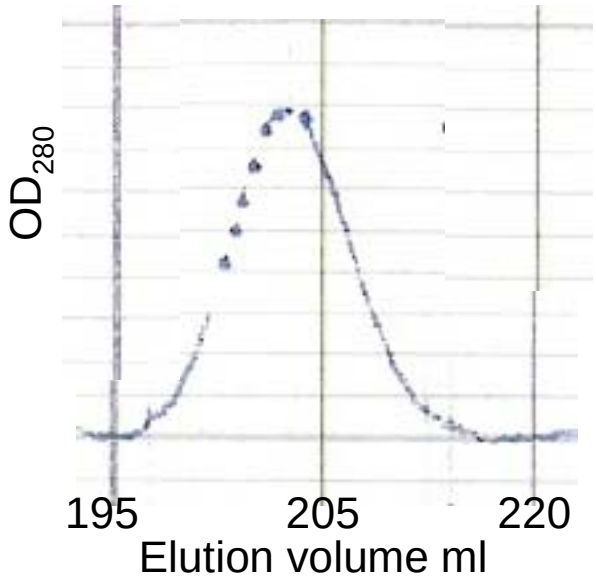


Figure 3.4 RevC Superdex G-75 elution curve profile. The RevC elution curve profile showed a symmetric peak which excludes the possibilities of RevC self-association or the matrix adsorption of RevC.

3.3 The circular dichroism analysis of RevC secondary structure

Circular dichroism, or CD, is defined as the differential absorption of left and right hand circularly polarized (CP) light. At a given wavelength, the difference between absorbance of left circularly polarized and right circularly polarized light is provided by ΔA:

$\Delta A = AL - AR = (\varepsilon L - \varepsilon R)*C*L$

Where εL and εR are the molar extinction coefficients for RCP and LCP light, C is the molar concentration, L is path length. Then

$\Delta\varepsilon = (\varepsilon L - \varepsilon R)$ - is the molar Circular dichroism (reported in degrees of ellipticity)

The molar ellipticity is defined as:

$[\theta] = 3298\Delta\varepsilon$

In general, this phenomenon will be exhibited in absorption bands of any optically active molecule. The far UV CD spectrum of proteins can predict important characteristics of their secondary structure. CD spectra can be readily used to estimate the fraction of a molecule that is in the alpha-helix conformation where $\pi\rightarrow\pi^*$ transitions leads to positive $(\pi\rightarrow\pi^*)_{perpendicular}$ at 192 nm and negative $(\pi\rightarrow\pi^*)_{parallel}$ at 209 nm, negative at 222 nm and red shifted ($n\rightarrow\pi^*$), The fraction of the molecule in the beta-sheet conformation is characterized by a negative ($\pi\rightarrow\pi^*$) at 218 nm and a positive at 196 nm ($n\rightarrow\pi^*$), The amount of beta-turn conformation, or other (random) conformations are characterized by positive at 212 nm ($\pi\rightarrow\pi^*$) and negative at 195 nm ($n\rightarrow\pi^*$).

RevC protein (0.2 mg/ml) was dialyzed into 10 mM potassium phosphate buffer to minimize absorbance around 200 nm. The sample was filtered through a 0.2 μm filter. A 300 ul sample was loaded into a 1mm cuvette and far-UV CD data was collected ($\lambda_{200\text{-}250nm}$) at 20°C (see Figure 3.5). If the HT (high-tension voltage) value is over 800, the data should not be trusted. In that case either the sample concentration needs to be decreased or a shorter path length needs to be used. The secondary structure deconvolution program (ACDP) was used for the secondary structure calculation (Yang *et al.* 1986; Chang *et al.* 1978; Andrade *et al.* 1993; Chou and Fasman 1974).

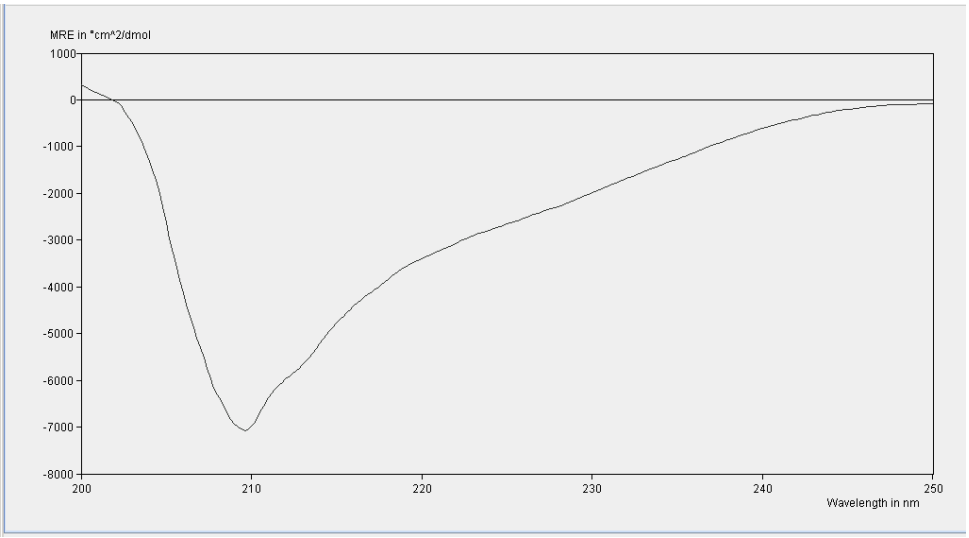


Figure 3.5 RevC Circular Dichroism spectrum. With the minimum at 210nm, the spectrum demonstrated that RevC is a disordered protein with some beta strand structure

The RevC CD spectrum shows the segmentally disordered conformation by a strong negative band near 210 nm and a negative shoulder at approximately 220nm. From the secondary structure deconvolution result of CD spectrum, RevC contains over 40% of its conformation as beta strand structure, (Table 3.2).

Table 3.2 RevC secondary structure prediction

structure	K2D backpropagation neural network
Alpha helix	0.082
Beta strand	0.411
Unordered	0.506
Sum	0.999

The low amount of alpha helix structure in RevC indirectly supports the concept that the Rev N-terminal region is dominated by alpha helix; because Rev contains about 50% alpha helix and RevC is the C-terminal half of the Rev protein. Because the disordered

part of RevC will not be well packed, it possibly explains its larger molecular weight in the gel filtration column.

3.4 The expression and preparation of ^{15}N labeled RevC NMR sample

The minimal media (M9) was prepared for the growth of 15N labeled RevC. The media contains 12gm Na_2HPO_4, 6gm KH_2PO_4 and 0.5gm NaCl per liter. After autoclaving, add mixed and filtered 100ul 1M $CaCl_2$, 2ml 1M $MgSO_4$, 1ml Trace metals (containing 14mg $ZnCl_2$, 11.4mg H_3BO_3, 5mg $MnCl_2.4H_2O$, 16.8mg $FeCl_2$, 1.6mg $CoCl_2$, 1.6mg $CuCl_2$ and 1.1mg $(NH_4)_6Mo_7O_{24}.4H_2O$ per milliliter), 10ml 100X Vitamins (GIBCO), 4gm glucose and 1gm ^{15}N-NH_4Cl per liter. After the clone was amplified overnight in 50ml LB media with 50mg/liter Ampicilin, the cells were gently spun down. The supernatant liquid was discarded, and the cell pellet was suspended in M9 minimal media. The growth followed the same protocol as the one for 2XYT rich media except the longer induce period (overnight at room temperature). The RevC-TEV fusion protein was purified by a Ni column, and the tag was removed by TEV protease. Final purification of ^{15}N-RevC was achieved by a Superdex G-75 gel filtration column. The purified ^{15}N-RevC was dialyzed in NMR buffer containing 20mM Bis-Tris, 100mM NaCl, 1mM TCEP at pH6.0. The ^{15}N-RevC NMR sample was concentrated to a final concentration of 0.5mM using Centriprep YM3 and 10% D_2O was finally added for the magnetic field locking process.

3.5 The ^{1}H-^{15}N HSQC spectrum of RevC

^{15}N-HSQC is a two dimensional NMR experiment that correlates the nitrogen atom of an NH_x group with the directly attached proton. Each signal in a HSQC spectrum represents a proton that is bound to a nitrogen atom. Each amino acid residue (except proline) would give rise to one signal in the ^{15}N-HSQC spectrum that corresponds to the backbone N-H amide group. In addition to the backbone amides, additional signals would be observed for the nitrogen-containing side chains, e.g. Gln, Asn and Trp. Lysine NH_3^+ groups can be rarely detected due to the rapid exchange of the NH_3^+protons with water.

Amide proton and ^{15}N chemical shifts are very sensitive to the state of the protein backbone: *i.e*., hydrogen bonding and local structural environment. Any change in the local or global conformation of the protein can be monitored by changes in chemical shifts of the signals in the HSQC spectrum. Folded proteins or protein domains display a broad distribution of NMR frequencies resulting in distinct signals in the ^{15}N-HSQC. Conversely, a protein that is unfolded or disordered may show amide proton peaks that overlap and cluster in the 8 – 8.5 ppm range.

The ^{15}N HSQC experiments of RevC were recorded on Varian Inova 800 MHz spectrometers. The FID data files were processed by NMR pipe and visualized by NMR view. Most of the amide proton signals of RevC cluster around 8-8.5 ppm, thereby indicating a predominantly random coil conformation of the protein. However, the possibility of partially structured regions could be inferred from previous CD analysis of RevC that indicated the presence of ~ 40% beta strand in the protein. Average beta strand shift relative to the random coil for amide proton is 0.38 ppm (Wishart D, 1991; Wishart D, 1992; Wishart D, 1994). This small difference may not shift the proton peaks out of the 8-8.5 ppm region. In the denatured state studies, proton peaks in HSQC spectra are limited to 8-8.5 ppm, but the local secondary structure still exists (Baker D, 2000).

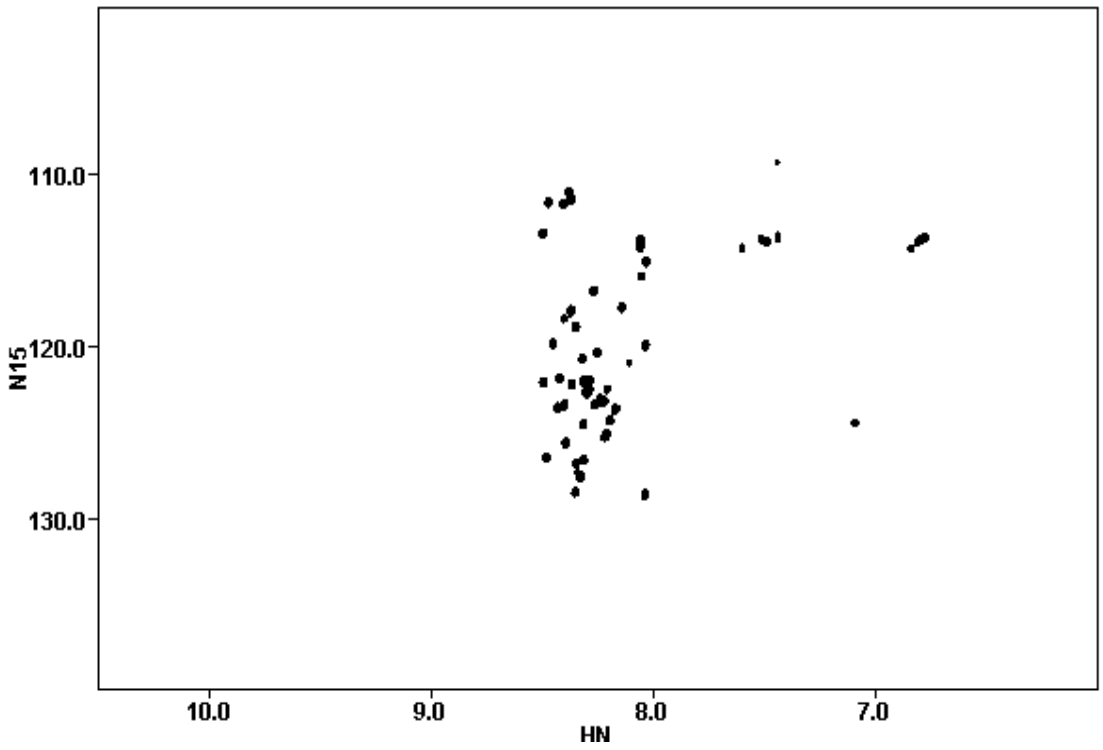


Figure 3.6 RevC ^{15}N HSQC spectrum. The amide proton signals of RevC HSQC were limited within 8.0 and 8.5 ppm range.

The RevC HSQC spectra were collected at variable temperatures of 5ºC, 15ºC and 25ºC. About 50 backbone NH peaks can be discerned out of a theoretical maximum of 53 peaks. Some peaks demonstrated a downfield (positive) shift, e.g. Peak number 6 and 32. The HSQC signal of a beta strand amide proton is shifted downfield compared to that of random coil (Richards F, 1992). This suggested that more beta strand structure was induced as the temperature decreased.

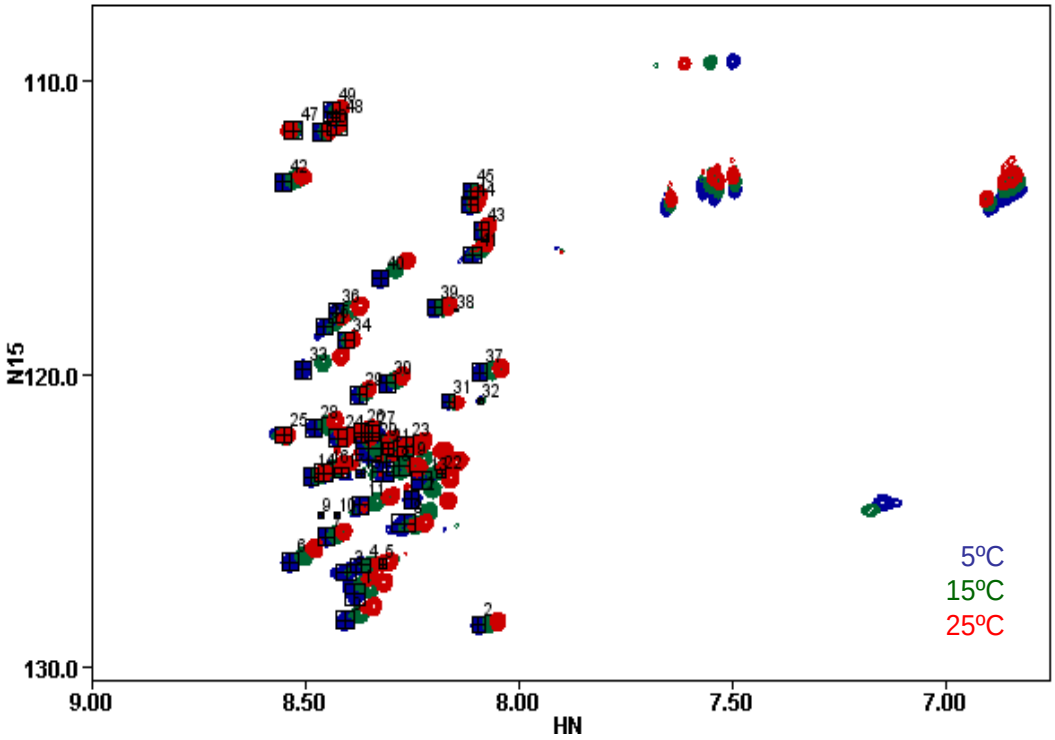


Figure 3.7 RevC 15N HSQC temperature variance overlay. 50 backbone HSQC peaks of RevC were observed under three different temperatures (5°C showing blue; 15°C showing green and 25°C showing red). Some amide protons, e.g. peak number 33, 40 and 6, demonstrated down field chemical shifts.

3.6 The assignment of RevC

Backbone resonances in proteins can be unambiguously assigned to the amino acid sequence using methods that correlate resonances from adjacent residues using J coupling (through bonds). The sequential assignments can be easily performed by a combination of two and three-dimensional experiments such as 2D HSQC, 3D HNCO, 3D HNCACB, CBCA(CO)NH and C(CO)NH experiments. The HNCACB resolves amide proton/nitrogen correlations in the same fashion as the 2D 1H-15N HSQC. The third dimension of the spectrum contains the ^{13}C chemical shifts of the C^{α} and C^{β} resonances of a given residue and of the residue before it in the sequence. The experiment exploits

the fact that each amide nitrogen is coupled to its own C^{α} (^{1}Jcn ~ 12 Hz) more strongly than to the C^{α} behind it (^{2}Jcn ~ 7 Hz). As a result, the strip for each N-H contains intense peaks for its own C^{α} and C^{β} resonances and weaker peaks for the C^{α} and C^{β} resonances of the previous residue. The way that magnetization is transferred between C^{α} and C^{β} causes the correlations for these carbons to have opposite signs. Whereas, proline residues interrupt this chain due to their missing amide proton. The assignment is much easier if the intra- and inter-residual cross signal can be distinguished. This discrimination can be made using the 3D CBCACONH experiment, which correlates the amide proton and nitrogen of a given amino acid residue with the alpha and beta carbons of the preceding residue. Therefore, a combination HNCACB and CBCA(CO)NH spectra allows the discrimination of the intra- and interresidual C^{α} and C^{β}chemical shifts. The strips from the two spectra show superpositions of a HNCACB spectrum and a CBCA(CO)NH spectrum (Figure 3.8).

