MUTATION OF Trp131RESIDUE OF HIV-1 INTEGRASE TO T131F AND STUDYING ITS INTERACTIONS WITH LEDGF/P75

G SHRAVAN KUMAR

A Dissertation Submitted to
Indian Institute of Technology Hyderabad
In Partial Fulfillment of the Requirements for
The Degree of Master of Technology

Department of Biotechnology Engineering

JUNE, 2015

Declaration

I declare that this written submission represents my ideas in my own words, and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources that have thus not been properly cited, or from whom proper permission has not been taken when needed.

G. Shravan Kumar

G SHRAVAN KUMAR

BO13M1004

Approval Sheet

This thesis entitled **"MUTATION OF Trp131RESIDUE OF HIV-1 INTEGRASE TO T131F AND STUDYING ITS INTERACTIONS WITH LEDGF/P75"** by G SHRAVAN KUMAR is approved for the degree of Master of Technology

Dr. Basant Kumar Patel
Head of the Department
Department of Biotechnology
IIT Hyderabad

Dr.Anindya Roy
Department of Biotechnology
IIT Hyderabad

Dr.N. K. Raghavendra
(Adviser)
Department of Biotechnology
IIT Hyderabad

Dr.G. Satyanarayana
Department of Chemistry
IIT Hyderabad

CONTENTS

1. ABBREVIATIONS

LEDGF	Lens epithelium derived growth factor
HIV-1	Human immunovirus-1
CCD	Catalytic core domain
IBD	Integrase binding domain
AIDS	Acquired Immune Deficiency Syndrome
CD4	Cluster domain
SiRNA	small interfering RNA
IniI	Integrase interactor 1
BAF	barrier-to-auto integration-factor
HMGA1	High Mobility Group chromosomal protein A1
PC4	positive transcription co-factor 4
NLS	Nuclear localization sequence
NTD	N-terminal domain
CTD	C-terminal domain
SSC	Strand synaptic complex
BME	Beta mercaptoethanol
HECT	Homologous to E6AP C- terminus
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISG15	Interferon Stimulated Gene
MMS2	Methyl Methane Sulfonate
PCR	Polymerase Chain Reaction
PMSF	Phenylmethylsulfonyl fluoride
PRR	Post Replication Repair
RING	Really Interesting New Gene

SUMO	Small Ubiquitin Related Modifier.
SWI/SNF	Switch/Sucrose NonFermentable
TRAF6	Tumor Necrosis Receptor Associated Factor
UEV1A	Ubiquitin Enzyme Variant 1

2. ABSTRACT

Part I

Ubc13, an ubiquitin-conjugating enzyme (Ubc), requires the presence of an Ubc variant (Uev) for polyubiquitination. MMS2 (methyl methane sulfonate) and Uev1A has similar sequence homology. The difference in these two variants is 12 Amino acid residues. But these 12 amino acids residues are dictating two different pathways in which one pathway leads to innate immunity (Uev1a with UBC13) and other leads to DNA Repair pathway (MMS2 with UBC13).Generating mutations in MMS2 and analyzing the MMS2 mutants with UBC13, followed by ubiquitin assays will help us to understand which amino acid residue is important for the pathway specificity.

Part II

LEDGF/p75 is an important transcriptional cofactor which binds specifically to HIV-1 integrase. Two regions of the IN CCD dimer were identified to be involved in the interaction with LEDGF/p75, one among them which centers around on residues Trp131 and Trp132 while the second extends from Ile161 up to Glu170. The side chain of LEDGF residue Ile-365 projects into a hydrophobic pocket formed by IN B-chain residues Leu-102, Ala-128, Ala- 129, and Trp-132 and A-chain residues Thr-174 and Met-178. The side chain of HIV-1 IN Trp-131 is involved in hydrophobic contacts with Phe-406 and Val-408 in LEDGF in the form of strong π-π interactions. Trp-131 was the only CCD residue that underwent significant rearrangement upon formation of the IN-LEDGF/p75 complex. And it is also having interactions which consist of a hydrogen bridge between Trp131 and

the backbone Carbonyl of Arg405 in the IBD protein. So mutations in T131 residue by removing its Indole ring might create some conformational changes and helps in reducing the interactions with LEDGF/p75.

3. INTRODUCTION

Part I

Ubiquitin, a 76 Amino acid residue protein is primarily attached to Lysine residues in target proteins through an enzymatic cascade reactions catalyzed by E1, E2, and E3 enzymes.

Post-translational modification of proteins is one of major mechanisms employed by eukaryotic cells to expand the functional diversity of the proteome. The covalent modification of proteins with Ubiquitin is highly conserved in eukaryotes. The human genome encodes over 300 enzymes dedicated to Ubiquitin conjugation and approximately 100 Ubiquitin proteases, termed as Deubiquitinating enzymes, which remove Ubiquitin from modified proteins (Nijman et al., 2005; Bengtson et al., 2008). Ubiquitin conjugation is necessary for fundamental cellular processes, including cell cycle progression, transcriptional regulation, DNA repair, apoptosis, and protein trafficking, endocytosis and signal transduction. Thus represents an important form of post-translational modification in the cell (Pick art 2001; Felberbaum et al., 2006; Karin et al., 2008)

3.1 Ubiquitin Family

Ubiquitin is the member of small protein modifiers. There are two possibilities for ubiquitin to bind proteins. One way of binding is with C-terminal of Gly76 is attached to the € amino group of a lysine by isopeptide linkage with target proteins. The other way of ubiquitin binding is binding to Ubiquitin itself. Because it contains seven lysine residues, which function as acceptor sites for building polyubiquitination chains. Proteins can therefore be modified at a single lysine with one Ubiquitin molecule (monoubiquitination), at several lysine residues with Ubiquitin monomers (multi-monoubiquitination), or through chains of lysine-

linked polyUbiquitin (polyubiquitination). The modification of proteins with a single Ubiquitin molecule, can serve as a signal for endocytosis (Hicke 2001)

3.2 Ubiquitin-like modifiers

A family of proteins termed Ubiquitin-like modifiers shares structural similarity with Ubiquitin. Total nine Ubiquitins including ISG15 and SUMO; have demonstrated capability of covalently modifying protein targets (Hochstrasser 2009). Conjugation of most Ubiquitin-like modifiers follows a similar mechanistic framework to Ubiquitin and requires catalysis by dedicated E1, E2, and E3 enzymes.

3.3 Ubiquitin ligases (E3)

E3 Ubiquitin ligases catalyze the final step of Ubiquitin transfer to substrate. With over 300 members encoded in the human genome, E3 ligases comprise the largest class of Ubiquitin conjugation enzymes (Bengtson et al., 2008). They are divided into three classes based on their mechanism of function (1) RING and RING-like (2) HECT (3) neither RING nor HECT

3.3.1 RING E3 ligases

RING domains contain a cross-brace structure of eight conserved cysteine or histidine residues coordinating two zinc ions (Wing et al., 1999; Jensen et al., 1999) Variations in the number of zinc ions distinguish related domains. The U-box domain lacks zinc ions. Charged and polar residues, instead of the cysteine and histidine residues, forming salt bridges and hydrogen bonds that stabilize a similar cross-brace structure. The RING-H2 domain contains three zinc-binding sites. The RING domain of the E3 ligase recruits E2~Ubiquitin thioester complex. The region of interaction between RING E3s and E2 enzymes were elucidated through

structural studies of RING-E2 complexes (Dye & Schulman 2007; Deshaies & Joazeiro 2009)

The role of RING E3 during catalysis is to bring the E2~Ubiquitin thioester in proximity of a lysine residue for modification. Evidence indicates that merely binding of E2 and RING E3s is not sufficient for catalysis (Deshaies and Joazeiro 2009) and it is possible that the RING domain induces a conformational change in E2~Ubiquitin to promote Ubiquitin discharge. A conserved Asn residue located near the E2 enzyme active site aids in RING-dependent or E3-independent transfer of Ubiquitin to a substrate lysine. This Asn residue functions to stabilize the oxyanion intermediate that forms during nucleophilic attack of the substrate lysine residue on the E2~Ubiquitin thioester (Hanlon et al., 2003). RING E3s may function to allosterically activate the E2 by promoting a conformation change that positions the Asn residue in the oxyanion hole to stabilize the charged intermediate (Hanlon et al., 2003; Yu et al., 2005)

3.3.2 HECT E3 ligases

The HECT (*h*omologous to *E*6AP *C*-*t*erminus) domain is a 350 residue C-terminal domain catalytic defined member of E6AP family (Scheffner et al., 1994) .In contrast to the RING domain E3s, the human genome encodes 28 HECT E3s (Scheffner et al., 1995).This class of enzymes is differed from RING E3s by its catalytic mechanism. HECT E3s form an E3~Ubiquitin intermediate on a conserved catalytic cysteine contained within the HECT domain prior to catalyzing isopeptide bond formation.

The catalytic and substrate-binding functions of HECT E3s are typically separated in different domains, with substrate binding mediated by N-terminal protein-protein interaction modules.

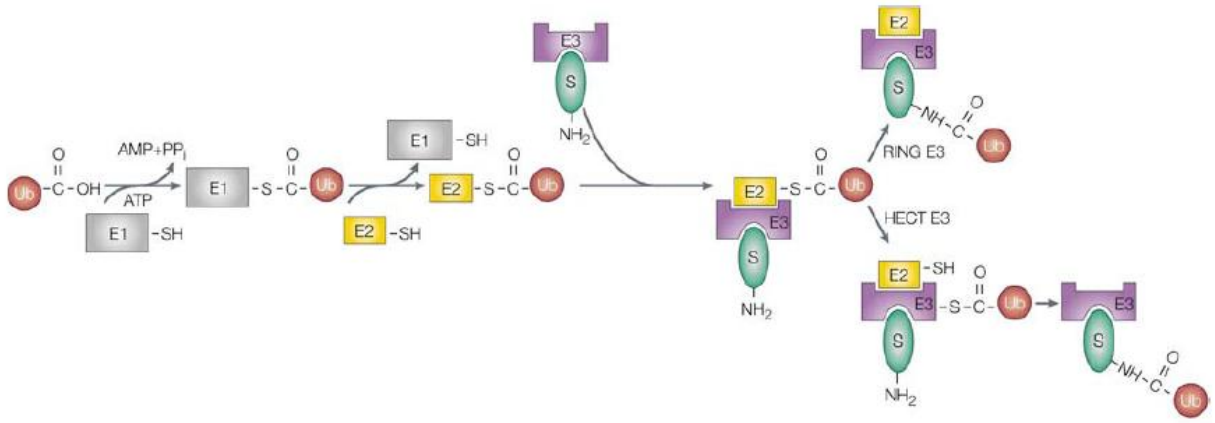

Figure 3.3

The ubiquitylation pathway: E1 enzyme activates Ubiquitin by adenylating Ub's C-terminal in an ATP-Dependent manner and forms E1 Ub~thioester complex. E1 Ub~thioester complex was associated with E2 and transfers Ub to the E2 catalytic cystiene The RING E3 brings E2 Ub~thioester complex in close proximity to the substrate and HECT E3 forms an E3~Ubiquitin intermediate on a catalytic cysteine.(Adapted from Fang & .Weissman 2004)

3.4 Uev1A

UEV1a encodes an ubiquitin (Ub) conjugating enzyme (Ubc) variant (Uev). Uev1 is a member of the protein family that serves as a co-factor for Ubc13-catalyzed Lys63-linked poly-Ub chains (McKenna et al., 2001; Moraes et al., 2001; Pastushok et al., 2003)

Tumor necrosis factor receptor associated factor 6 (TRAF6) is an important signaling molecule involved in a diverse array of physiological processes. It has been proposed that TRAF6, a RING finger-containing protein, acts as an ubiquitin ligase (E3) and a target for Lys-63 linked polyubiquitination mediated by Ubc13–Uev, an ubiquitin conjugating (E2) complex are required for the interaction. The TRAF6 RING finger domain mediates physical interaction with Ubc13 (Xiao et al., 2004)

3.5 MMS2

MMS2 encodes a 145-amino acid, 15.2-kDa protein with significant sequence homology to a conserved family of Ubiquitin-conjugating proteins. MMS2 interacts with RAD6 to direct the RAD6–RAD18 complex results in PRR and this interaction requires the amino terminal of RAD6. It is believed to be targeted to the sites of damage by the DNA-binding RING finger protein RAD18 (Von Borstel 1978). RAD6 system is mediated by another chromatin-associated RING finger protein, the SWI/SNF homolog RAD5 (Williamson & Fogel 1985)

```
gi|40806164|ref|NP_068823.2|    MPGEVQASYLKSQSKLSDEGRLEPRKFHCKGVKVPRNFRLLEELEEGQKG 50
gi|2947301|gb|AAC05381.1|       ----------MAVS----------TGVKVPRNFRLLEELEEGQKG 25

gi|40806164|ref|NP_068823.2|    VGDGTVSWGLEDDEDMTLTRWTGMIIGPPRTIYENRIYSLKIECGPKYPE 100
gi|2947301|gb|AAC05381.1|       VGDGTVSWGLEDDEDMTLTRWTGMIIGPPRTNYENRIYSLKVECGPKYPE 75

gi|40806164|ref|NP_068823.2|    APPFVRFVTKINMNGVNSSNGVVDPRAISVLAKWQNSYSIKVVLQELRRL 150
gi|2947301|gb|AAC05381.1|       APPSVRFVTKINMNGINNSSGMVDARSIPVLAKWQNSYSIKVVLQELRRL 125

gi|40806164|ref|NP_068823.2|    MMSKENMKLPQPPEGQCYSN 170
gi|2947301|gb|AAC05381.1|       MMSKENMKLPQPPEGQTYNN 145
```

Figure 3.5.Sequence alignment of human UEV1a and MMS2 are shown. Important non identical residues in both variants are labeled with colors.(MMS2 and Uev1a sequence adapted from uniprot and multiple alignment sequence was done by using ClustalW tool)

Part II

INTRODUCTION

1st time AIDS was reported in 1981 and the discovery of its causing Agent is human immunodeficiency virus (HIV). In 1983, this virus has killed more than 2.9 million people. About 4.3 million were infected in 2006. At the end of 2006 approximately 39.5 million people are living with the disease. The emergence of HIV strains resistant to the presently available anti-HIV drugs is an important factor in therapy failure. Therefore it is extremely important that to develop new compounds which helps to target HIV replication.

HIV was discovered in 1983 as the causative agent of Acquired Immune Deficiency Syndrome, AIDS (Gallo et al., 1984; Barré-sinoussi et al., 1983; Levy et al., 1984) and is a member of the Lentiviruses, which belongs to a subfamily of the retroviruses. Retroviruses are enveloped RNA viruses and are characterized by a reverse transcription step in their replication cycle. Unlike other retroviruses, which require cell division for productive infection, lentiviruses can stably integrate their proviral DNA into the genome of dividing as well as non-dividing cells. Based on genome organization and phylogenetic relationship with other lent viruses, HIV can be divided into 2 types: HIV-1 and HIV-2. Worldwide infection with HIV-1 is more common.