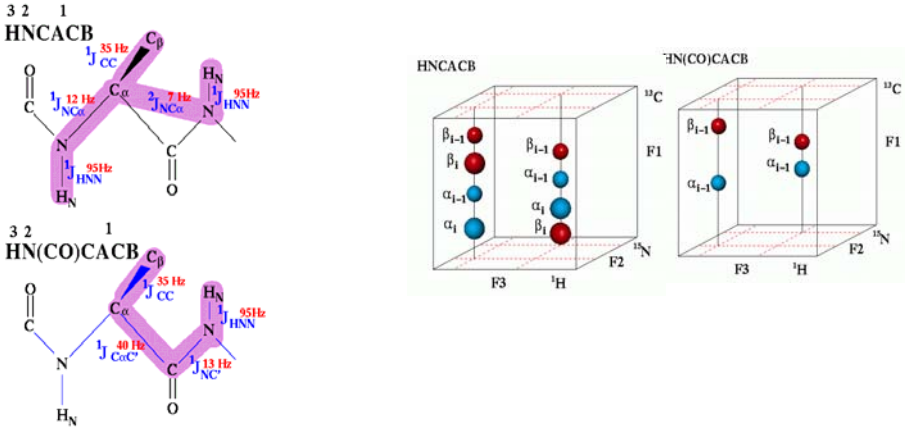


Figure 3.8 NMR sequence assignment scheme. Left panel demonstrated the NMR triple resonances of HNCACB and HN(CO)CACB. Right panel demonstrated the triple resonance signals of i and i-1 residues from HNCACB experiment and only i-1 residue's triple resonance signal could be observed by HN(CO)CACB experiment.

Several of these strips are placed in a row to show the sequential connectivities from each amino acid to the preceding one. Usually the sequential assignment by HNCACB and CBCACONH is ambiguous because the C^{α} and C^{β} frequencies of totally different amino acids are accidentally degenerate. Thus, the predecessor of a given amino acid often cannot be defined because there are several possibilities. Therefore, correlations via other types of nuclei have to be used for resolution of ambiguities in these experiments. Thus, 3D HNCO experiment is used in conjunction with CBCACONH and HNCACB for sequential assignments. The 3D HNCO experiment correlates the frequency of an amide proton with the frequencies of the inter-residual CO atom. Besides, the HNCO is very useful for resolving accidental signal degenerations in the HSQC projection: In proteins every amide proton is covalently bonded to a single CO group. Therefore, only one cross signal per frequency is observed in the HNCO. However, if two cross signals are observed at the frequency of an amide proton there have to be two amino acids with accidentally degenerated amide protons. The sequential assignment by triple resonance experiments yields only the connectivities of the individual spin systems with each other but not their amino acid types. Once the backbone assignments (C^{α}, C^{β}, N and NH) are completed, the proton assignments can be obtained using a combination of HBHA(CO)NH and HC(CO)NH experiments.

In order to assign the RevC protein using triple resonance spectra, $^{15}N^{13}C$ double labeled RevC protein was prepared. The same protocol for ^{15}N labeled RevC was followed except the M9 minimal media was supplemented with 3gm/liter of ^{13}C glucose. The experiments of triple resonance of RevC (CBCACONH, HNCACB, CCCONH and HNCO) were collected using Varian Inova 600 MHz spectrometer. The data were processed by NMR pipe and analyzed by NMR view, (Figure 3.9).

All the RevC backbone atoms could not be assigned due to variation in the ^{15}N and NH chemical shifts with time as monitored by ^{15}N-HSQC experiment. The amino acid sequence of RevC consists of six proline residues. The change in the conformation of the

protein as a result of slow isomerizations of X-Pro peptide bonds could possibly explain the changes observed in the ^{15}N HSQC spectrum of the protein with time.

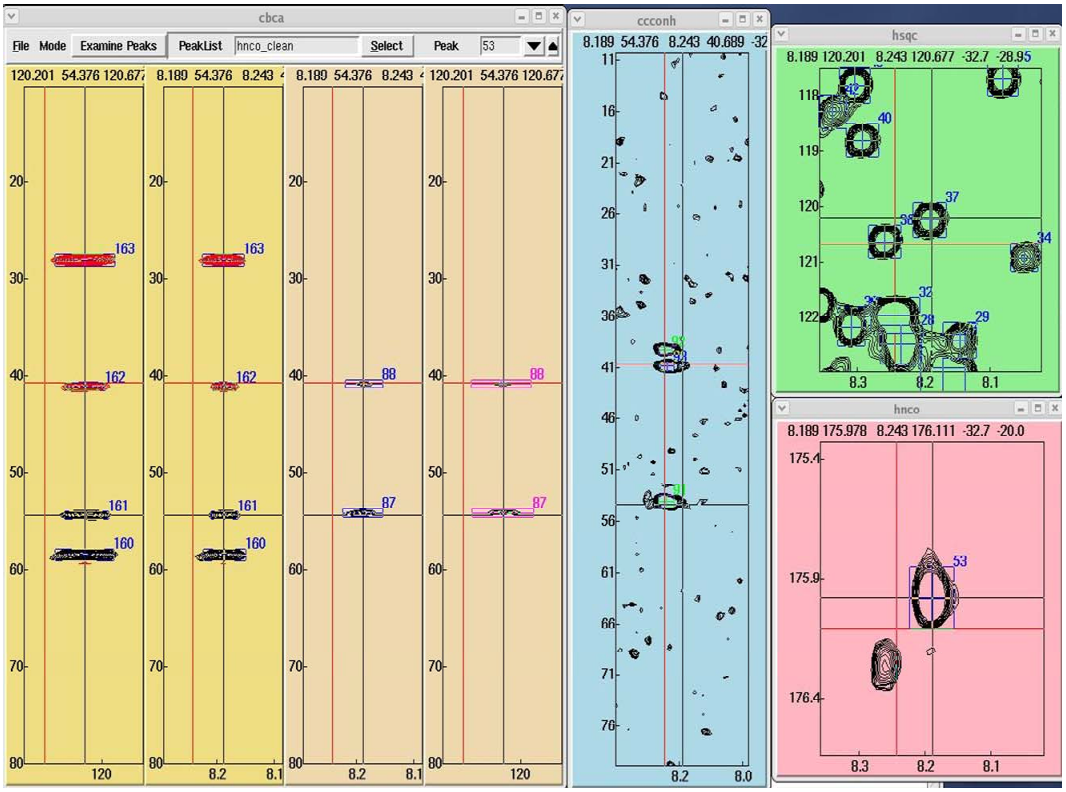
Compared to random coil, the amide proton chemical shift often has a downfield shift for beta strand (from http://bmrb.protein.osaka-u.ac.jp/ref_info/csishift.txt), (Table 3.3).

Table 3.3 Average chemical shift for amide proton in HSQC

Residue	Atom_type	All_chem_shift	Coil_chem_shift	β_chem_shift	α_chem_shift
LEU	proton of NH	8.25	7.99	8.58	8.11
GLU	proton of NH	8.30	8.40	8.49	8.20
GLN	proton of NH	8.23	8.19	8.51	8.11
GLY	proton of NH	8.33	8.36	8.44	8.07

From the assigned HSQC spectrum of RevC, the chemical shifts of Gly90, Gly93, Leu75, Glu105 and Gln95 are close to average beta strand value rather than average random coil value. The chemical shift of Glu116 is close to alpha helix value rather than random coil. These data suggest that RevC contains beta strand secondary structure.

(A)



(B) GHMERIL STYLGRSAEP VPLQLPPLER LTLDCNEDCG TSGTQGVGSP QILVESPTVL ESGTKE
61 71 81 91 101 111

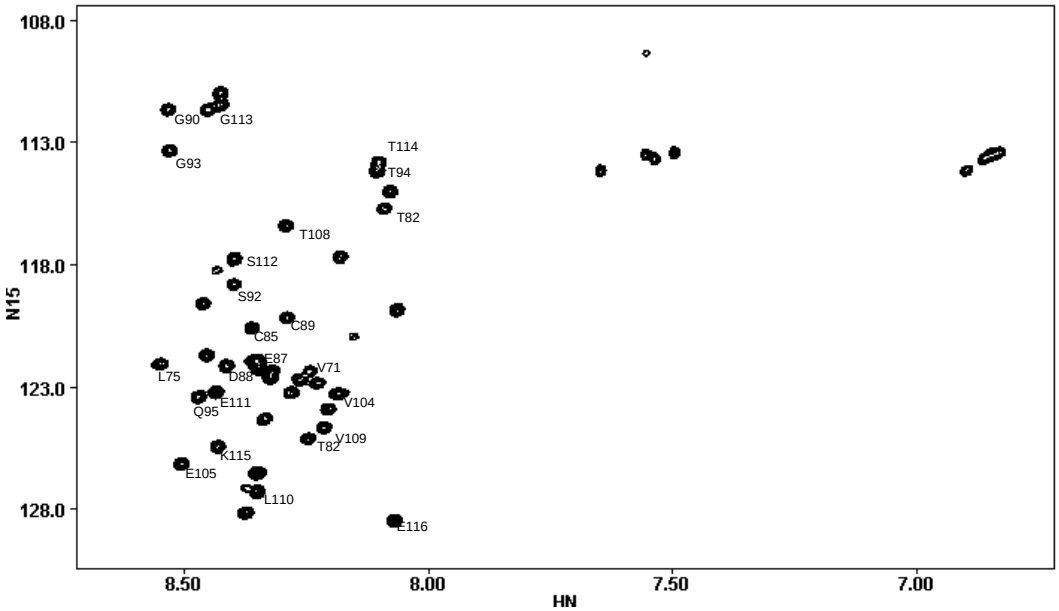


Figure 3.9 RevC sequence assignment. (A) NMR view strips for RevC assignment (B) partially assigned RevC HSQC

CHAPTER 4: THE STRUCTURAL STUDIES OF REV LOOP DELETION MUTANT

The mutant of Rev loop deletion was designed based on the Blasco helical projection model of a Rev monomer. The Rev model was built based on the structural constraints from the elegant genetic studies and atomic coordinates previously determined by NMR analysis for the arginine-rich Rev peptide (Jain and Belasco 2001) (Figure 4.1).

The Rev loop deletion construct was made by replacing Rev Ser25-Arg44 with Gly-Ser-Gly-Ala as a small loop linker between two helices. The removal of the loop and most of the arginine-rich domain of Rev was thought to impede or diminish filament formation. What is more, sequence analysis using FoldIndex of the Rev loop deletion indicated that it is more prone to fold compared to Rev wild type (Figure 4.2). FoldIndex is a program (http://bioportal.weizmann.ac.il/fldbin/findex), based on the algorithm of a combination of low overall hydrophobicity and large net charge representing a unique structural feature of "natively unfolded" proteins (Uversky *et al.* 2000). FoldIndex estimates the local and general probability for the provided sequence to fold (Prilusky *et al.* 2005).

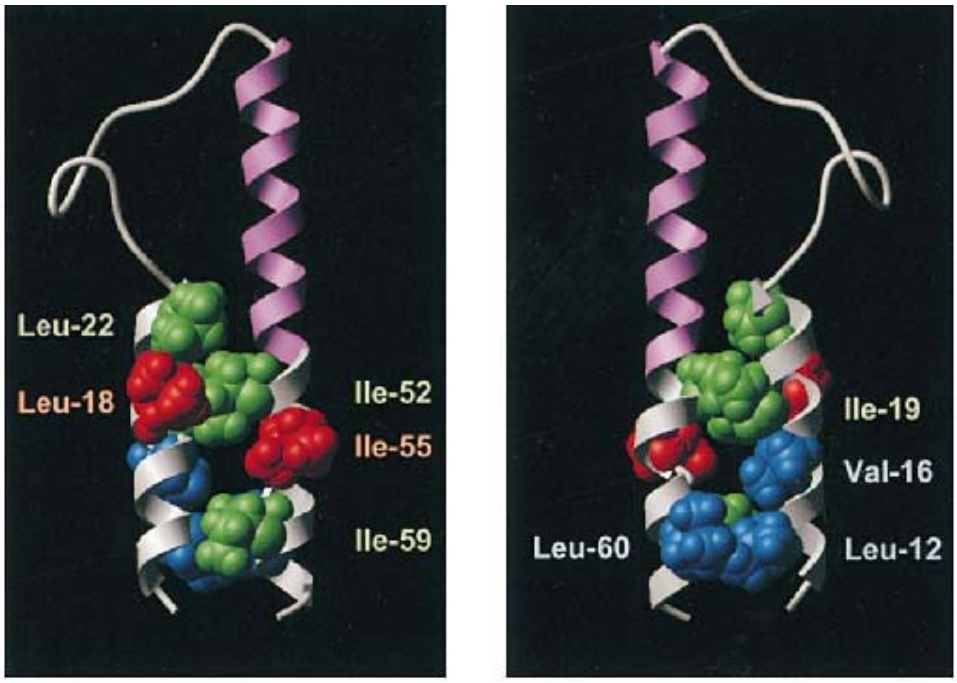


Figure 4.1 Helical projection model of a Rev monomer. (Jain and Belasco 2001). Rev N-terminus backbone is shown as two helical ribbons (residues 9–24 and 34–62) connected by a loop of undefined conformation (residues 25–33). The backbone of the arginine-rich RNA binding domain (residues 34–50) is colored violet. The α helices cross at an angle of +18º. Residues involved in class 1 (Leu-18 and Ile-55, red), class 2 (Leu-12, Val-16, and Leu-60, blue), or class 3 (Ile-19, Leu-22, Ile-52, and Ile-59, green) Rev multimerization mutants are shown with amino acid side chains.

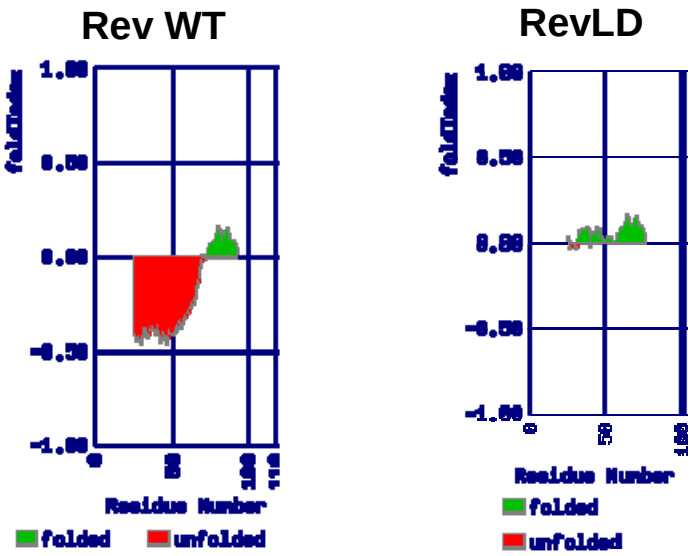


Figure 4.2 comparisons of Rev wild type and RevLD with foldindex program. The red area in the graph shows the predicted unfolded region, the green area shows the predicted folded region. The window set for prediction calculation is 50 residues.

4.1 The expression and purification of Rev Loop Deletion mutant (RevLD)

RevLD TEV plasmid was transformed into E. coli BL21DE3 Code plus cells. One clone was picked up, and the cell culture were raised in 2L 2XYT with 50mg/L Ampicillin at 37ºC. When the absorbance of OD600 of cell culture reaches between 0.4 and 0.6, IPTG was added to 1mM final concentration for induction. After 3-hour induction, the cells were spun down at 2,000 × g for 30minutes, and the cell palette was stored at -80ºC before purification.

RevLD TEV was purified by Ni-NTA column under denaturing conditions. First, the cell pellet was thawed for 15 minutes on ice. Then, the cell pellet was resuspended in 30 mL of Denaturant Binding Buffer (8M urea, 10 mM imidazole, 100 mM NaCl, 20 mM Tris, and 5mM 2-mercaptoethanol at pH 7.5). The urea was pretreated with deion resin

AG 501-X8 (Bio-Rad) right before usage for 45 minutes to prevent the possible protein backbone damage. After shaking the suspension on a vortex mixer for 5 minutes to allow for cell lysis, the cell suspension was sonicated for 30 seconds on ice for 10 cycles with 30 second cooling periods between cycles. Next, the cell suspension was centrifuged at 18,000 rpm for 30 minutes to pellet the cellular debris. The supernatant was loaded onto a Ni-NTA column, which has pre-equilibrated by Denaturant Binding Buffer. The supernatant liquid was allowed to flow through by gravity and the column washed with the Denature Binding Buffer. After extensive washing of the column (10 column volumes), RevLD TEV was eluted with an imidazole gradient from 10mM to 1M over 5 column volumes.