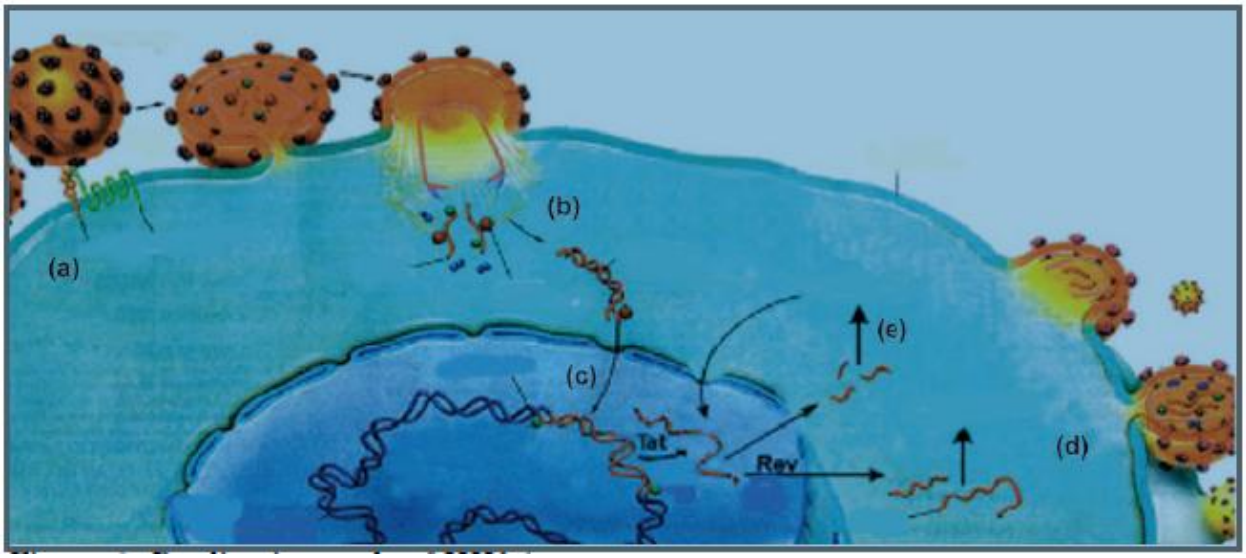

Figure 1: Replication cycle of HIV-1

HIV-1 replication is of two steps.

1st step nonspecific attachment of the virus to cell surface and enter into the cells (step a) the reverse transcription (step b) and the integration of the proviral DNA (step c).

2nd step includes all events from transcription of the integrated DNA (step e) to budding (step d) and maturation and results in the formation of infectious viral particles. (Adapted from Meadows et at., 2009)

3.1 Cellular Co-Receptors

One of the first cellular co-factors used in the replication cycle of HIV are the cellular receptor CD4 and co-receptors CXCR4 or CCR5. The first efforts at targeting cellular co-factors for antiretroviral therapy have focused on these receptors. A 32 bp deletion in the gene coding for the CCR5 co-receptor of HIV-1 results in a receptor that is severely truncated and cannot be detected at the cell surface. Individuals, homozygous for this deletion, are resistant to infection with an HIV-1 CCR5 strain (Benkirane, et al., 1997)

3.2 Cellular co-factors of Integrase

Besides these, cellular co-factors for HIV-1 integration have also been identified. Among them (1) barrier-to-auto integration-factor (BAF) (2) the High Mobility Group chromosomal protein A1 (HMGA1) (3) Integrase interactor 1 (IniI) and (4)Lens epithelium-derived growth factor (LEDGF/p75).

3.2.1 Lens epithelium-derived growth factor (LEDGF/p75)

LEDGF/p75 was identified as a HIV-1 Integrase-interacting protein by co-immunoprecipitation of nuclear extracts of cells stably over expressing HIV-1 integrase from a synthetic gene (Cherepanov, et al., 2003) during the course of purifying the general transcriptional co-activator PC4 a ~75 kDa polypeptide was co-purified. This polypeptide was identified as p75 or the positive transcription co-factor 4 (PC4)-interacting protein. LEDGF/p75, 530 AA long, is a member of the hepatoma-derived growth factor (HDGF) family and contains, a conserved N-terminal PWWP-domain (Amino acid from 1-91) thought to be involved in protein-protein interactions (Stec, et al., 2000) and chromatin binding (Gee, et al., 2004). Besides the PWWP domain, LEDGF/p75 also contains a nuclear localization signal (NLS) (Amino acid from 148-156). A conserved domain of LEDGF/p75 that is necessary and sufficient for the binding of LEDGF/p75 to the HIV-1 integrase was mapped to just 83 amino acids, spanning the residues (347 – 429): the integrase binding domain or IBD. This IBD is essential but not sufficient to enhance the strand transfer activity of HIV-1 IN (Cherepanov, et al., 2004) LEDGF/p75 proved to be essential for nuclear localization and chromosomal association of HIV-1 IN since transient and stable RNA-interference (RNAi)-mediated knockdown of endogenous LEDGF/p75expression abolished

nuclear/chromosomal localization of integrase [Maertens, et al., 2003; Llano, et al., 2004). LEDGF/p75 is probably present in the reintegration complex (Llano, et al., 2004).

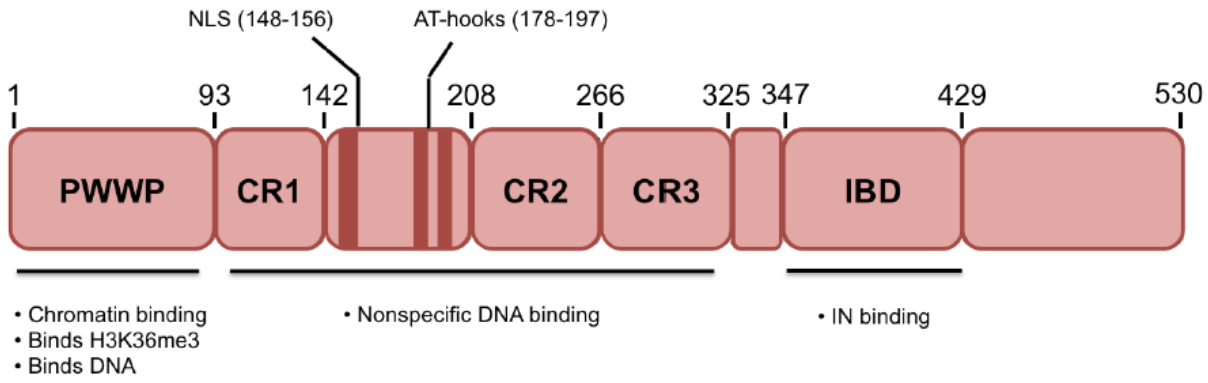

Figure 3.2.1: LEDGF/p75 structural organization.

The N-terminal portion of LEDGF/p75 is responsible for binding chromatin. The PWWP domain binds to methylated histones as well as DNA nonspecifically. The AT-hooks CR1, CR2 and CR3 each contribute to nonspecific DNA binding. The C-terminal end harbors an integrase binding domain (IBD) which binds IN in a hydrophobic pocket at the CCD-CCD dimer interface

3.2.2 Importance of IBD residues in IN-LEDGF/p75 interaction

.

The C-terminal of LEDGF/p75, an 83-amino acid domain interacts with lentiviral INs, including HIV-1 IN, and is thus named the integrase binding domain (IBD) (Gee, et al., 2004). The LEDGF/IBD is comprised of 5 alpha-helices, of which 4 (α1, α2, α4 and α5) form a right-handed helical bundle which are connected by α3. A cluster of hydrophobic residues at the top of the LEDGF/IBD structure formed by the loops between α1-α2 and α4-α5 have been shown to interact with IN. Mutagenesis these amino acids, Ile366, Asp366 and Phe406 abrogated LEDGF/IBD-IN interactions while Val408 severely reduced binding. This confirmed that the V-shaped loop is critical for protein-protein interaction (Cherepanov, et al., 2004) Biochemical studies have also demonstrated that the IN

CCD in necessary and sufficient to bind LEDGF/p75. Additionally, the crystal structure of the LEDGF/IBD in complex with the HIV-1 IN CCD dimer revealed that the LEDGF/IBD binds at the CCD-CCD dimer interface. In the crystal structure the Ile366 of LEDGF/IBD forms hydrophobic interactions with HIV-1 IN residues Leu102, Ala128, Ala129, Try132, Thr174 and Met178. Additionally, LEDGF/IBD residues Phe406 and Val408 have hydrophobic interactions with HIV-1 IN Try131 and LEDGF/IBD Asp366 forms hydrogen bonds to the backbone amides of HIV-1 IN Glu170 and His171. (De rijck, et al., 1997) Although the IN CCD is necessary and sufficient to bind LEDGF/IBD, the IN NTD contributes to the high-affinity binding of IN to LEDGF/p75 (Cherepanov, et al., 2004).