To remove the fusion tag, RevLD TEV was first dialyzed in TEV cleavage buffer (5 mM 2-mercaptoethanol, 100 mM NaCl and 20 mM Tris at pH 8.0). The concentration of the protein was then measured by UV spectrometer at OD_{280} with the extinction coefficient as 10930 M^{-1} cm^{-1}. The cleavage reaction was performed by adding TEV protease in 1:10 molar ratio with 1mg/ml RevLD TEV at 16ºC overnight. The fusion tag, TEV protease and any uncleaved RevLD TEV were removed by again passing the reaction mixture through Ni-NTA column. The purity of RevLD was checked by SDS PAGE, (Figure 4.3).

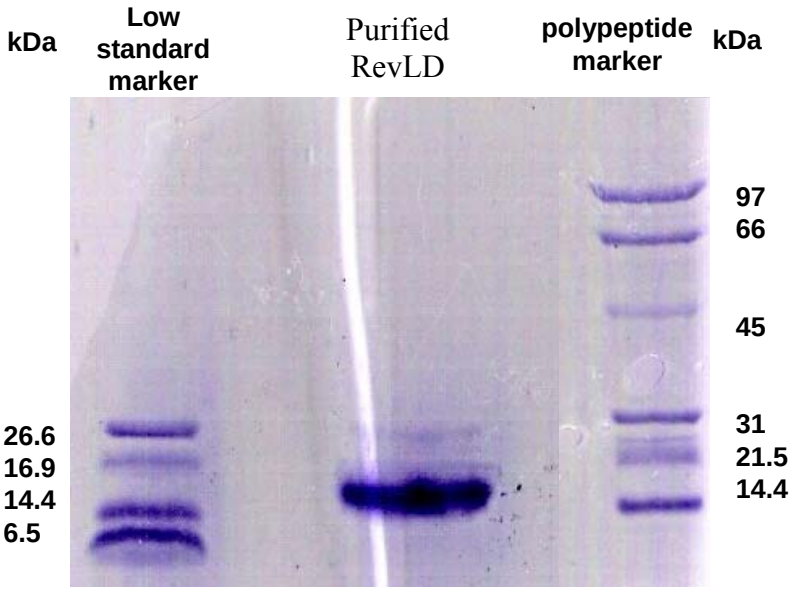


Figure 4.3 SDS-PAGE of RevLD. After the cleavage of TEV protease, RevLD were collected from the Ni-NTA column flow-through fraction. The protein was loaded onto 15% SDS-PAGE gel to check the purity. RevLD was about 90% pure with little large contaminates which might come from self-assembly of the protein.

4.2 MALDI TOF-MS for purity check of RevLD

MALDI (matrix-assisted laser desorption/ionization), is one of the most successful ionization methods for mass spectrometric analysis and investigation of large molecules. The sample is embedded in a chemical matrix (ca. 1000 x molar excess) that greatly facilitates the production of intact gas-phase ions from large, nonvolatile, and thermally labile compounds such as proteins. The matrix plays a key role in this technique by absorbing the laser light energy and causing a small part of the target substrate to vaporize. It is essential for the matrix to be in excess thus leading to the analyte molecules being completely isolated from each other. The formation of the homogenous 'solid solution' is required to produce a stable desorption of the analyte. When the chromophore of the matrix couples with the laser frequency causing rapid vibrational excitation, the clusters are ejected from the surface. These clusters consist of analyte

molecules surrounded by matrix and salt ions. The matrix molecules evaporate away from the clusters to leave the free analyte in the gas-phase. The photo-excited matrix molecules are stabilized through proton transfer to the analyte. Cation attachment to the analyte is also encouraged during this process. These ionization reactions take place in the desorbed matrix-analyte cloud just above the surface. The ions are then extracted into the mass spectrometer for analysis. MALDI TOF-MS has developed into a valuable tool in the biosciences for obtaining accurate mass with the sensitivity typically in the 1-10 picomole range.

The RevLD sample was desalted by Zip Tip (Millipore). The C-18 Zip Tip was preequilibrated with ACN 80%, 0.1% TFA, and washed in the same Zip Tip with aqueous 0.1% TFA. Then the ^{15}N RevLD sample was loaded on the Zip Tip by aspirating and dispensing the 5μL 2mg/ml sample 10 times. The tip was aspirated by wash solution (0.1% TFA in water) and dispensed for 5 times. The protein was eluted by dispensing 1μL of elution solution (ACN 80%, 0.1% TFA) into a clean vial and aspirated and dispensed the sample 10 times through the Zip Tip. Then the sample was loaded on the MALDI plate directly from the Zip Tip. One μL of MALDI matrix (alpha-cyano) was dispensed on the same spot. The matrix and the sample were mixed by aspirating and dispensing twice. Finally, the matrix was allowed to completely dry for 2 hours before MALDI analysis. It should turn white if there is enough matrix. Then the sample was loaded into the machine, and started with a standard method. The laser power was adjusted until a peak, or peaks, is observed.

There are three peaks that were observed. The major one (11138.20) is RevLD with an error of ±1.7 Da. The 5563.84 Da peak is RevLD protein in the +2 charge state. The 22284.06 Da peak is a RevLD protein dimer. The molecular weight of RevLD calculated from MALDI result is consistent with the expected molecular weight calculated from primary sequence, (Figure 4.4).

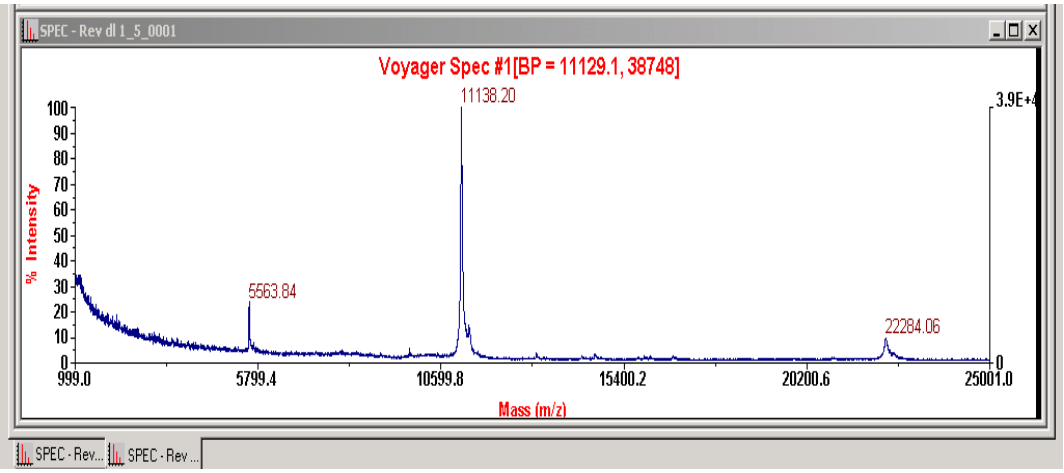


Figure 4.4 MALDI spectrum of RevLD. The estimated molecular weight of RevLD is 11136.5. The major one (11138.20) came from RevLD with an error of ±1.7 Da. The 5563.84 Da peak came from RevLD protein in the +2 charge state. The 22284.06 Da peak should be a RevLD protein dimmer.

4.3 The CD analysis of RevLD

The CD analysis of RevLD will provide the secondary structure information. 1ml of 0.1mg/ml RevLD Sample was dialyzed in 10mM potassium phosphate buffer. The sample was filtered before loading into a 0.1cm path length cuvette, which had been washed with the distilled water, ethanol, lab soap solution and then distilled water again. The JASCO J-810 CD spectropolarimeter was started by turning on the water and nitrogen first, and then turning on spectropolarimeter and computer. The temperature control was set at 20°C. The spectrum measurement was carried out from 250nm to 200nm after waiting for 30 minutes to stabilize the UV lamp. The data was only reliable if the HT number is under 800 volts.

RevLD CD spectrum contains two minima in ellipticity, one at 222 nm and the other at ~212 nm, a signature of α-helical secondary structure, (Figure 4.5). The secondary structure calculation was also carried out by the secondary structure deconvolution program (ACDP). The calculation indicates that there is about 60% of α helix, and 18%

of β strand (Table 4.1). The Rev wild type circular dichroism spectrum shows roughly 50% of α-helical structure and 25% of β strand. Thus, RevLD contains more α-helical structure, and less β strand structure than Rev wild type.

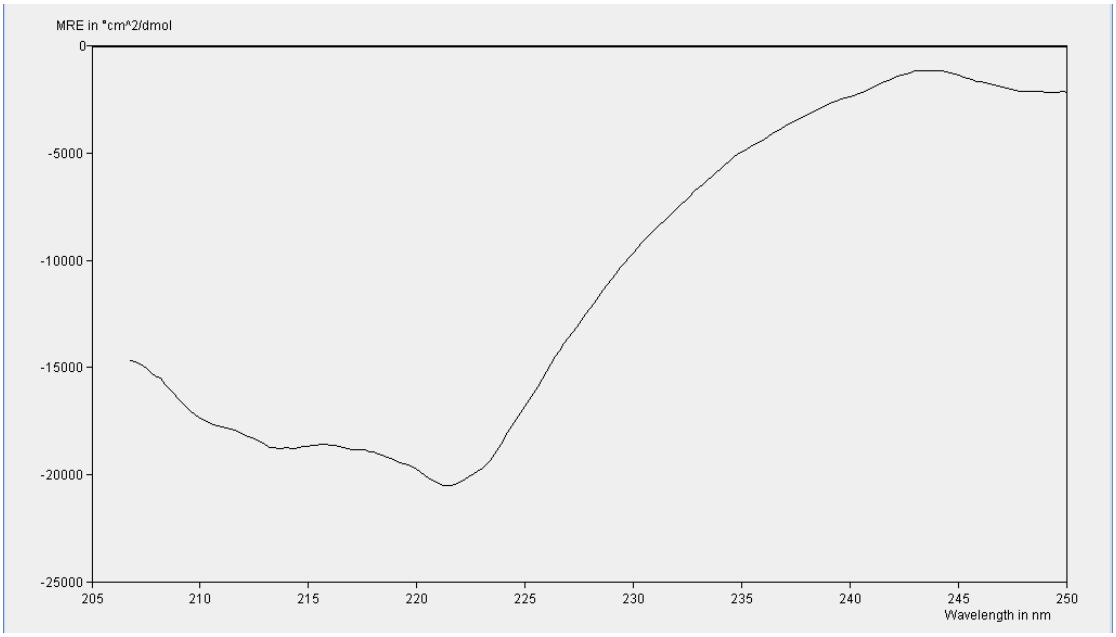


Figure 4.5 Circular Dichroism of Rev loop deletion. Two drops at 222nm and 212nm demonstrated that RevLD mainly contains alpha helix structure.

Table 4.1 Rev loop deletion secondary structure prediction

structure	K2D backpropagation neural network
Alpha helix	0.592
Beta strand	0.181
Unordered	0.227
Sum	0.999

4.4 The analytical ultracentrifugation of RevLD

Analytical ultracentrifugation (AUC) is a widely used biophysical tool for characterizing the association mechanisms of biomolecular complexes and determining

complex stoichiometry. One big advantage of AUC is that it allows one to characterize the macromolecules' behavior in solutions. In contrast to many biophysical techniques, biomolecules are characterized during AUC in their native state under relevant biophysical conditions. Because the experiments are performed in solution, there are no complications caused by interactions with matrices or surfaces that can obscure interpretation of certain types of commonly used experiments, such as gel filtration.

AUC can be used to perform two types of experiments, sedimentation velocity and sedimentation equilibrium. Sedimentation velocity is a hydrodynamic technique and is sensitive to the mass and shape of the macromolecular species. In a sedimentation velocity experiment, a moving boundary is formed under a strong centrifugal field. A series of scans with measurements of sample concentration (c) as a function of radial distance (r) are recorded at regular intervals to determine the rate of movement and broadening of the boundary as a function of time.

In a sedimentation velocity experiment, application of a sufficiently large centrifugal force field leads to movement of solute molecules toward the bottom of the centrifuge cell. The rate of sedimentation is described by the Svedberg equation:

$$s = \frac{u}{\omega^2 r} = \frac{M(1-\bar{v}\rho)}{N_A f} = \frac{MD(1-\bar{v}\rho)}{RT}$$

Where u is the observed radial velocity of the macromolecule, ω is the angular velocity of the rotor, r is the radial position, $\omega^2 r$ is the centrifugal field, M is the molar mass, υ over-bar is the partial specific volume of the protein, ρ is the density of the solvent, N_A is Avogadro's number, f is the frictional coefficient, D is the diffusion coefficient, and R is the gas constant. The s-values are commonly reported in Svedberg (S) units, which correspond to 10^{-13} sec.

Movement of the solute away from the air-solvent interface, the meniscus, in a sedimentation velocity experiment leads to formation of a solute concentration gradient, called the boundary. Because it is a concentration gradient, the boundary sediments and diffuses with time, leading to "boundary spreading" over the course of the experiment. The evolution of the concentration distribution of macromolecular species χ as a function of time and radial position under the influence of sedimentation and diffusion in the sector-shaped ultracentrifugal sample cell is described by the transportation equation.

$$\frac{\partial \chi(r,t)}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[rD\frac{\partial \chi(r,t)}{\partial r} - s\omega^2 r^2 \chi(r,t) \right]$$

A sedimentation coefficient distribution $c(s)$ can be defined with $a(r,t)$ denoting the observed sedimentation data, $c(s)$ the concentration of species with sedimentation coefficients between s and $s + ds$, and $\chi(s,D(s),r,t)$ is defined by the transportation equation above.

$$a(r,t) = \int c(s)\chi(s,D(s),r,t)ds + \varepsilon$$

SEDFIT software is used for the analysis of AUC data by direct fitting with numerical solutions of the transportation equation for c (s) analysis. C (s) analysis is possible with several variants using prior knowledge, such as conformational change models, or in general model that reveals a weight-average frictional ratio of the molecules. For the discrete transportation equation fitting, different models could be implemented.

RevLD was examined at four protein concentrations of 9.37, 3.64, 1.55 and 0.44 mg/ml after dialysis against 100 mM NaCl, 20 mM Bis-Tris, pH 6.8, 1 mM TCEP, at 4ºC, 20,000 RPM in a Beckman Optima XL-I analytical ultracentrifuge. Data were analyzed with the SEDFIT package. The general steps are: loading data from the entire

sedimentation process, use of systematic noise decomposition (and subtraction), modeling with finite element solutions of the transportation equation. If we expand the scale of the continuous sedimentation distribution c(s) with maximum entropy regularization shown above, it can be seen that the c(s) analysis reveals the presence of several oligomeric species and a smaller species. Sedimentation velocity analysis using c(s) analysis with SEDFIT showed a single, unimodal boundary at about 60 S (Figure 4.6). The precise location and amplitude of the smaller peaks in the c(s) analysis are dependent on the exact choice of data span and number of points used in the analysis and, therefore, no definite assignments could be made.

Table 4.2 Rev loop deletion velocity sedimentation.

RevLD concentration (mg/ml)	Sedimentation speed (rpm)	Estimated S value
9.37	20,000	58
3.64	20,000	59
1.55	20,000	59

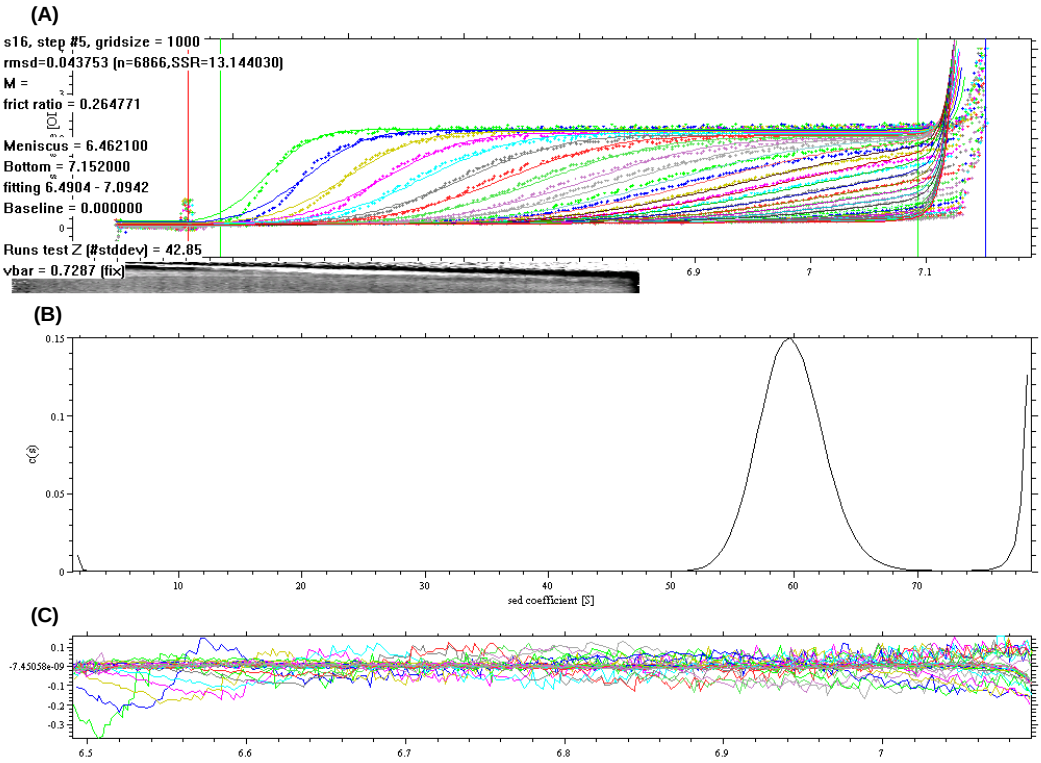


Figure 4.6 SEDFIT of Rev loop deletion (A) 31 scan profile of RevLD (B) fitted S value of RevLD (C) Residual check of individual fitting

4.5 The HSQC spectrum of RevLD

The AUC sedimentation velocity experiment shows that the RevLD protein stays in solution as a 60S oligomer. Whereas RevC, which is part of RevLD, behaves as an unfolded monomer in solution. One way to see the flexible region of a big complex is HSQC NMR experiment.

The ^{15}N labeled RevLD protein was prepared the same way as ^{15}N RevC. The labeled RevLD was purified the same way as unlabeled RevLD protein. The pure ^{15}N labeled RevLD protein was dialyzed in NMR buffer 20mM Bis-Tris, 100mM NaCl, 1mM TCEP at pH6.8. After the protein was concentrated to 0.5mM, D_2O was added to 10% final

concentration. The NMR sample was filtered by 0.2um filter before loading into NMR tube.

The HSQC experiment was set on Varian INOVA 800 NMR machine. The data was processed by NMR pipe and analyzed by NMR view (Figure 4.7).

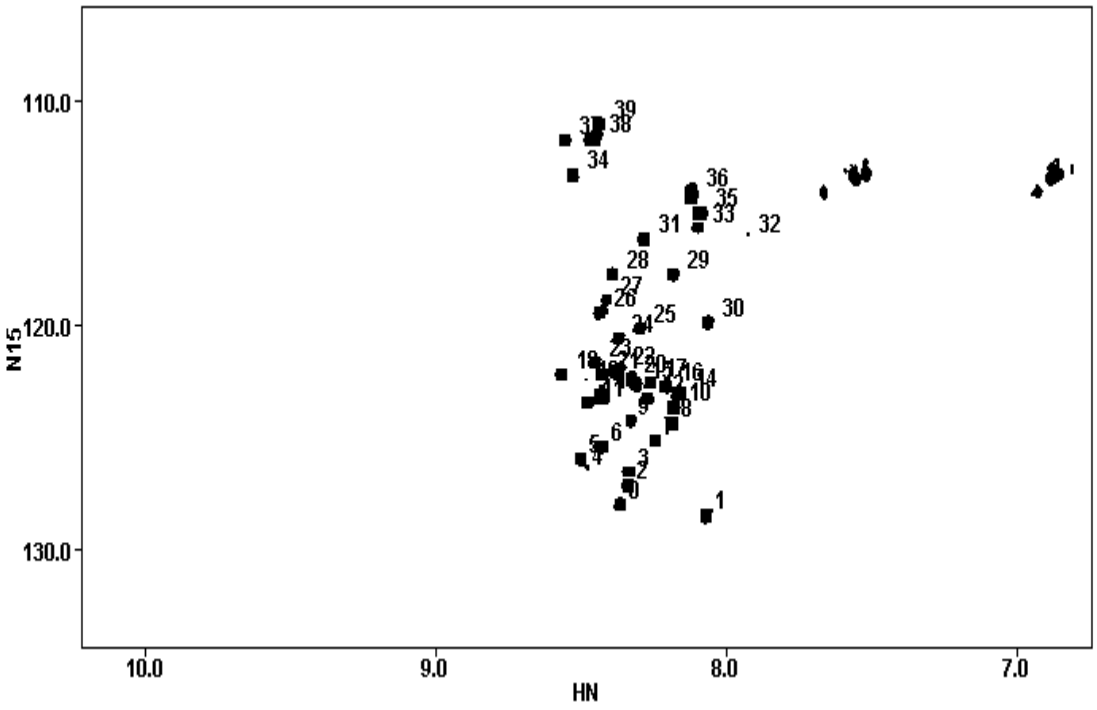


Figure 4.7 Rev loop deletion ^{15}N HSQC. 40 backbone peaks of RevLD were observed from HSQC. The amide proton signals were in the 8.0 to 8.5 region.

Considering the large size of the oligomer (60S), the RevLD complex would tumble very slowly and broaden all the HSQC peaks. But any flexible part of the complex, which can move freely in solution, can be seen. In the RevLD HSQC spectrum, there are 40 backbone peaks. Of all of these peaks, their amide proton signals are within the 8.0 to 8.5 ppm region, which indicates an unordered character. The next question is what is the location of the unordered region.

When the HSQC spectra of RevLD and RevC are overlaid, all the RevLD peaks are under the RevC peaks (Figure 4.8). There are no RevLD peaks which are not shown in the RevC spectrum. This indicates the flexible region of RevLD is located in the Rev C-terminus. But some RevC HSQC peaks are missing in the RevLD HSQC spectrum. This can be explained that part of the Rev C-terminal is interacting with the N-terminal helical region, and thus tumbles with the oligomer and their HSQC signals are broadened.

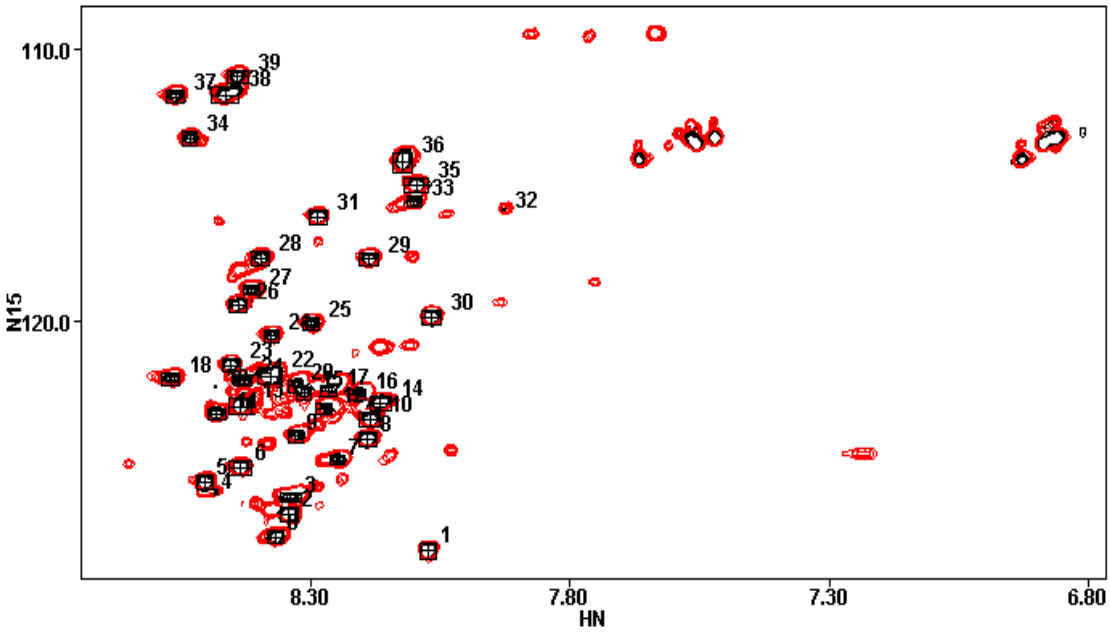


Figure 4.8 ^{15}N HSQC comparison of Rev loop deletion (black) and RevC (red). RevC HSQC spectrum includes all RevLD peaks, which indicated the RevLD HSQC signals were from the protein's C-terminus.

4.6 The TEMPO modification of RevLD

Using a spin labeled reagent, distance information can be obtained for a labeled protein. The most used is nitroxide spin label, which is incorporated in chemically stable molecules such as 2,2,6,6-tetramethyl-1-piperidine-1-oxyl (TEMPO). The distance information is characterized by the paramagnetic relaxation of NMR resonances caused

by TEMPO. The 4-(2-Iodoacetamido)-TEMPO group contains a nitroxide moiety. The nitroxide radical is known to be a good broadening reagent for NMR signals due to its long electronic relaxation time. It presents an unpaired electron, providing an efficient paramagnetic mechanism for the relaxation of neighboring nuclei by electron–nuclear dipolar coupling. The relaxation rate enhancement depends on the distance of the nuclei from the unpaired electron, which can be used to map distances of nuclei from the spin labeled residue. Compared to the nuclear Overhauser effect, which is proportional to l/r^6 and thus is short range in nature (<5 Å), the unpaired electron–nucleus interaction is relatively long-range with distances up to 20 Å. We use TEMPO -labeled RevLD protein to perturb the cross-peaks in ^{15}N–^{1}H HSQC spectrum in a distance-dependent fashion. The perturbed cross- peaks would show appropriate distances to the labeled residues (cysteine).

To label ^{15}N RevLD, 4-(2-Iodoacetamido)-TEMPO was dissolved in DMF. Then, ^{15}N Rev Loop Deletion protein was dialyzed into PBS buffer and concentrated by YM3 centriprep to 1.5mg/ml. The label reaction (Figure 4.9), was carried out by adding 4-(2-Iodoacetamido)-TEMPO in 10-fold molar ratio to the Rev Loop deletion concentration. The reaction solution was kept in the dark at 25°C for 2 hours. The reaction is stopped by dialyzing with the NMR buffer, 20mM Bis-Tris, 100mM NaCl, and 1mM TCEP at pH 6.8 to remove the unbound radicals.

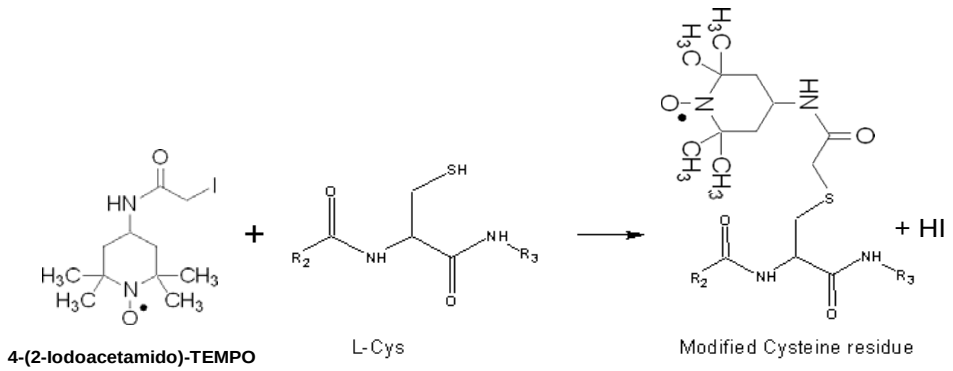


Figure 4.9 4-(2-iodoacetamido)-TEMPO modification of RevLD cysteine residues.

After concentrating with YM3 centriprep, the TEMPO labeled ^{15}N RevLD was checked by ElectroSpray Ionization Mass Spectrometry (ESI-MS). In the ESI-MS, the TEMPO labeled ^{15}N RevLD is directly injected into the machine in a solution of 1μL sample with 19ul 50% methanol and 2% acetic acid with a syringe pump. Flow rates are 0.04μl/min with 50% acetonitrile and 0.1% TFA. The high potential difference forces the spraying of charged droplets containing analyte (TEMPO labeled ^{15}N RevLD) from the needle. As the droplets traverse the space between the needle tip and the cone, solvent evaporation occurs. When the solvent evaporation occurs, the droplet shrinks until it reaches the point that the surface tension can no longer sustain the charge (the Rayleigh limit), at which point a "Coulombic explosion" occurs and the droplet is ripped apart. This produces smaller droplets that can repeat the process as well as naked charged analyte molecules. These charged analyte molecules can be singly or multiply charged. The peaks observed are due to the multiple charging affect, and are roughly Gaussian distributed. MassLynx (Micromass UK Ltd., Manchester, UK) software was used to control the machine and analyze the data. Although the sample contains salts and buffers which contaminate the spectrum, the multiply charged signals of TEMPO labeled ^{15}N RevLD could be identified (Figure 4.10). The calculated molecular weight of TEMPO

labeled ^{15}N RevLD from ESI-MS spectrum is 11722 Da with the error of 256ppm to the estimated molecular weight (11717 Da). From the spectrum, over 75% of RevLD had been fully labeled by TEMPO.

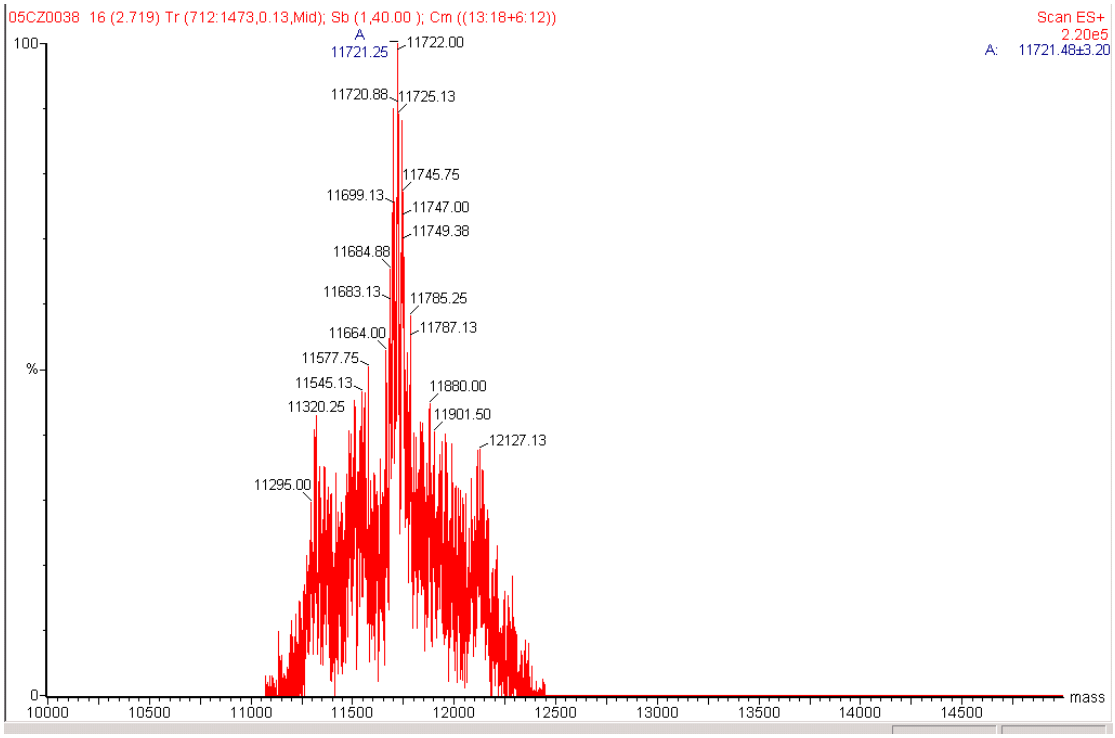


Figure 4.10 ESI-MS analysis of TEMPO labeled Rev loop deletion. Over 75% of RevLD was fully labeled with two TEMPO molecules.

4.7 The HSQC analysis between RevLD and TEMPO labeled RevLD

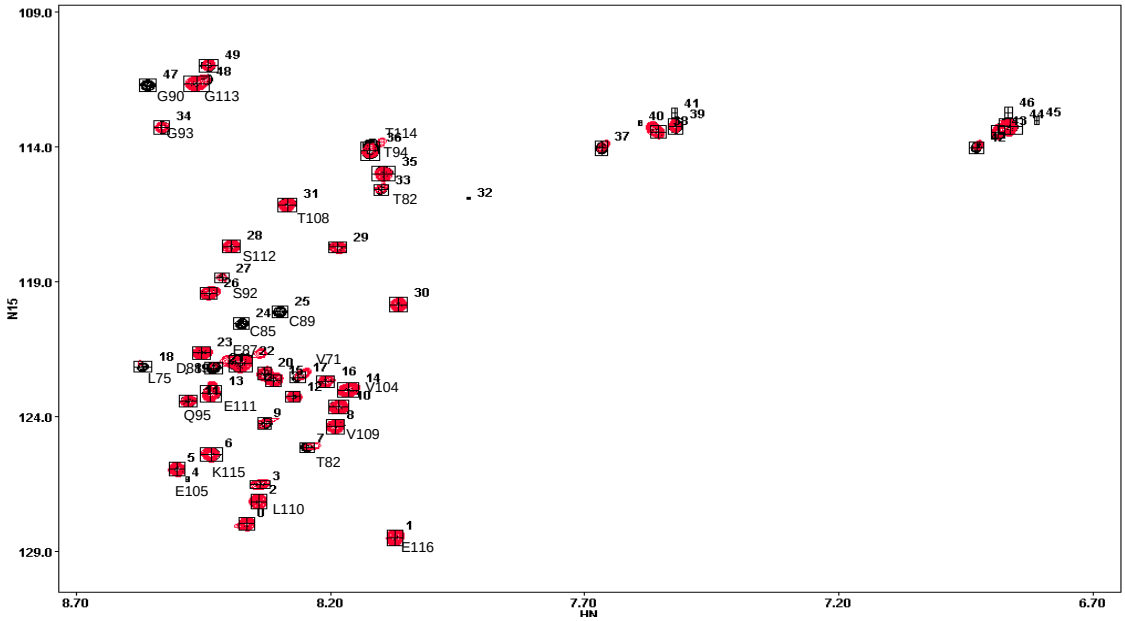


Figure 4.11 15N HSQC overlay of Rev (black) and TEMPO-RevLD (red). Two cysteines peaks were broadened off in the TEMPO-RevLD spectrum. Some other residues peaks were also broadened in the TEMPO-RevLD spectrum.