3.3 HIV-1 Integrase protein

Integrase is a 288-amino acid protein (32 kDa) encoded by the end of the pol gene. It is produced as part of the Gag-Pol polypeptide precursor, from which it is released by viral protease-mediated cleavage (Olivier delelis, et al., 2008) It is comprised of three domains: the N-terminal domain (NTD) which contains the Zn-binding motif (HH-CC type), the catalytic core domain (CCD) which contains the DDE catalytic triad that coordinates essential Mg2+ ions, and the C-terminal domain (CTD) which contains an SH3-like fold (Andarke et al., 1995). Each of these domains contributes to protein multimerization(Wang jy, et al., 2001; Chen jc, et al., 2000; Jenkins et al.,1999; Andarke et al., 1995).Integrase is an important therapeutic target as its function is essential for viral replication. A tetramer of IN assembles on the viral DNA ends to form then stable synaptic complex (SSC) and catalyzes two reactions necessary for the integration of linear viral DNA into the host chromatin (Brown et al,.) IN initially removes a GT dinucleotide from the 3′-

terminus of each viral DNA end (3′-processing), and then integrates recessed viral DNA ends into the target DNA in a staggered fashion through concerted transesterification reactions (DNA strand transfer). IN is comprised of three domains: the N-terminal domain (NTD) which contains the Zn-binding motif (HH-CC type), the catalytic core domain (CCD) which contains the DDE catalytic triad that coordinates essential Mg2+ ions, and the C-terminal domain (CTD) which contains an SH3-like fold (Engelman et al., 2011). Each of these domains contributes to protein multimerization (Wang jy, et al., 2001; Chen jc, et al., 2000; Jenkins et al., 1999; Andarke et al., 1995)

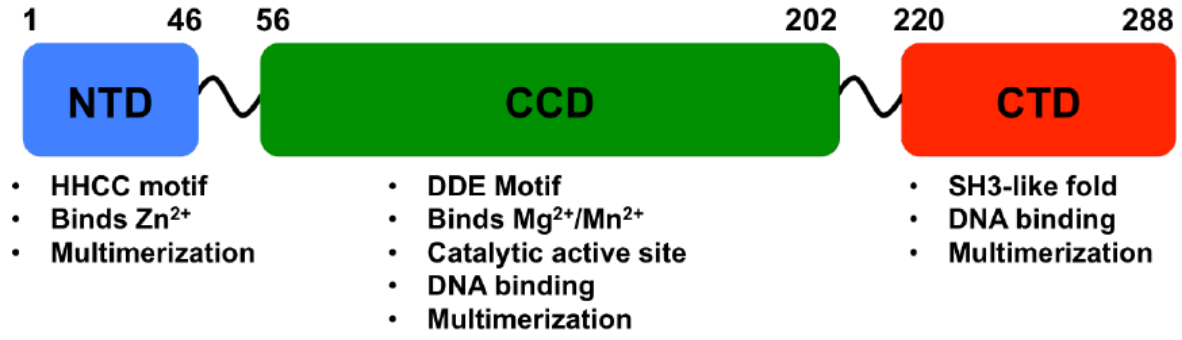

Figure 3.3: IN structural organization

IN is comprised of 3 domains, an N-terminal domain (NTD), a catalytic core domain C-terminal domain (CCD) and a C-terminal domain (CTD). The NTD contains an HHCC motif for the coordination of a Zn2+ ion that is important for the multimerization of the full-length protein. The CCD contains the DDE active site and contributes to DNA binding. The CTD contributes to DNA binding as well as the multimerization of the full-length protein.

3.4 INTERACTION OF T131 WITH PHE406, VAL408 AND ARG405 OF LEDGF/p75

Structure obtained from PDB and interactions were analyzed by using PyMol.

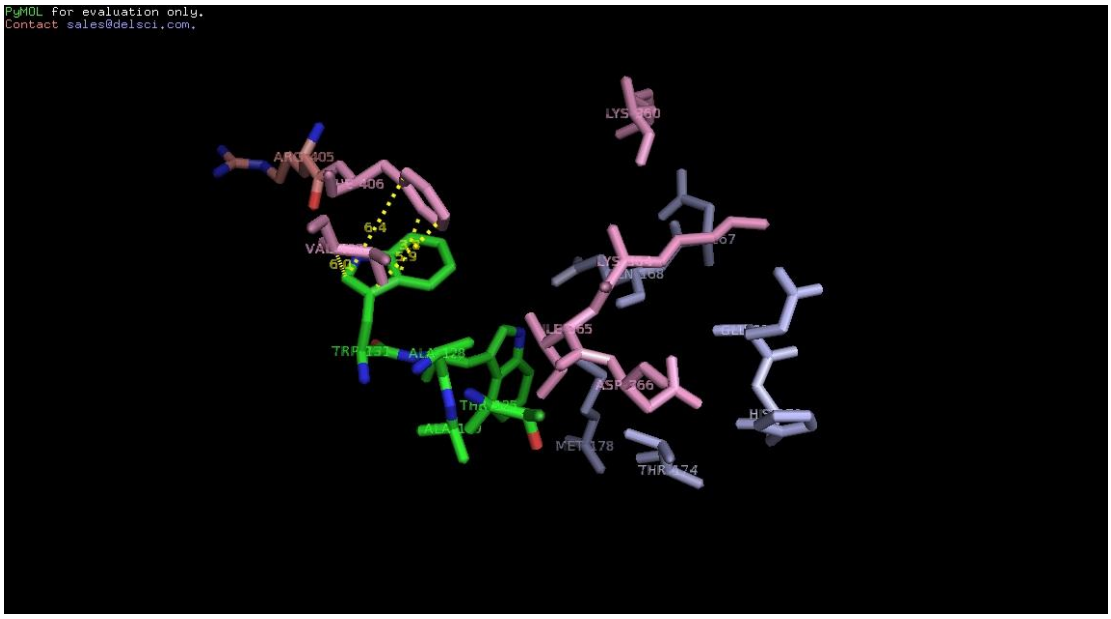

Figure 3.4A: Trp-131 is involved in hydrophobic contacts with Phe-406 and Val-408 in the form of strong π-π interactions.

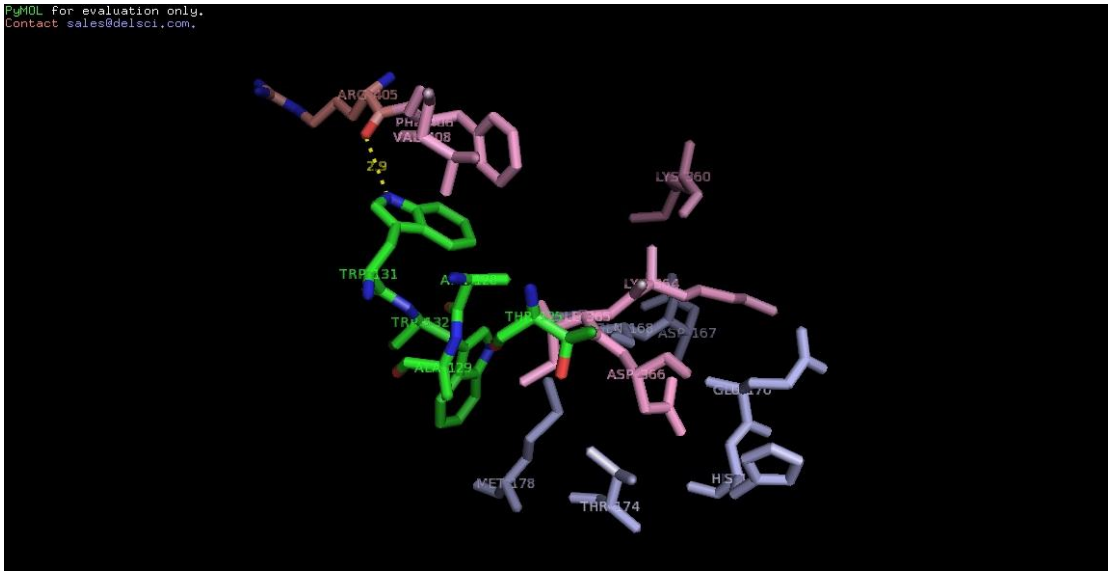

Figure3.4B: Hydrogen bond between Trp131 and the backbone Carbonyl of Arg405

4. OBJECTIVES

Part I

1) Mutation of N57I residue of MMS2 by overlap PCR
2) To clone the mutant gene and purify the mutant protein.

Part II

1) Mutation of T131F residue of IN by using Pymol
2) Dock both wt LEDGF IBD and mutant HIV-1 IN CCD by using Cluspro tool
3) Analyzing the interactions with LEDGF/p75 by using Discovery studio 4.0 Software

5. RATIONALE

Part I

MMS2 and Uev1A have similar sequence homology. The only change is 12 amino acid residues in both the proteins. But these 12 amino acids change, dictating two different pathways in which one pathway leads to innate immunity (Uev1a with UBC13)and other goes to DNA Repair pathway(MMS2 with UBC13). By creating mutations in one of the E2 variants (MMS2) will help us to understand which amino acid residue is important for specificity of signaling pathway when it interacts with UBC13.

Part II

The side chain of HIV-1 IN Trp-131 is involved in hydrophobic contacts with Phe-406 and Val-408 in LEDGF in the form of strong π-π interactions. And it is also having interactions which consist of a hydrogen bridge between Trp131 and the backbone Carbonyl of Arg405 in the IBD protein. So mutations in T131 residue by

removing its indole ring might create some conformational changes and helps in reducing the interactions with LEDGF/p75.

6. METHODS

Part I

6.1 Cloning of wild type MMS2 and Mutant MMS2

The plasmids carrying ORF coding for Uev1a, MMS2 and UBC13 proteins in the study have been purchased from Origene technologies U.S.A. The MMS2 gene was PCR amplified and cloned into pRSETA vector having in frame HIS–tag at the N-terminus with a protease cleavage site. Followed by screening, PCR amplification of desired insert with wild type and mutant primers of both forward and backward(Wild type MMS2 NR-7 5` CGT GGA TCC ATG GCG GTC TCC ACA GGA G 3` NR-8 5` CCG CTC GAG TTA ATT GTT GTA TGT TTG TCC TTC TG 3` N57I mutant MMS2 primers NR-22 GGGCCACCTAGGACAATTTATGAAAACAG 3`NR-23 5` CTG TTT TCA TAA ATT GTC CTA GGT GGC CC 3` which creates AvrII site) and Insert release strategy by using Restriction Enzymes(Bam-XhoI) was performed.

6.2 Recombinant protein expression and purification

6.2.1 Protein expression

The positive clone was transformed into BL21 (DE3) pLys cells in LB Agar media, with ampicillin (50µg/ul) and chloramphenicol (30µg/µl) and incubated at 37^0C at 250rpm for 13-14hrs.Picked 3 colonies from BL21 (DE3) pLysS transformed plate and inoculated into 3ml culture media by adding 50µg/ul of

ampicillin, incubated at 37^0C at 250rpm for overnight which serve as a pre-inoculum. Took 5% preinoculum into fresh 3ml culture tubes with 50µg/ul of ampicillin and grow the bacterial culture till it reaches 0.5-0.6 OD at 600nm.Induce with 1mM IPTG and incubated at 37^0C at 250rpm for 4hrs. Loaded the samples in 12% SDS PolyAcrylamideGel with reference to protein ladder

6.2.2 Protein purification

Added 1% of pre-inoculum into 125ml LB media by adding 50µg/ul of ampicillin and grown the bacterial culture till OD reaches 0.5-0.6. Induced with 1mM IPTG and incubated at 37ºC at 250rpm. Mutant Protein purification was performed by using Ni NTA column, in which cells were lysed in lysis buffer(50mMTris-Cl (pH8), 150mM NaCl, 20mM imidazole,1mM BME, 0.5% Triton-X100,1mM PMSF) followed by sonication and centrifuged at 14.8k rpm for 10 min. Added cell lysate to Ni NTA column and collected Flow through followed by adding wash buffer (50mMTris-cl (pH8), 150mM NaCl, 20mM imidazole) and collected wash which then followed by elution buffer and collected eluted fractions(50mMTris-cl (pH8), 150mM NaCl, 400mM imidazole) loaded on in 12% SDS Poly Acryl amide Gel with reference to protein ladder.

Part II

6.1 Computational methods

Solution structure of IBD of LEDGF (1ZE9) and HIV Integrase protein (4DMN) was obtained from PDB. Structure was visualized and analyzed using PyMol. Initially mutation of T131F was done in Pymol by removing non bonded molecules and ligands.

Steps to mutate one amino acid on a protein structure
Open from the menu Wizard > Mutagenesis next Click on the residue for choosing the mutant followed by selecting a conformational state in the new mutagenesis panel menu .Click on the No Mutation button and select a new amino acid (e.g. PHE). Click Apply and click done when finished with the Mutagenesis Wizard.

6.1.1 Cleaning the crystal structure of protein before docking

Open the crystal structure of IBD of LEDGF (1ZE9) and T131F mutant of HIV Integrase protein (4DMN) in Discovery Studio and change the display style. Change the protein representation to flat ribbon and colour by Secondary type. Click on Structure in the toolbar and Crystal Cell in the dropdown and click on Remove Cell. Click on Macromolecules Tab, Protein Report and if any ligands or water molecules found, delete them. In the same tab, click on Prepare Protein followed by Clean Protein and Show Clean Report. Alternate confirmations of proteins will be reported.

This is followed by Prepare Protein Step. In the window that appears, change the pH and Ionic strength according to conditions and Click on Run. On job

completion, a file opens as protein name prep and reports about missing residues etc. Validate the structure on clicking X-ray (or Macromolecules) Tab and click on Validate Protein Structure followed by Perform Validations, Check Structure.

6.1.2 Docking with Cluspro2.0

Docking was performed by using Cluspro 2.0 version (http://cluspro.bu.edu/home.php). It is an online server which represents a web-based program for the computational docking of protein structures. Upload the coordinate files of two protein structures through ClusPro web interface, or enters the PDB codes. Both Wt and mutant protein structure were uploaded separately and ran the job.

Cluster is the group of docked models in which high no of models will be more favorable and stable. These models were opened in Discovery studio 4.0 and analyzed interactions.

6.1.3 Analyzing the interactions by using Discovery studio

Open the cluspro model in Discovery studio 4.0 and select all interactions and distances from Ligand interaction tab and analyze the interactions with desired residues.

7. RESULTS AND DISCUSSIONS:

Part I

7.1 Screening of wild type MMS2:

Digested PCR product amplified product of wild type MMS2 and digested pRSETA vector (Bam-XhoI) was incubated at 37^0C for 4hrs followed by Ligation with T4 DNA ligase and Transformed 100ul on DH5α plate. Picked colonies from DH5α plate and incubated at 37^0C for 13hrs followed by screening the colonies and DNA was run on 0.6% agarose for 90min at 50v. Lane 1 & 3 showed slower mobility shift when compared with pRSETA vector of on 2.9kb.