The HSQC spectra of freshly made RevLD and TEMPO-RevLD were overlaid by the NMRview program (Figure 4.11). Because RevLD and RevC HSQC can be well overlaid, the assignment of RevC can be used to analyze the RevLD HSQC spectra. As expected, two cysteine residues are fully broadened by TEMPO. The residues close to labeled cysteines also are broadened, such as Gly90, Asp88 and Glu87. The relative HSQC peak intensity changes from RevLD to TEMPO-RevLD were plotted (Figure 4.12). Within the assigned HSQC peaks, Thr94, which is the fifth residue from Cys89 has been assigned,

has the relative intensity decrease to little less than 0.8. Considering this decrease is significant caused by TEMPO, the cut off is set at 0.8 and excludes the side chain signals, there are 12 HSQC peaks which have the relative intensity decrease for 0.8 or less. There are 14 residues from Arg80 to Cys85-TEMPO to Cys89-TEMPO to Thr84. Although most of the broadened peaks are from the residues flanking the two cysteines, there is one broadened peak from residue Leu75, which is 10 residues away from the nearest cysteine, Cys85. The broaden effect of Leu75 is even more than Glu87, which is only two residues away from both cysteines. It demonstrates that Leu75 is close to Cys85 and/or Cys 89. The average Leu chemical shift of amide proton for beta strand is 8.58, which is far from the average chemical shift for random coil (7.99). The Leu75 chemical shift for amide proton is over 8.5, thus placing Leu75 within a beta strand region.

In summary, RevLD forms oligomers in solution with a flexible C-terminal region. This flexible C-terminus behaves similar to RevC. And within the C-terminus, there is at least a beta strand region containing Leu75, which is close to Cys86 and/or Cys89.

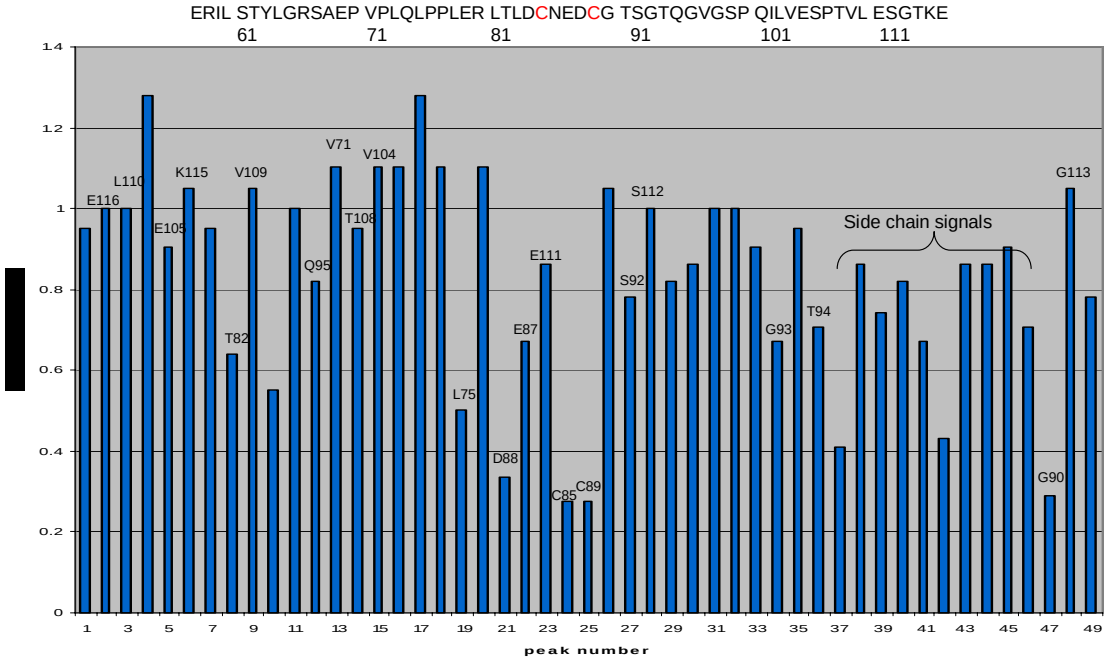


Figure 4.12 the relative ^{15}N HSQC peak intensity changes from RevLD to TEMPO labeled RevLD. The intensities of the two TEMPO labeled cysteines decreased most. The residues next to the cysteines were also observed intensity decreases. Leu75, which is 10 residues away from Cys85 and 14 residues away from Cys89, was also observed an intensity decrease.

CHAPTER 5: THE INTERACTION OF HUMAN B23 AND REV

The nucleolar protein, human B23 (hB23), is the major component in a crude nuclear extract which specifically binds to the HIV-1 Rev protein. Sedimentation analysis indicates that the hB23-Rev complex behaves like a multimer with a S value of 9S. But the binding region of hB23 to Rev is not clear.

Human B23 has several distinctive regions in its primary sequence. The non-polar N-terminal domain has three acidic regions, whereas the C-terminal contains a basic region and aromatic rich region (Figure 5-1).

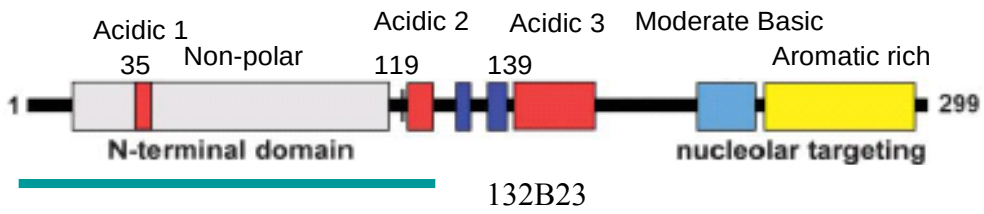


Figure 5.1 human B23 primary sequence. The numbers above the primary sequence are the start positions of three acidic regions. The truncation mutant of B23, 132B23, is shown as green bar.

The hB23 protein associates with nucleolar ribonucleoprotein structures and binds single-stranded nucleic acids. It also functions in the assembly and/or transport of ribosomes. The nucleic acid binding activities of hB23 are mainly involved in its C-terminal. The alignment of the N-terminal part of hB23 and Xenopus NO38 primary core sequence using the Clustal W program (Thompson *et al.* 1994) revealed that the two

sequences are over 76% identical (Figure 5.2) and thus probably adopt a similar tertiary structure.

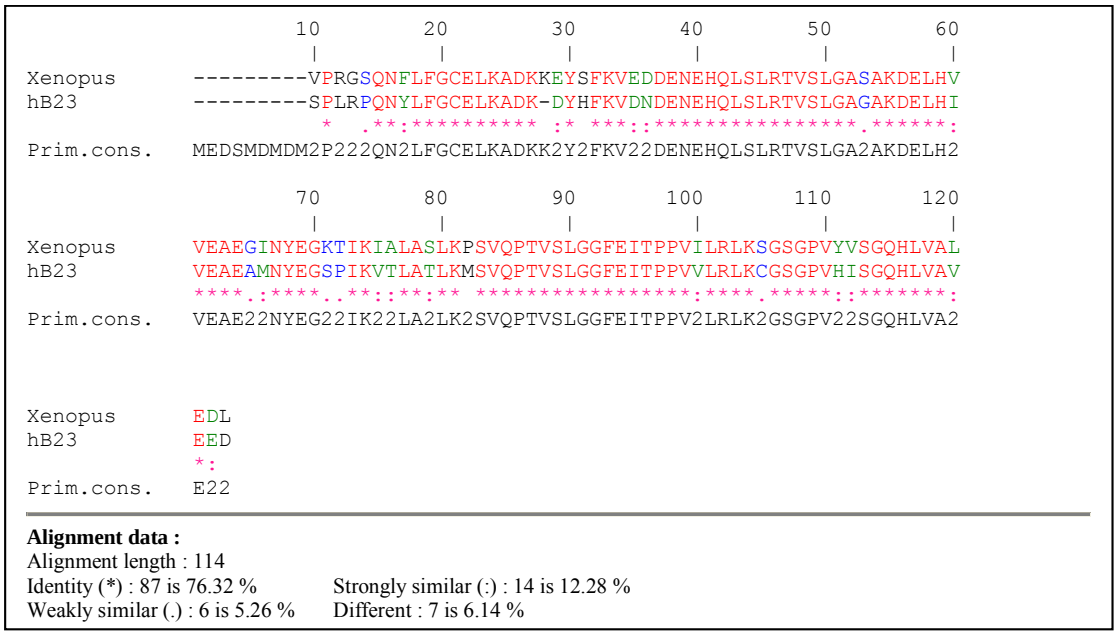


Figure 5.2 result of the alignment of N-terminal core of human B23 and NO38 by using the Clustal W program.

The structure of the N-terminal Xenopus NO38 was solved by X-ray diffraction (Namboodiri *et al.* 2004). The model of its decamer binding with histone illustrates its molecular chaperone activity. The hB23 protein would act as a chaperone in the nucleus, and the Rev-hB23 complex in the nucleolus may also serve as a storage device for Rev. We propose that N-terminal B23 also acts as a chaperone in binding with Rev protein (Figure 5.3).

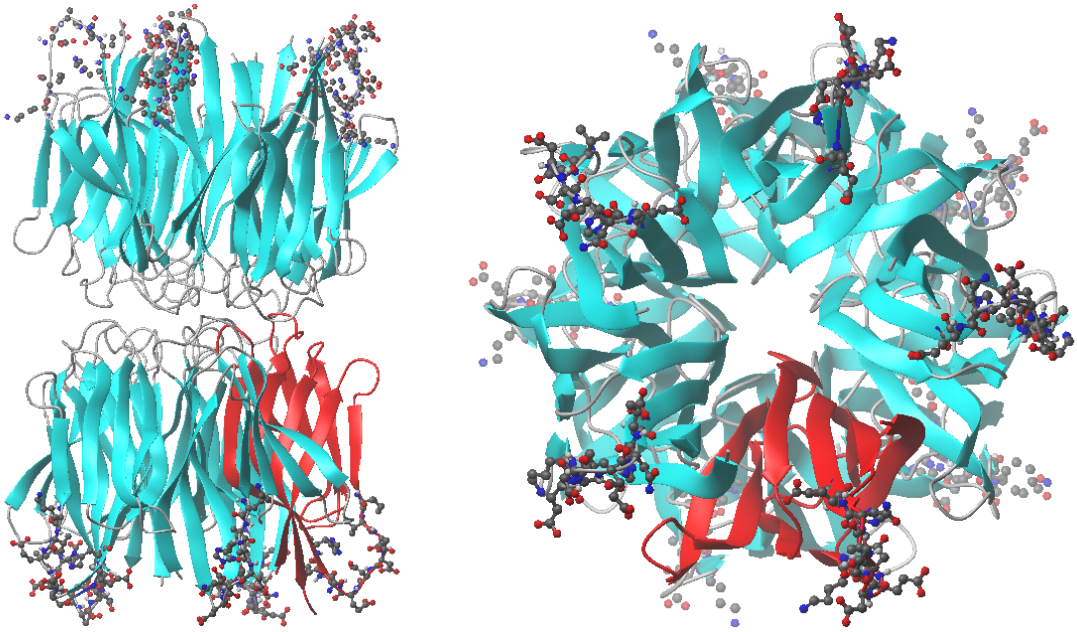


Figure 5.3 NO38 core structure, drawn from PDB ID 1XB9. One member of decamer is shown in red, the ball-stick model indicates the acidic 1 region. Left panel is side view and right panel is top view.

5.1 N-terminal B23 construct (132B23TEV) and protein purification

The hB23 plasmid was generously provided by Dr. Mark O. J. Olson at the University of Mississippi Medical Center. Two primers with the following sequence were designed to PCR the N-terminal part of B23 (1-132):

132B23TEV5′:

5′-gTA-gCA-CAC-ATA-Tgg-AAg-ACT-CgA-Tgg-ACA-Tgg-3′

132B23TEV3′:

5′-TGA-TCG-ATC-CAT-GGG-CTA-ATC-TTC-CTC-ATC-TTC-ATC-TTC-3′

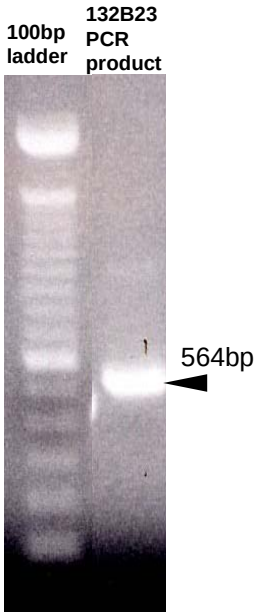


Figure 5.4 132B23 PCR product. The estimated product is 564bp. The arrow pointed the PCR product of 132B23 after PCR reaction. Compared to the 100bp ladder, the size of the PCR product of 132B23 matched the estimated.

The PCR product was purified using the QIAquick PCR purification kit (Qiagen). Both the PCR product and the vector were cleaved using the restriction enzymes, NdeI and NcoI (Figure 5.4). The correct sized insert and vector were separated by an agarose gel and further purified by the Qiagen Gel purification kit. The ligation was carried out overnight using a vector to insert ratio of 1:3. The ligation products were transformed into Subcloning Efficiency DH5 Competent Cells. Finally, the positive clones were sent to the Sealy Center for Molecular Science for sequencing. After sequence confirmation, the construct was saved in both plasmid and BL21DE3 Code plus cell forms.

After the plasmid was transformed into E. coli BL21DE3 Code plus cells, the cells were plated onto LB plate with 50mg/L Ampicillin at 37°C. After overnight growth, one clone was picked up and amplified in 50ml LB medium with 50ml/L Ampicillin for overnight growth at 37°C. Then the cells were pooled into 2L 2XYT with 50ml/L Ampicillin at 37°C. When the absorbance OD_{600} of the cell culture reached between 0.4 and 0.6, isopropyl-beta-D-thiogalactopyranoside (IPTG) was added to a 1mM final concentration for induction. After 3-hour induction, the cells were spin down at 2,000 × g

for 30 minutes. The cell palette was stored at -80°C before purification. In order to remove the fusion tag, 132B23TEV was first concentrated with Centriprep YM-3 (Millipore) by centrifugation. The concentration of the protein was then measured by UV absorption at OD_{280} with the extinction coefficient as 6520 M^{-1} cm^{-1}. The cleavage reaction was performed by adding TEV protease in 1:10 molar ratio with 1mg/ml 132B23TEV at 16C over night.

Another Nickel affinity column was needed to remove the fusion tag, TEV protease and any uncleaved 132B23TEV proteins. The fractions containing the pure 132B23 were dialyzed against storage buffer with 5 mM 2-mercaptoethanol, 100 mM NaCl, and 20 mM Tris pH 7.5.

5.2 Rev protein purification

Rev wild-type TEV plasmid was transformed into HMS174DE3 cells. The cells were initially amplified in 50ml LB medium overnight and finally transferred into 8 liter of 2XYT medium. The cells were induced at OD_{600} 0.4-0.6 by adding 1mM IPTG. After 5 hours, the cells were harvested by centrifugation at 2000g for 30 minutes and the pellets were stored at -80°C for further purification.

Rev wild-type TEV protein (RevWT-TEV) has to be purified under denatured condition with 8M urea because it is prone to aggregate at very low concentration. The urea was pretreated with de-ionizing resin AG 501-X8 (Bio-Rad) immediately before usage for 45 minutes to prevent possible protein backbone damage. Since RevWT-TEV contains his-tag in its N-terminal end, it can be easily purified by the Ni-NTA column. After adding the lyses buffer containing 10 mM imidazole, 100 mM NaCl, 20 mM Tris, 8M urea and 5mM BME at pH 7.5, the cell pellets were sonicated 10 times for 30 seconds each and with 30 seconds interval in ice. The cell lysates were centrifuged at 38000g for 30 minutes to remove cell debris. The supernatant liquid was loaded onto the Ni-NTA column which was pre-equilibrated with the lyses buffer. The column was washed until the absorbance OD_{280} reached baseline which was followed by an imidazole

gradient wash using 10mM to 1M imidazole. The RevWT-TEV fraction eluted at around 600mM imidazole. The UV spectrum of the collected protein fraction indicated a high ratio of OD_{260} : OD_{280} which strongly indicated the presence of nucleic acid contamination. The strong anion ion-exchange column, SP Sepharose column, was used to remove the excess nucleic acid. The Rev WT-TEV fraction with 8M urea was directly loaded onto SP column after the imidazole elution. The SP column was extensively washed with SP washing buffer containing 100 mM NaCl, 20 mM Tris, 8M urea and 5mM BME at pH 7.5. When the absorbance at OD_{280} reached baseline value, the REV WT-TEV protein was dissociated from the column matrix by the elution buffer with 4M guanidine and 100 mM NaCl, 20 mM Tris and 5mM BME at pH 7.5. The residual urea was further dialyzed away with the elution buffer. Finally the protein was concentrated to 10mg/ml in Centriprep YM3 by centrifugation and stored at 4°C.