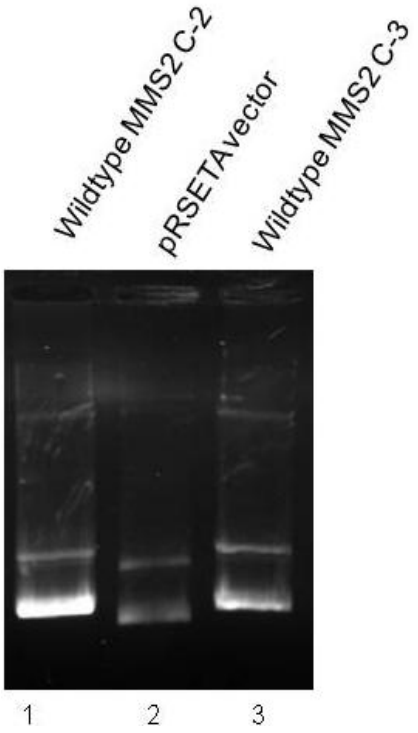

Figure 7.1: **Sreening of wild type MMS2.**Lane1 & 3 shows wild type MMS2 screened colonies with pRSETA vector backbone and MMS2 insert showing slower mobility; Lane2: shows only pRSETA vector.

7.2 Bam-XhoI Restriction Digestion for MMS2 insert release

The restriction enzymes for screening wild type MMS2 was used for screening strategy of insert release. Prepared master-mix (autoclaved water,10x BSA,Buffer-3,BamHI and XhoI) to that added screened wild type MMS2 and incubated at 37^{0}C for 8hrs.Uncut DNA and Digested DNA with reference to 1kb DNA Ladder was run on 0.6% agarose at 50v for 90min. Insert release was observed at 435 bp in Lane2 of Wild type MMS2 which was digested with Bam-XhoI.

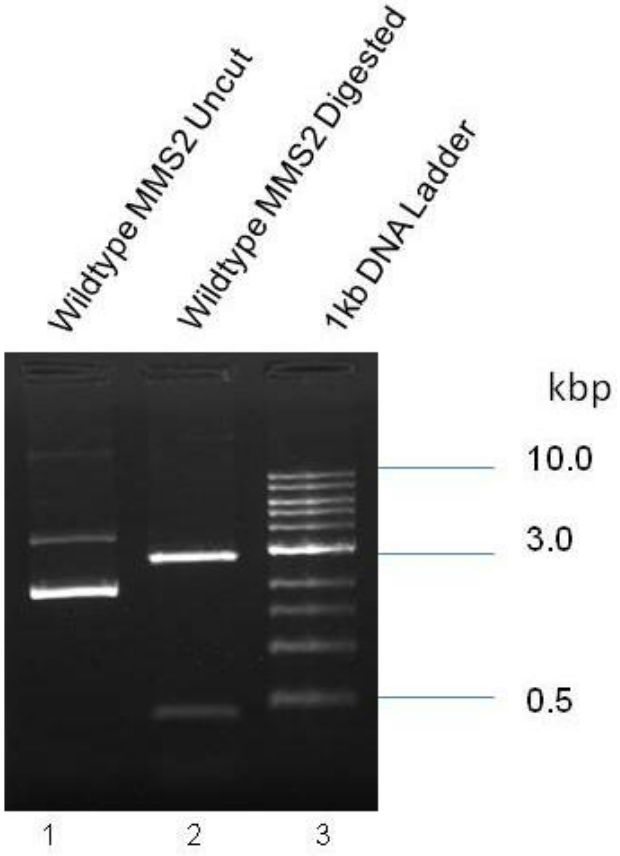

Figure7.2: **Bam-XhoI Restriction Digestion for insert release:**Lane1: shows wild type MMS2 Uncut DNA; Lane2: shows wild type MMS2 Digested DNA with Bam-XhoI; Lane3: shows 1kb DNA Ladder; Insert release was observed at 435 bp in Lane2 of Wild type MMS2

7.3 Screening of mutant MMS2 (N57I)

Digested MMS2 mutant (N57I) PCR amplified product of and digested pRSETA vector (Bam-XhoI) was incubated at 37^0C for 4hrs followed by Ligation with T4 DNA ligase and Transformed 100ul on DH5α plate. Picked 9 colonies from DH5α plate and incubated at 37^0C for 13hrs followed by screening the colonies and DNA was run on 0.6% agarose for 90min at 50v. Lane4 showed slower mobility shift when compared with pRSETA vector of on 2.9kb

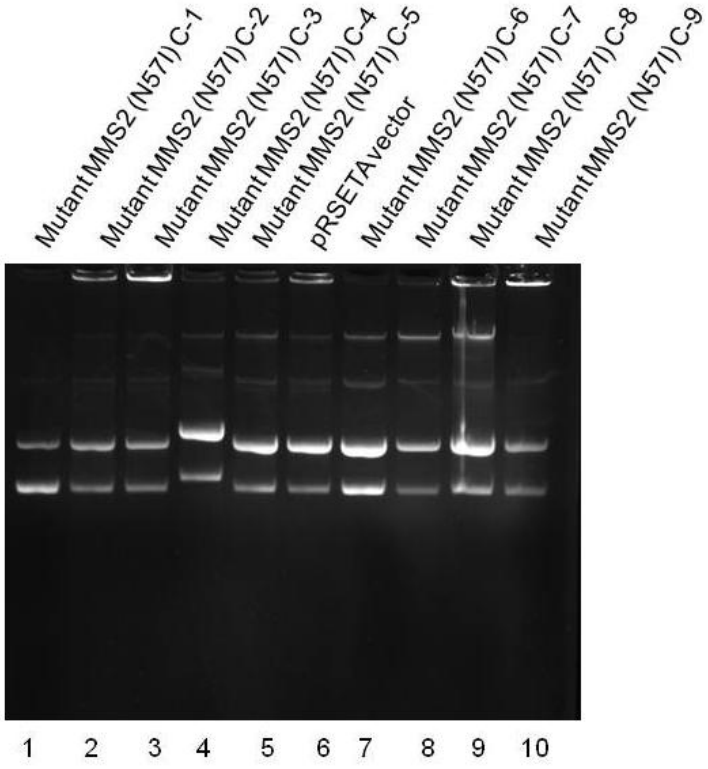

Figure 7.3: Screening of mutant MMS2 (N57I). Lane1-5 & 7-10 shows mutant MMS2 screened colonies; Lane4: shows mutant MMS2 insert (N57I) along with vector backbone showing slower mobility shift; Lane6: shows pRSETA vector (2.9kb);

7.4. Bam-XhoI Restriction Digestion for mutant MMS2 (N57I) insert release

The restriction enzymes for screening wild type MMS2 was used for screening strategy of insert release. Prepared master-mix (autoclaved water,10x BSA,Buffer-3,BamHI and XhoI) to that added screened MMS2 mutant N57I and incubated at 37^0C for 3hrs.Uncut DNA and Digested DNA with reference to 1kb DNA Ladder was run on 0.6% agarose at 50v for 90min. Insert release was observed at 435bp in Lane2 of mutant MMS2(N57I) which was digested with Bam-XhoI