The fusion tag of Rev WT-TEV was removed by TEV protease. Initially, the Rev WT-TEV protein stock was diluted ten times with the cleavage buffer, 5 mM 2-mercaptoethanol, 100 mM NaCl and 20 mM Tris at pH 8.0. This was followed by the addition of TEV protease at a molar ratio of 1 (TEV protease) : 10 (Rev WT-TEV). The cleavage reaction was carried out over night at 16°C. The reaction was terminated by adding 4M guanidine. The solution containing the protein was reloaded onto the Ni-NTA column and pure Rev WT protein collected as the flow through fraction.

5.3 The 132B23 and Rev WT interaction

The binding of 132B23 to Rev wild type can be observed by the precipitation test. The Rev wild type protein easily precipitates under 10mM HEPES, 100 mM NaCl, 5mM BME at pH7.2. However in the presence of an equimolar ratio of Rev WT with 132B23 solution in the same buffer, Rev WT can stay in solution for days. The precipitation test was applied by concentrating proteins with YM-3. In the tube containing Rev WT alone, most protein was precipitated as the pellet; whereas in the tube containing equimolar ratio of Rev WT and 132B23, the solution stayed clear. The supernatant liquid and pellet

(washed by the buffer) were checked by SDS PAGE (Figure 5.5) that indicated that 132B23 binds to Rev WT and prevents it from precipitation.

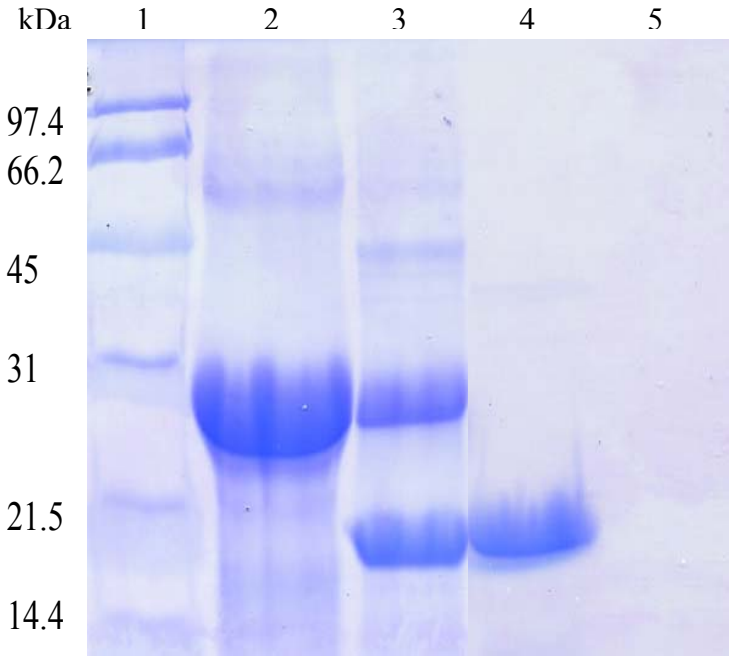


Figure 5.5 132B23 prevents RevWT precipitation. Lane 1, low standard marker; lane 2, 132B23 protein supernatant; lane 3 132B23 and Rev WT (1:1 molar ratio) supernatant; lane 4, Rev WT supernatant; lane 5, Rev WT pellet.

5.4 Biacore experiment to determine the dissociation constant

To use Surface plasmon resonance (SPR) biosensors, we can measure the interactions of biomolecules (Canziani *et al.* 1999; Malmqvist 1993; Myszka and Rich 2000; Homola *et al.* 2002; Rich and Myszka 2000). The experiments involve immobilizing one reactant on a surface and monitoring its interaction with a second component (analyte) in the solution. SPR detectors are capable of measuring the amount of a complex formed in real time without the need for fluorescent or radioisotopic labels. The instruments (BIACORE)

monitor the change in refractive index of the solvent near the sensor surface caused by the association and dissociation of the analyte–ligand complex (Fagerstam *et al.* 1992). As a quantitative tool, SPR biosensors can be used to examine molecular interactions with a wide range of affinity constants (*K*D = 1 mM to 1 pM) and kinetic rate constants (*k*a = 10^3–10^8 M^{-1} s^{-1}, *k*d = 1–10^{-6} s^{-1}) (Myszka, 2000; Joss *et al.* 1998).

Ligands are attached to the sensor chip surface using one of two methods: direct ligand immobilization or ligand capture. In the direct immobilization method, the ligand is covalently attached to the surface through chemical modification of its amine, carbohydrate, thiol or aldehyde groups. To immobilize Rev protein, the free carboxyl groups in the dextran surface of CM5 was chemically activated using a mixture of *N*-hydroxysuccinimide and 3-(*N*,*N*-dimethyl-amino) propyl-*N*-ethylcarbo-diimide. Rev was injected through one flow cell to form a covalent bond to the surface through amide linkages. This was followed by passing ethanolamine over both surfaces to deactivate the remaining activated carboxyl groups. After the immobilization of Rev wild type on the surface of flowcell 1, the surface of flow cell 2 was activated and inactivated without an immobilization step to serve as a reference surface that has undergone similar coupling chemistry. The difference in the response of flow cells 1 and 2, indicated that 80 Surface Plasmon Resonance unit (RU) of Rev wild type was immobilized on the surface of flow cell 1.

To study the interaction of Rev and 132B23, 132B23 was introduced into the Biacore X at the injection port and flowed across the sensor chip surface. An increase in SPR response during the association phase correlates with the number of Rev-132B23 complexes formed over time. When the instrument automatically switched back to running buffer after the injection, the dissociation phase started to collect information about the stability of the complex. After analyzing the data by using the Biacore evaluation kit, the dissociation constant of 132B23-Rev WT complex was calculated to be 0.1μM.

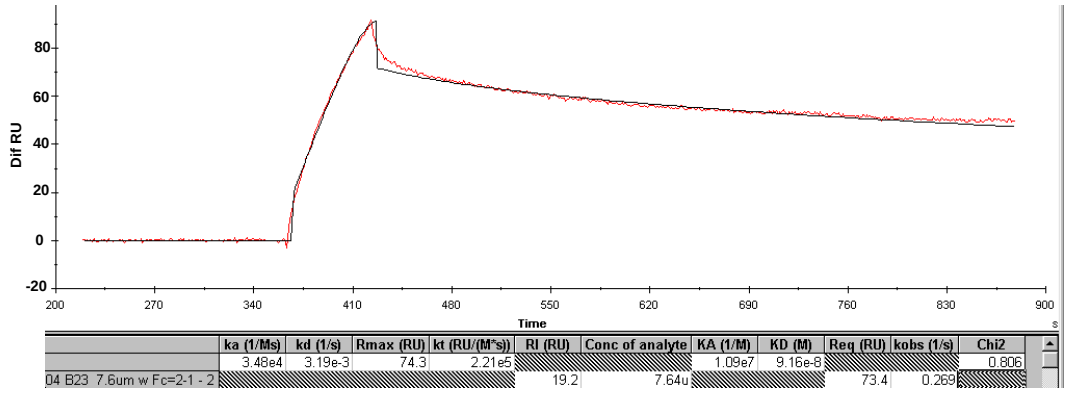


Figure 5.6 Rev WT interacts with 132B23 monitored by Biacore X. the RU changes were monitored as 200μL 7.64μM 132B23 first run though the CM5 chip, which had immobilized Rev WT on it, then the washing buffer washed the chip. The curve was fitted by Biacore analysis pack. The calculated Kd is 91.6nM with Chi^2 0.806.

5.5 MALDI-MS of Rev WT and 132B23 complex

To analyze the purity of Rev WT-132B23 complex, MALDI-MS analysis was performed (Figure 5.7). The estimated molecular weight for 132B23 and RevWT is 14660 Da and 13443 Da, respectively. In the MALDI-MS spectrum, the peaks observed for 132B23 are 14673 Da (monomer) and 29331 (dimer). The peak for Rev WT is 13434 Da. Another peak at 28090 Da correlates well to the molecular weight of a hetero-dimer of RevWT and 132B23 (28103 Da).

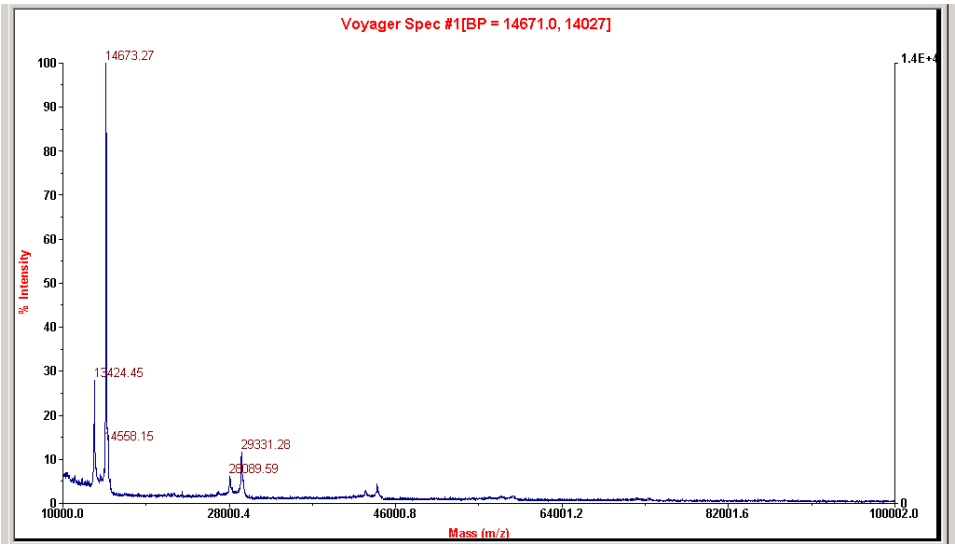


Figure 5.7 132B23-Rev wild type MALDI-MS spectrum. Peak 14673 demonstrated 132B23 monomer, 14660 Da; peak 13424 came from Rev WT monomer, 13443 Da; peak 29331 came from 132B23 dimer, 29320 Da, and peak 28090 was hetero-dimer of RevWT and 132B23, 28103 Da.

5.6 The analytical centrifugation experiment to calculate the size of the complex

To further characterize the complex stoichiometry, the analytical centrifugation experiments were performed. Firstly, the complex was examined by the velocity sedimentation experiment to estimate the S value. 3mg/ml sample of the RevWT-132B23 complex was loaded with 20mM Bis-tris, 100 mM NaCl, 1 mM TCEP at pH6.0. The experiment was carried out with 30,000 rpm at 20ºC. The data were analyzed by SEDFIT. After fitting the 133 scan curves, the S value was estimated as 8.8S (Figure 5.8), which is much smaller than that of the Rev loop deletion (50S). The estimated molecular weight is around several hundred kDa, the exact molecular weight can be calculated from equilibration sedimentation experiment.

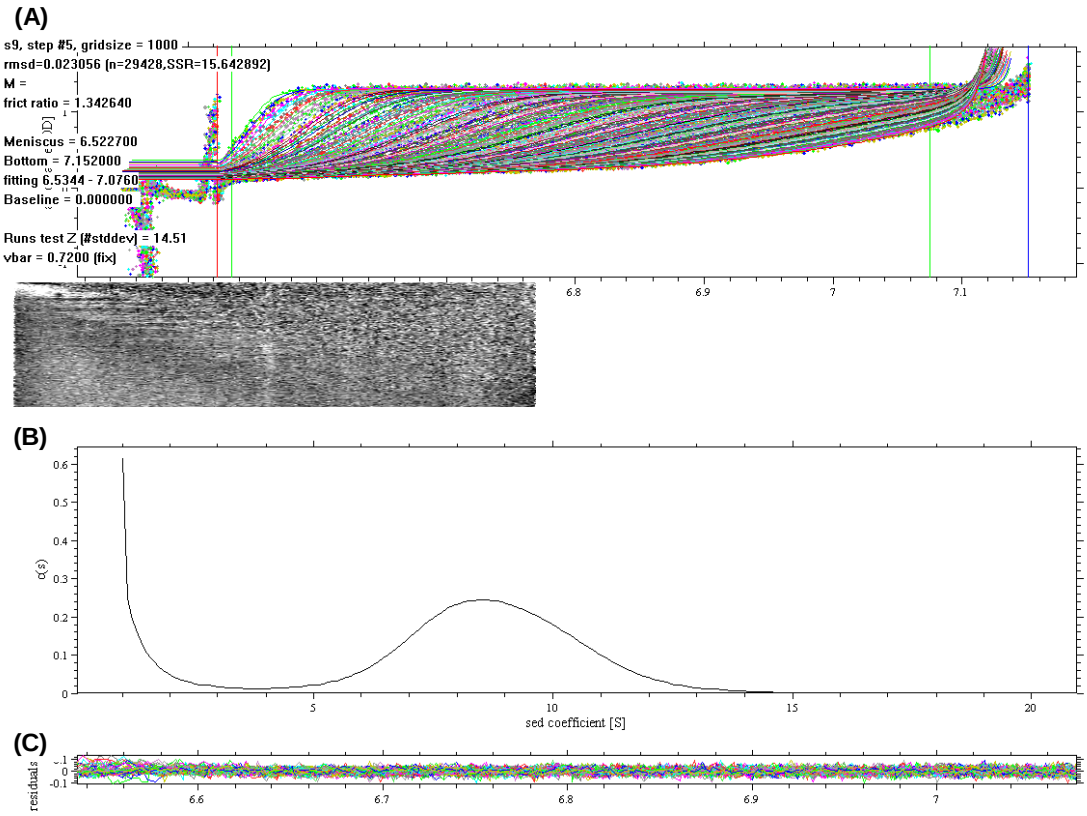


Figure 5.8 SEDFIT of Rev WT-132B23 complex (A) 133 scan profile of the complex (B) fitted S value of the complex (C) Residual check of individual fitting

In contrast to sedimentation velocity, sedimentation equilibrium is a thermodynamic technique that is sensitive to the mass but not the shape of the macromolecular species. These experiments are performed at lower speeds and measure the equilibrium concentration distribution of macromolecules that eventually forms when sedimentation is balanced by diffusion. When the centrifugal force is sufficiently small, the process of diffusion significantly opposes the process of sedimentation, and an equilibrium concentration distribution of macromolecules is eventually maintained throughout the cell.

The concentration distribution generally approaches an exponential, and for a mixture of noninteracting ideally sedimenting solutes, the measured signal as a function of radial position, $a(r)$, takes the following form:

$$a(r) = \sum_n c_{n,0}\varepsilon_n d \exp\left[\frac{M_n (1 - \bar{v}_n\rho)\omega^2}{2RT}(r^2 - r_0{}^2)\right] + \delta$$

where the summation is over all species n; $cn,0$ denotes the molar concentration of species n at a reference position $r0$; Mn, vn, and εn denote the molar mass, partial specific volume, and the molar extinction coefficient, respectively; d is the optical path length (usually 1.2 cm); and δ is a baseline offset, which compensates for differences in non-sedimenting absorbing solutes between sample and reference compartments and small non-idealities in the cell assemblies and data acquisition. SE analysis of both self-associating and hetero-associating proteins can be greatly facilitated if the total amount of soluble protein remains constant during the time course of the experiment.

Nonlinear least-squares parameter estimation is the major numerical method for SE data analysis. Methods for analysis of sedimentation equilibrium data can be divided into model-independent and model-dependent approaches. Graphical model-independent data analysis methods date from the time of the Model E centrifuge, before the advent of digital data collection and fast computers capable of global nonlinear least squares fitting. Model-independent methods are most useful at the initial stages of sample analysis, when the goal is to survey sample behavior, or for comparative analysis of samples that are too complex to be fit directly by model-dependent methods. In contrast, model-dependent analysis involve direct fitting of the sedimentation equilibrium concentration gradients to mathematical functions describing various physical models, such as single ideal species, a monomer --> n-mer self-associating system, or an A + B --> C hetero-associating system. Direct fitting is the method of choice for detailed quantitative analysis of sedimentation equilibrium data. This approach provides the best-fit values and the

associated statistical uncertainties in the fitting parameters (e.g., molecular mass, oligomer stoichiometry, association constants) and a statistical basis to discriminate among alternative physical models. The global fitting approach helps to ensure that a unique solution is obtained and greatly reduces the statistical uncertainty in the parameters.