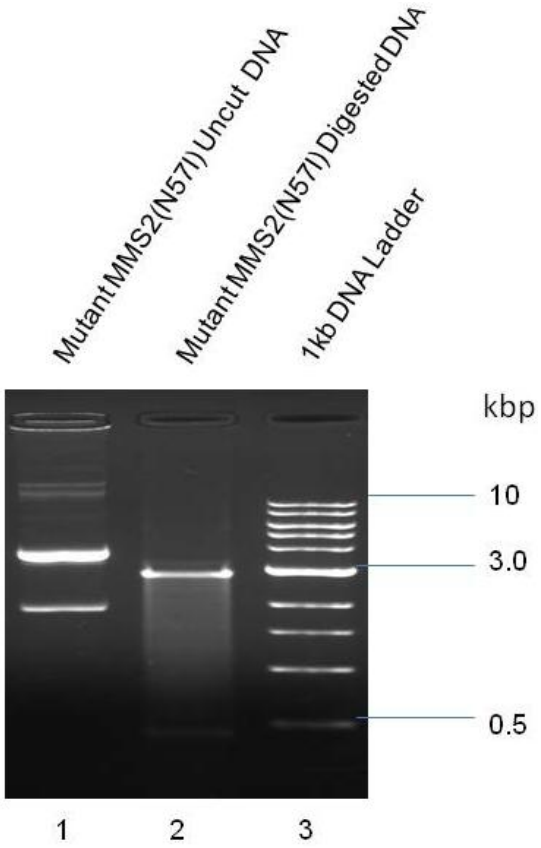

Figure 7.4: Bam-XhoI Restriction Digestion for mutant MMS2 (N57I) insert release

Lane1: shows mutant MMS2 (N57I) Uncut DNA; Lane2: shows mutant MMS2 (N57I) Digested DNA with Bam-XhoI; Lane3: shows 1kb DNA Ladder. Insert release was observed at 435bp in Lane2 of mutant MMS2 (N57I)

7.5 Mutant MMS2 N57I screening with AvrII digestion with respect to wild type MMS2

Mutant MMS2 N57I creates AvrII restriction site. Wild type MMS2 doesn't have AvrII restriction site. So to confirm mutant MMS2 N57I insert AvrII Digestion was done. Prepared master-mix (autoclaved water, 10x BSA, Buffer-4 and AvrII enzyme) to that added screened colonies of both wild type and mutant MMS2. Uncut and digested wild type MMS2 and Uncut and digested mutant MMS2 N57I along with 1kb DNA ladder was run on 0.6% agarose at 50v for 90min.Lane2 & 3 of wild type didn't showed any linearized band. But Lane 4 of mutant MMS2 which gained AvrII restriction site showed linearized band.

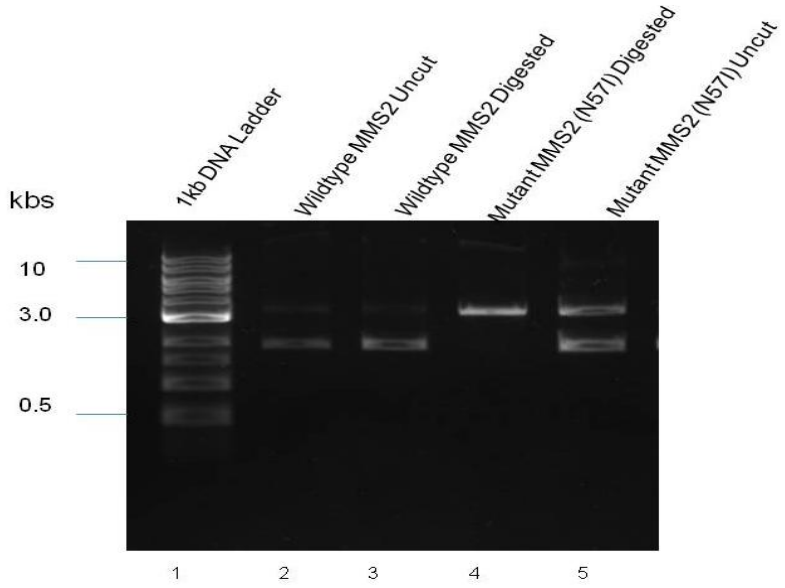

Figure 7.5: AvrII digestion for mutant MMS2 N57I insert release

Lane:1kb DNA ladder;Lane2 Uncut wild type MMS2; Lane3 Digested wild type MMS2;Lane4:Digested Mutant MMS2 N57I; Lane5:Uncut Mutant MMS2 N57I; Lane 4 of mutant MMS2 which gained AvrII restriction site showed linearized band.

7.6 Mutant MMS2 N57I protein purification

Positive clone of mutant MMS2 was transformed into BL21 (DE3) pLys Cells followed by picking colonies and incubated at 37^0C in shaker at 250rpm for 13hrs, grew the bacterial culture till it reaches to 0.5-0.6 OD at 600nm and induced with 1mM IPTG. Induced culture was pelleted and added lysis buffer to the pellet followed by sonication and centrifugation. Collected flow through from cell lysate through NI NTA column and also collected wash. Added elution buffer to the column and eluted the protein for 5 fractions. The N57I mutant MMS2 protein was run at 10% SDS-PAGE with 80v (till the protein samples goes out from the wells) for 30 min and 110v for 76 min. The N57I mutant protein band was observed in elution fraction 3 & 4 with (approx) 23kDa.

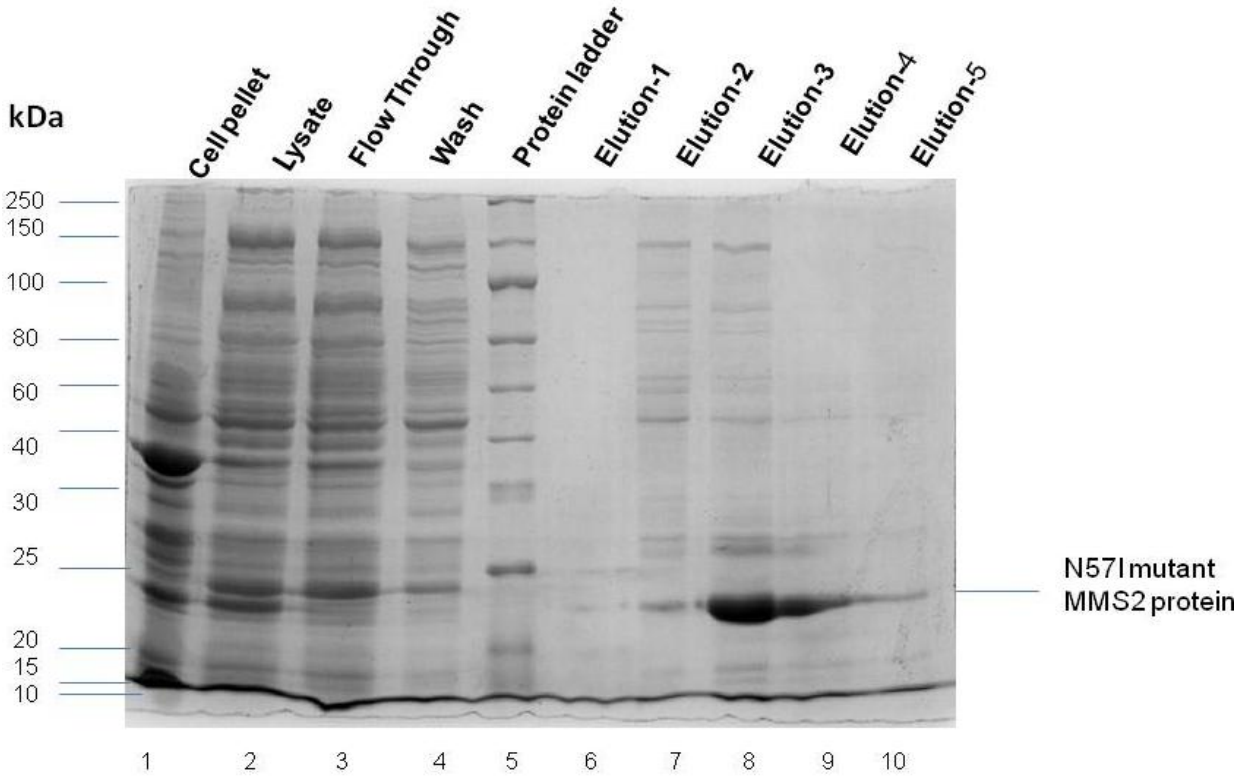

Figure 7.6: **Mutant MMS2 N57I protein purification:**Lane1: cell pellet was lysed in lysis buffer and followed by sonication and centrifuged at 14.8k rpm for 10 min; Lane2 was cell lysate; Lane3: was collected as flow through; Lane4: was collected as was; Lane5: 250kDa

Protein Ladder;Lane6-10 Elution fractions. Mutant protein was seen in elution fraction 3 & 4 with (approx) 23kDa

Part II

7.1 PyMol Analysis:

Solution structure obtained from PDB was analyzed using PyMol and Distance from T131 residue of HIV Integrase and Phe406, Val408 and Arg405 of LEDGF IBD was measured.