The sedimentation equilibrium experiment was performed under 7,000 rpm with 6mg/ml RevWT-132B23 complex loaded. Beckman ORIGIN software was used to analyze the data. After curve fitting, the molecular weight of the Rev WT-132B23 complex was calculated as 275839 Da (ranges from 256898 Da and 294728 Da) (Figure 5.9). The closest stoichiometry for the complex is 10 with an expected molecular weight of 281030 Da. Rev WT could act in a similar way as histones in binding B23 core and form a decamer in the nucleus. The acid 1 region would be involved in this interaction. The model was built based on these data (Figure 5.10).

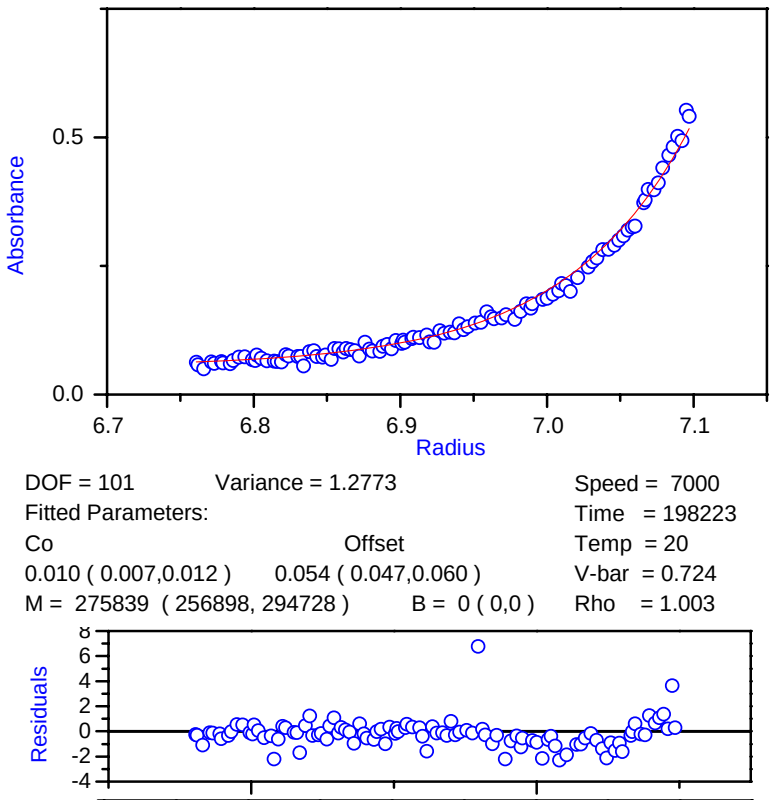


Figure 5.9 molecular weight of Rev WT-132B23 complex estimation by sedimentation equilibrium experiment. The curve fits was shown by top graph, the bottom graph is residuals check for the fit.

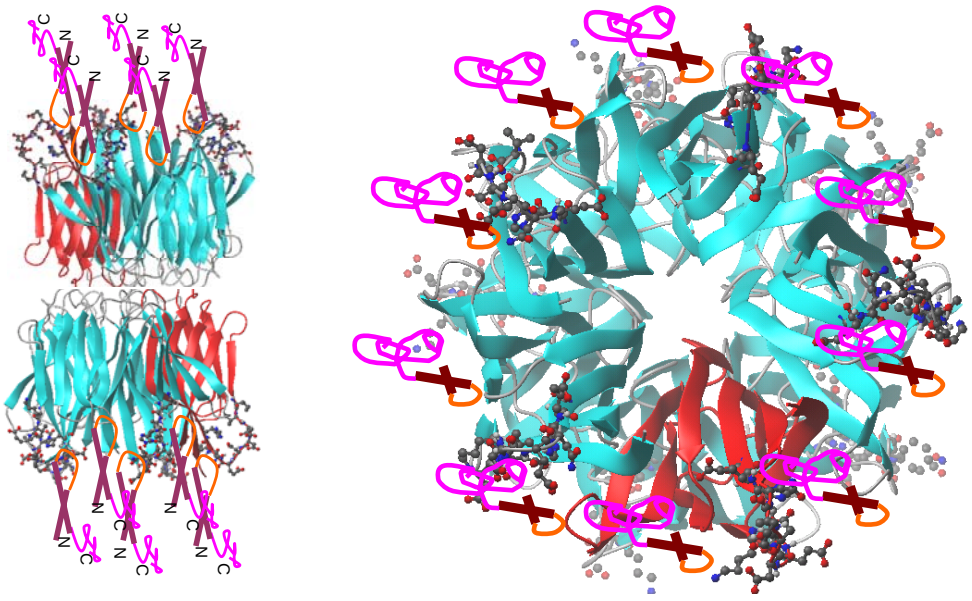


Figure 5.10 models of Rev WT-132B23 complex. The 132B23 was drawn using the NO38 core structure. The Rev WT was drawn as cartoon to show helix-loop-helix N-terminal region and disordered C-terminal region. The complex is shown as decamer. Left is side view, right is top view.

5.7 HSQC of ^{15}N RevWT and 132B23

The ^{15}N-HSQC of the complex with ^{15}N labeled RevWT and unlabeled 132B23 would show the signals only coming from RevWT. The sample was loaded onto a Varian INOVA 600 NMR machine. The HSQC spectrum was recorded at 10ºC. The FID was processed by NMR pipe and analyzed by NMR view. The ^{15}N RevWT HSQC signals show little chemical shift dispersion from 8.0 ppm to 8.5 ppm, which is quite similar to what is observed in the ^{15}N HSQC spectrum of RevC (Figure 5.11). The N-terminal signals would be broadened due to slow tumbling of the large complex.

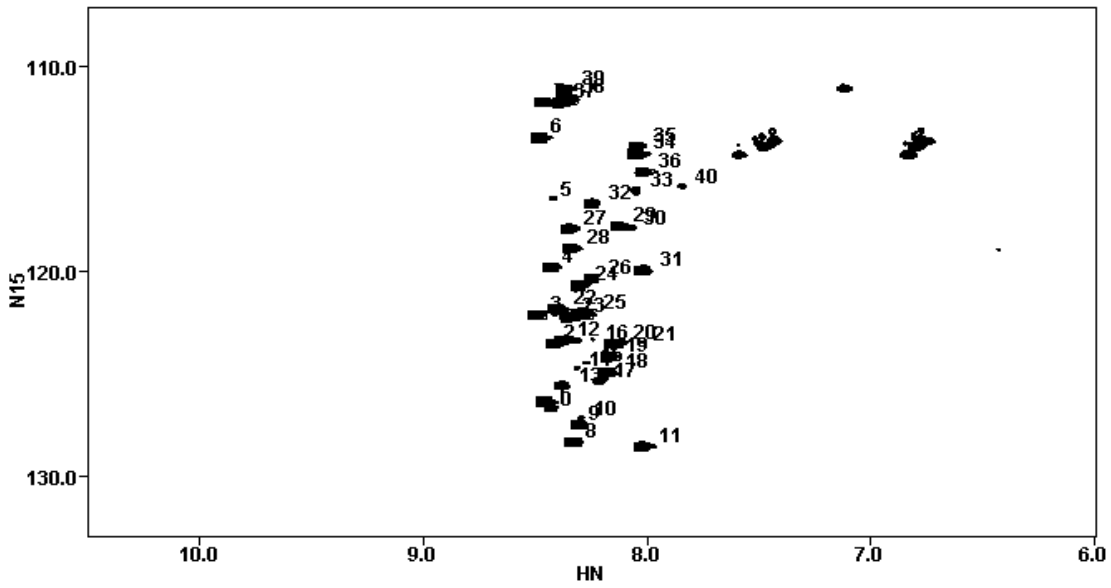


Figure 5.11 Rev WT-132B23 ^{15}N HSQC. 40 backbone signals were observed in Rev WT-132B23 complex HSQC. The amide proton chemical shifts were within 8.0 to 8.5 ppm region.

The overlay of ^{15}N HSQC spectra of Rev WT-132B23 and RevC indicates that there are no observed signals due to the additional residues present in ^{15}N Rev WT (Figure 4.11). This evidence supports the concept that Rev wild type contains a flexible C-terminal region. Also the missing signals of RevWT HSQC spectrum compared to that of RevC in the overlay would be explained as part of RevC could transform from flexible to rigid when forming complex with 132B23.

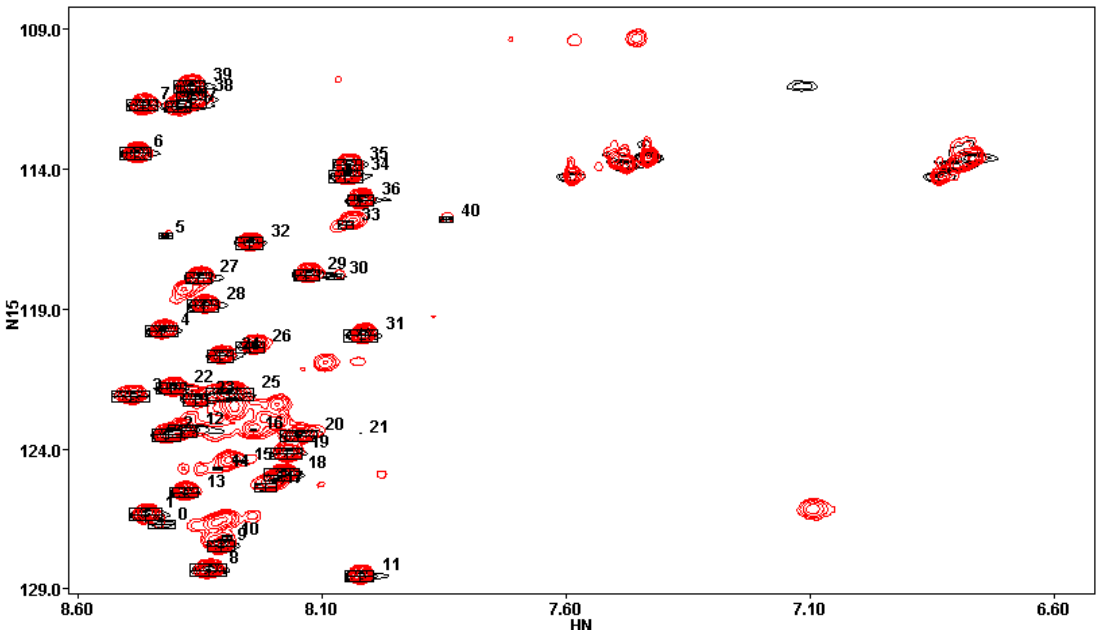


Figure 5.12 ^{15}N HSQC overlay of ^{15}N Rev WT-132B23 and ^{15}N RevC. RevC spectrum included all ^{15}N RevWT signals in the complex.

CHAPTER 6: SUMMARY

In my dissertation studies, I explored the Rev C-terminal domain (RevC), which before now has not been extensively studied. RevC behaves as a monomer in solution. The far-UV circular dichroism spectrum of RevC illustrated that over 50% of the protein is random coil and about 40% is beta strand. The partially disordered character was also seen in the gel filtration column profile of RevC. The amide proton signals in HSQC spectrum of RevC are within 8.0 and 8.5 ppm region, which is normally observed for disordered protein. The backbone assignment of RevC is in progress. Due to the fact that 10% of the residues are prolines and unstable NMR sample, the assignment was hard to finish. From the backbone assignment, the amide proton chemical shifts of some residues, e.g. , Gly93, Leu75, Glu105 and Gln95 are in the range of the average beta strand value and far from the average random coil value. This suggests that local secondary structure exists for RevC. The data collected for RevC suggest that the C-terminal domain of Rev does not form filament or high order structure in solutions. In its native state, RevC is unfolded with some local beta strand structure.

In order to see the unfolded character of RevC occurs in the context of the whole Rev protein, Rev loop deletion mutant (RevLD) was also examinated. RevLD stays in solution at as high as 9.6mg/ml as a polymer with an S value of 60. This high concentration is suitable for NMR studies. The overlay HSQC of RevC and RevLD demonstrated very similar spectra but with fewer peaks for RevLD HSQC. The data suggest two points: one is that RevLD has a region that forms the filament, and the HSQC signals of that region were broaden out; the other is the C-terminus of RevLD behaves quite the same as RevC. Based on these two points, the N-terminus of RevLD, which is suggested as a helical rich domain by far-UV circular dichroism spectroscopy, is the only component for filament formation. Considering that the N-terminal half is well

packed, whereas the C-terminal half is partially disordered, the in-side part should be the N-terminal half with flexible C-terminal half outside. This orientation was also reported in the structure of HIV-1 Rev filaments (Steven *et al* 1998). However, the peaks showing in RevLD HSQC is coming from freely moved C-terminal region, which should not contribute to the filament formation, as reported by Steven *et al*(1998). The well packed N-terminal half of Rev is the only region involved in filament formation. One Rev mutant, which contains only the Rev N-terminus, is more prone to aggregate compared to Rev wild type (data not shown). The data suggest a model for the Rev filament that contains a core from the Rev N-terminus, which is well packed and associated with each other, and a segmentally disordered C-terminus on the outside. In order to map the spacal interactions within RevLD, 4-(2-Iodoacetamido)-TEMPO was used to label ^{15}N RevLD cysteine residues to broaden the NMR signals up to 20 Å away. The overlay of HSQC with labeled and unlabeled RevLD illustrated that two labeled cysteines had the largest broaden effect. The residues flanking the two labeled cysteines (up to 5 residues) in primary sequence were also noticed to exhibit the expected broaden effect. One significant broadening effect was observed for the Leu75 signal. Because Leu75 is 10 residues away from nearest cysteine, the perturbed HSQC signal of Leu75 suggests that Leu75 is close to Cys85 and/or Cys89 in space. This implied that the local beta strand region (containing Leu75) has some interactions with the cysteine(s)-containing region.

Proteins which are misfolded are candidates for degradation. To prevent degradation, the misfolded proteins have to bind chaperones, which will assist the misfolded protein to achieve a proper fold (Goldberg 2003). The data suggest that for Rev to contain a partially disordered C-terminus, there must be some cell factor to act as chaperone to prevent Rev from degradation. The nucleolar B23 protein is one cell factor that associates specifically with Rev. The N-terminus of B23, 132B23, was observed to associate with Rev wild type. The eqilibration sedimentation experiment demonstrated the complex as a decamer of Rev-132B23 hetero-dimer. When the HSQC spectra of complex and RevC are overlaid, they are similar to each other, except that the spectrum of the complex had

fewer peaks. This suggests that most of Rev C-terminal region in the complex behaves like the C-terminus alone in solution. It also suggests that some residues interact with the bulk part of the complex and the NMR signals were broadened out. Because the N-terminus of B23 is over 70% identical to that of NO38, the structure of 132B23 should follow the same fold as the core of NO38. Based on those facts, a model of Rev-132B23 complex was built (Figure 5.10). The Rev N-terminal helix-loop-helix region interacts with acid 1 and 2 region of 132B23. The Rev C-terminal segmentally disordered region stretches out freely.

REFERENCE LIST

Goldberg A.L. (2003) Protein degradation and protection against misfolded or damaged proteins *Nature* **426**, 895-899

Alexandrescu A.T., Abeygunawardana C., Shortle D. (1994) Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: a heteronuclear NMR study. *Biochemistry*. **33**, 1063–1072.

Andrade M. A., Chacon P., Merelo J. J. and Moran F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng* **6,** 383-390.

Andrus L., Szabo P., Grady R. W., Hanauske A. R., Huima-Byron T., Slowinska B., Zagulska S. and Hanauske-Abel H. M. (1998) Antiretroviral effects of deoxyhypusyl hydroxylase inhibitors: a hypusine-dependent host cell mechanism for replication of human immunodeficiency virus type 1 (HIV-1). *Biochem. Pharmacol.* **55,** 1807-1818.

Askjaer P., Bachi A., Wilm M., Bischoff F. R., Weeks D. L., Ogniewski V., Ohno M., Niehrs C., Kjems J., Mattaj I. W. and Fornerod M. (1999) RanGTP-regulated interactions of CRM1 with nucleoporins and a shuttling DEAD-box helicase. *Mol. Cell Biol.* **19,** 6276-6285.

Auer M., Gremlich H. U., Seifert J. M., Daly T. J., Parslow T. G., Casari G. and Gstach H. (1994) Helix-loop-helix motif in HIV-1 Rev. *Biochemistry* **33,** 2988-2996.

Bartel D. P., Zapp M. L., Green M. R. and Szostak J. W. (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. *Cell* **67,** 529-536.

Battiste J. L., Mao H., Rao N. S., Tan R., Muhandiram D. R., Kay L. E., Frankel A. D. and Williamson J. R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273,** 1547-1551.

Blanco F. J., Hess S., Pannell L. K., Rizzo N. W. and Tycko R. (2001) Solid-state NMR data support a helix-loop-helix structural model for the N-terminal half of HIV-1 Rev in fibrillar form. *J. Mol. Biol.* **313,** 845-859.

Bogerd H. and Greene W. C. (1993) Dominant negative mutants of human T-cell leukemia virus type I Rex and human immunodeficiency virus type 1 Rev fail to multimerize in vivo. *J. Virol.* **67,** 2496-2502.

Canziani G., Zhang W., Cines D., Rux A., Willis S., Cohen G., Eisenberg R. and Chaiken I. (1999) Exploring biomolecular recognition using optical biosensors. *Methods* **19,** 253-269.

Chang C. T., Wu C. S. and Yang J. T. (1978) Circular dichroic analysis of protein conformation: inclusion of the beta-turns. *Anal. Biochem.* **91,** 13-31.

Chang D. D. and Sharp P. A. (1989) Regulation by HIV Rev depends upon recognition of splice sites. *Cell* **59,** 789-795.

Cho HS, Liu CW, Damberger FF, Pelton JG, Nelson HCM, Wemmer DE. (1996) Yeast heat shock transcription factor N-terminal activation domains are unstructured as probed by heteronuclear NMR spectroscopy. *Protein Sci*. **5**, 262–269

Chou P. Y. and Fasman G. D. (1974) Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry* **13,** 211-222.

Cingolani G., Bednenko J., Gillespie M. T. and Gerace L. (2002) Molecular basis for the recognition of a nonclassical nuclear localization signal by importin beta. *Mol. Cell* **10,** 1345-1353.

Cingolani G., Petosa C., Weis K. and Muller C. W. (1999) Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature* **399,** 221-229.

Cochrane A. W., Chen C. H., Kramer R., Tomchak L. and Rosen C. A. (1989) Purification of biologically active human immunodeficiency virus rev protein from Escherichia coli. *Virology* **173,** 335-337.

Coffin J. M. (1996) HIV viral dynamics. *AIDS* **10 Suppl 3,** S75-S84.

Dahlberg J. E. and Lund E. (1998) Functions of the GTPase Ran in RNA export from the nucleus. *Curr. Opin. Cell Biol.* **10,** 400-408.

Daly T. J., Cook K. S., Gray G. S., Maione T. E. and Rusche J. R. (1989) Specific binding of HIV-1 recombinant Rev protein to the Rev-responsive element in vitro. *Nature* **342,** 816-819.

Fagerstam L. G., Frostell-Karlsson A., Karlsson R., Persson B. and Ronnberg I. (1992) Biospecific interaction analysis using surface plasmon resonance detection applied to kinetic, binding site and concentration analysis. *J. Chromatogr.* **597,** 397-410.

Felber B. K., Drysdale C. M. and Pavlakis G. N. (1990) Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein. *J. Virol.* **64,** 3734-3741.

Felber B. K., Hadzopoulou-Cladaras M., Cladaras C., Copeland T. and Pavlakis G. N. (1989) rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 1495-1499.

Fischer U., Huber J., Boelens W. C., Mattaj I. W. and Luhrmann R. (1995) The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* **82,** 475-483.

Fischer U., Meyer S., Teufel M., Heckel C., Luhrmann R. and Rautmann G. (1994) Evidence that HIV-1 Rev directly promotes the nuclear export of unspliced RNA. *EMBO J.* **13,** 4105-4112.

Fornerod M., Ohno M., Yoshida M. and Mattaj I. W. (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90,** 1051-1060.

Fukuda M., Asano S., Nakamura T., Adachi M., Yoshida M., Yanagida M. and Nishida E. (1997) CRM1 is responsible for intracellular transport mediated by the nuclear export signal. *Nature* **390,** 308-311.

Gast K, Damaschun H, Eckert K, Schulze-Foster K, Maurer HR, Mu¨ ller-Frohne M, Zirwer D, Czarnecki J, Damaschun G. (1995) Prothymosin a: A biologically active protein with random coil conformation. *Biochemistry*. **34**,13211–13218.

Ghosh S, Lowenstein JM. (1996) A multifunctional vector system for heterologous expression of proteins in Escherichia coli. Expression of native and hexahistidyl fusion proteins, rapid purification of the fusion proteins, and removal of fusion peptide by Kex2 protease. *Gene.* **176**: 249-255.

Gorlich D. (1997) Nuclear protein import. *Curr. Opin. Cell Biol.* **9,** 412-419.

Gosser Y., Hermann T., Majumdar A., Hu W., Frederick R., Jiang F., Xu W. and Patel D. J. (2001) Peptide-triggered conformational switch in HIV-1 RRE RNA complexes. *Nat. Struct. Biol.* **8,** 146-150.

Guatelli J. C., Gingeras T. R. and Richman D. D. (1990) Alternative splice acceptor utilization during human immunodeficiency virus type 1 infection of cultured cells. *J. Virol.* **64,** 4093-4098.

Hakata Y., Yamada M. and Shida H. (2001) Rat CRM1 is responsible for the poor activity of human T-cell leukemia virus type 1 Rex protein in rat cells. *J. Virol.* **75,** 11515-11525.

Hanly S. M., Rimsky L. T., Malim M. H., Kim J. H., Hauber J., Duc D. M., Le S. Y., Maizel J. V., Cullen B. R. and Greene W. C. (1989) Comparative analysis of the HTLV-I Rex and HIV-1 Rev trans-regulatory proteins and their RNA response elements. *Genes Dev.* **3,** 1534-1544.

Heaphy S., Finch J. T., Gait M. J., Karn J. and Singh M. (1991) Human immunodeficiency virus type 1 regulator of virion expression, rev, forms nucleoprotein filaments after binding to a purine-rich "bubble" located within the rev-responsive region of viral mRNAs. *Proc. Natl. Acad. Sci. U. S. A* **88,** 7366-7370.

Henderson B. R. and Percipalle P. (1997) Interactions between HIV Rev and nuclear import and export factors: the Rev nuclear localisation signal mediates specific binding to human importin-beta. *J. Mol. Biol.* **274,** 693-707.

Himathongkham S. and Luciw P. A. (1996) Restriction of HIV-1 (subtype B) replication at the entry step in rhesus macaque cells. *Virology* **219,** 485-488.

Homola J., Dostalek J., Chen S., Rasooly A., Jiang S. and Yee S. S. (2002) Spectral surface plasmon resonance biosensor for detection of staphylococcal enterotoxin B in milk. *Int. J. Food Microbiol.* **75,** 61-69.

Hope T. J. (1999) The ins and outs of HIV Rev. *Arch. Biochem. Biophys.* **365,** 186-191.

Jain C. and Belasco J. G. (1996) A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. *Cell* **87,** 115-125.

Jain C. and Belasco J. G. (2001) Structural model for the cooperative assembly of HIV-1 Rev multimers on the RRE as deduced from analysis of assembly-defective mutants. *Mol. Cell* **7,** 603-614.

Jensen T. H., Jensen A., Szilvay A. M. and Kjems J. (1997) Probing the structure of HIV-1 Rev by protein footprinting of multiple monoclonal antibody-binding sites. *FEBS Lett.* **414,** 50-54.

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.

Jones T., Sheer D., Bevec D., Kappel B., Hauber J. and Steinkasserer A. (1997) The human HIV-1 Rev binding-protein hRIP/Rab (HRB) maps to chromosome 2q36. *Genomics* **40,** 198-199.

Joss L, Morton TA, Doyle ML, Myszka DG. (1998) Interpreting kinetic rate constants from optical biosensor data recorded on a decaying surface. *Anal Biochem.* **261**, 203-210.

Kubota S., Siomi H., Satoh T., Endo S., Maki M. and Hatanaka M. (1989) Functional similarity of HIV-I rev and HTLV-I rex proteins: identification of a new nucleolar-targeting signal in rev protein. *Biochem. Biophys. Res. Commun.* **162,** 963-970.

Lee S. J., Imamoto N., Sakai H., Nakagawa A., Kose S., Koike M., Yamamoto M., Kumasaka T., Yoneda Y. and Tsukihara T. (2000) The adoption of a twisted structure of importin-beta is essential for the protein-protein interaction required for nuclear transport. *J. Mol. Biol.* **302,** 251-264.

Legrain P. and Rosbash M. (1989) Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell* **57,** 573-583.

Lucast, L. J., Batey, R. T., and Doudna, J. A. (2001). Large-scale purification of a stable form of recombinant tobacco etch virus protease. *Biotechniques* **30**, 544-550.

Madore S. J., Tiley L. S., Malim M. H. and Cullen B. R. (1994) Sequence requirements for Rev multimerization in vivo. *Virology* **202,** 186-194.

Malim M. H. and Cullen B. R. (1993) Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. *Mol. Cell Biol.* **13,** 6180-6189.

Malim M. H. and Cullen B. R. (1991) HIV-1 structural gene expression requires the binding of multiple Rev monomers to the viral RRE: implications for HIV-1 latency. *Cell* **65,** 241-248.

Malim M. H., Hauber J., Le S. Y., Maizel J. V. and Cullen B. R. (1989) The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. *Nature* **338,** 254-257.

Malmqvist M. (1993) Surface plasmon resonance for detection and measurement of antibody-antigen affinity and kinetics. *Curr. Opin. Immunol.* **5,** 282-286.

Marques S. M., Veyrune J. L., Shukla R. R. and Kumar A. (2003) Restriction of human immunodeficiency virus type 1 Rev function in murine A9 cells involves the Rev C-terminal domain. *J. Virol.* **77,** 3084-3090.

Marx J. L. (1989) New hope on the AIDS vaccine front. *Science* **244,** 1254, 1256.

McClure M. O., Sattentau Q. J., Beverley P. C., Hearn J. P., Fitzgerald A. K., Zuckerman A. J. and Weiss R. A. (1987) HIV infection of primate lymphocytes and conservation of the CD4 receptor. *Nature* **330,** 487-489.

Merelo, J.J., M.A. Andrade, A. Prieto and F. Morán. (1994) Proteinotopic Feature Maps. Neurocomputing. 6, 443-454

Mermer B., Felber B. K., Campbell M. and Pavlakis G. N. (1990) Identification of trans-dominant HIV-1 rev protein mutants by direct transfer of bacterially produced proteins into human cells. *Nucleic Acids Res.* **18,** 2037-2044.

Mogridge J, Legault P, Li J, Van Oene MD, Kay LE, Greenblatt J. (1998) Independent ligand-induced folding of the RNA-binding domain and two functionally distinct antitermination regions in the phage lambda N protein. *Mol Cell*. **1**, 265–275.

Muranyi W. and Flugel R. M. (1991) Analysis of splicing patterns of human spumaretrovirus by polymerase chain reaction reveals complex RNA structures. *J. Virol.* **65,** 727-735.

Myszka D. G. and Rich R. L. (2000) Implementing surface plasmon resonance biosensors in drug discovery. *Pharm. Sci. Technol. Today* **3,** 310-317.

Namboodiri VM, Akey IV, Schmidt-Zachmann MS, Head JF, Akey CW. (2004) The structure and function of Xenopus NO38-core, a histone chaperone in the nucleolus. *Structure*. **12**:2149-2160.

Olsen H. S., Nelbock P., Cochrane A. W. and Rosen C. A. (1990) Secondary structure is the major determinant for interaction of HIV rev protein with RNA. Science 247, 845-848.

Palmeri D. and Malim M. H. (1999) Importin beta can mediate the nuclear import of an arginine-rich nuclear localization signal in the absence of importin alpha. *Mol Cell Biol.* **19,** 1218-1225.

Park M. H., Wolff E. C. and Folk J. E. (1993) Hypusine: its post-translational formation in eukaryotic initiation factor 5A and its potential role in cellular regulation. *Biofactors* **4,** 95-104.

Peter F. (1998) HIV nef: the mother of all evil? *Immunity* **9**, 433-437.

Phan, J., Zdanov, A., Evdokimov, A. G., Tropea, J. E., Peters, H. P. K., Kapust, R. B., Li, M., Wlodawer, A., and Waugh, D. S. (2002). Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem*. **277**, 50564-50572.

Prilusky J., Felder C. E., Zeev-Ben-Mordehai T., Rydberg E. H., Man O., Beckmann J. S., Silman I. and Sussman J. L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatic.* **21,** 3435-3438.

Purcell D. F. and Martin M. A. (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. *J. Virol.* **67,** 6365-6378.

Rich R. L. and Myszka D. G. (2000) Advances in surface plasmon resonance biosensor analysis. *Curr. Opin. Biotechnol.* **11,** 54-61.

Richards S. A., Carey K. L. and Macara I. G. (1997) Requirement of guanosine triphosphate-bound ran for signal-mediated nuclear protein export. *Science* **276,** 1842-1844.

Robert-Guroff M., Popovic M., Gartner S., Markham P., Gallo R. C. and Reitz M. S. (1990) Structure and expression of tat-, rev-, and nef-specific transcripts of human

immunodeficiency virus type 1 in infected lymphocytes and macrophages. *J. Virol.* **64,** 3391-3398.

Ruhl M., Himmelspach M., Bahr G. M., Hammerschmid F., Jaksche H., Wolff B., Aschauer H., Farrington G. K., Probst H., Bevec D. and . (1993) Eukaryotic initiation factor 5A is a cellular target of the human immunodeficiency virus type 1 Rev activation domain mediating trans-activation. *J. Cell Biol.* **123,** 1309-1320.

Stade K., Ford C. S., Guthrie C. and Weis K. (1997) Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90,** 1041-1050.

Surendran R., Herman P., Cheng Z., Daly T. J. and Ching L. J. (2004) HIV Rev self-assembly is linked to a molten-globule to compact structural transition. *Biophys. Chem.* **108,** 101-119.

Szebeni A., Mehrotra B., Baumann A., Adam S. A., Wingfield P. T. and Olson M. O. (1997) Nucleolar protein B23 stimulates nuclear import of the HIV-1 Rev protein and NLS-conjugated albumin. *Biochemistry* **36,** 3941-3949.

Szebeni A. and Olson M. O. (1999) Nucleolar protein B23 has molecular chaperone activities. *Protein Sci.* **8,** 905-912.

Thomas S. L., Oft M., Jaksche H., Casari G., Heger P., Dobrovnik M., Bevec D. and Hauber J. (1998) Functional analysis of the human immunodeficiency virus type 1 Rev protein oligomerization interface. *J. Virol.* **72,** 2935-2944.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673-4680.

Thumb W., Graf C., Parslow T., Schneider R. and Auer M. (1999) Temperature inducible beta-sheet structure in the transactivation domains of retroviral regulatory proteins of the Rev family. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **55A,** 2729-2743.

Uversky V. N., Gillespie J. R. and Fink A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **41,** 415-427.

Venkatesh L. K., Mohammed S. and Chinnadurai G. (1990) Functional domains of the HIV-1 rev gene required for trans-regulation and subcellular localization. *Virology* **176,** 39-47.

Wishart, D.S., Sykes, B.D., and Richards, F.M. (1991) "Relationship between Nuclear Magnetic Resonance Chemical Shift and Protein Secondary Structure," *J. Mol. Biol.* **22**2, 311-333.

Wishart, D.S., Sykes, B.D., and Richards, F.M. (1992) "The Chemical Shift Index: A Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR Spectroscopy," *Biochemistry* **31**, 1647-1651.

Wishart, D.S. and Sykes, B.D. (1994) "The 13C Chemical-Shift Index: A Simple Method for the Identification of Protein Secondary Structure Using 13C Chemical-Shift Data," *J. Biomol. NMR* **4**, 171-180.

Watts N. R., Misra M., Wingfield P. T., Stahl S. J., Cheng N., Trus B. L., Steven A. C. and Williams R. W. (1998) Three-dimensional structure of HIV-1 Rev protein filaments. *J. Struct. Biol.* **121,** 41-52.

Weiss R. A., Clapham P. R., Weber J. N., Dalgleish A. G., Lasky L. A. and Berman P. W. (1986) Variable and conserved neutralization antigens of human immunodeficiency virus. *Nature* **324,** 572-575.

Wingfield P. T., Stahl S. J., Payton M. A., Venkatesan S., Misra M. and Steven A. C. (1991) HIV-1 Rev expressed in recombinant Escherichia coli: purification, polymerization, and conformational properties. *Biochemistry* **30,** 7527-7534.

Yang J. T., Wu C. S. and Martinez H. M. (1986) Calculation of protein conformation from circular dichroism. *Methods Enzymol.* **130,** 208-269.

Yi Q, Scalley-Kim ML, Alm EJ, Baker D. (2000) NMR characterization of residual structure in the denatured state of protein L. *J Mol Biol.* **299**, 1341-51.

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

Zolotukhin A. S. and Felber B. K. (1999) Nucleoporins nup98 and nup214 participate in nuclear export of human immunodeficiency virus type 1 Rev. *J. Virol.* **73,** 120-127.

VITA

Ye Peng was born on March 19th, 1973 to Xinsheng Peng and Deyuan Liu. After attending the University of Science and Technology of China for his Bachelor's degree of Biology, Ye matriculated at the University of Texas Medical Branch in 1999.

Ye gained significant research experience while at the University of Texas Medical Branch. Two publications are in the track of submission.

Ye can be contacted at 928 Post office Street, Apt #7, Galveston 77550.

Education

B.S., May 1997, the University of Science and Technology of China, Hefei, P.R.China.

Publications

The NMR studies of HIV-1 Rev structural mutants. (in preparation)

The interaction of HIV-1 Rev and human B23 N-terminal core (in preparation)