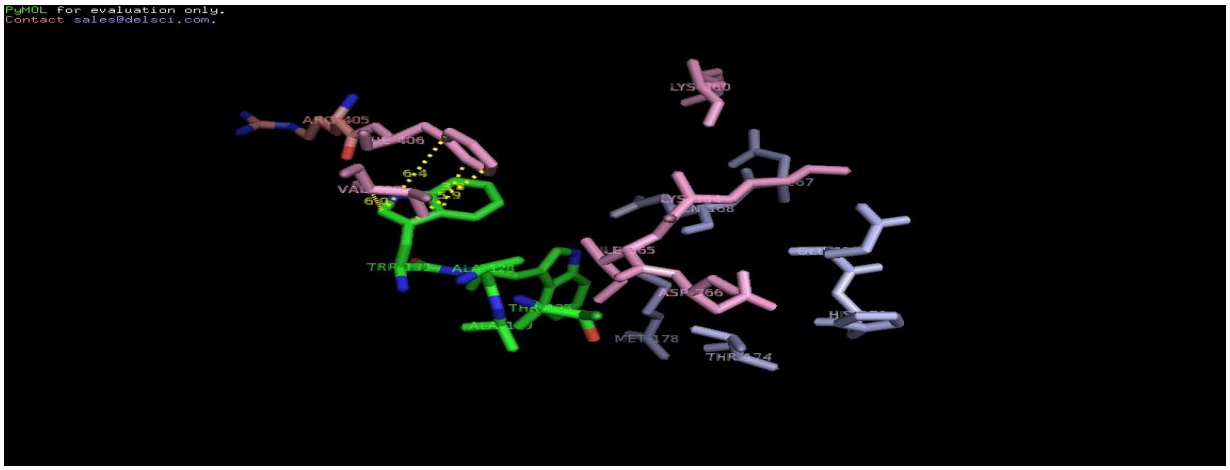

Figure 7.1A: Trp-131 (green) is involved in hydrophobic contacts with Phe-406 and Val-408(light pink) in the form of strong π-π interactions.

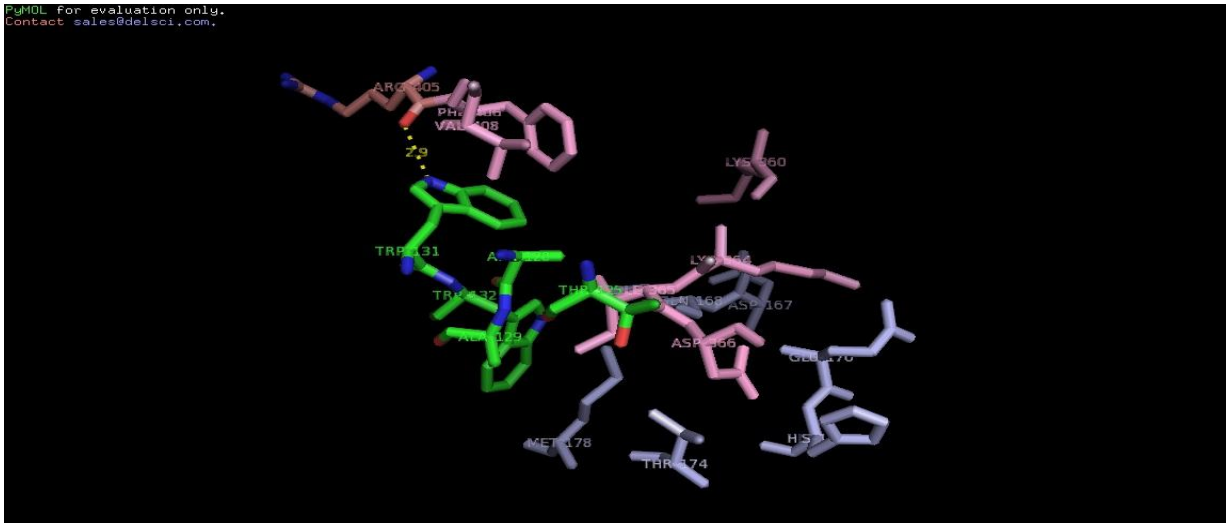

Figure7.1B: Hydrogen bond between Trp131 (green) and the backbone Carbonyl of Arg405 (pink)

After creating mutation from T131 to F131 in pymol the conformation of residue were changed. The changed conformation was observed by overlapping two structures on top of each other in pymol.

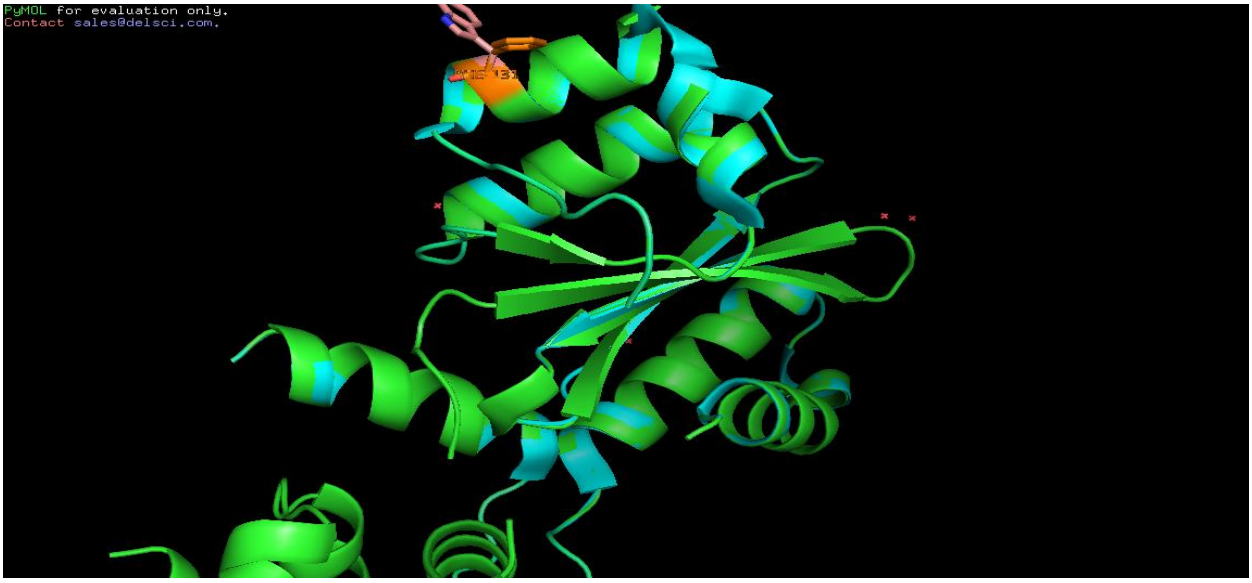

Figure7.1C: wt T131 (light pink) and mutant T131F (orange) was indicated.

7.2 Docking results

Total 9 models were generated with wt IBD and wt CCD which starts from 0-9 belongs 1 cluster. All 9 models posses interactions with T131 residue but Model 6th which posses least energy and having strong π-π interactions with wt IBD and wt CCD were observed in Discovery studio4.0.

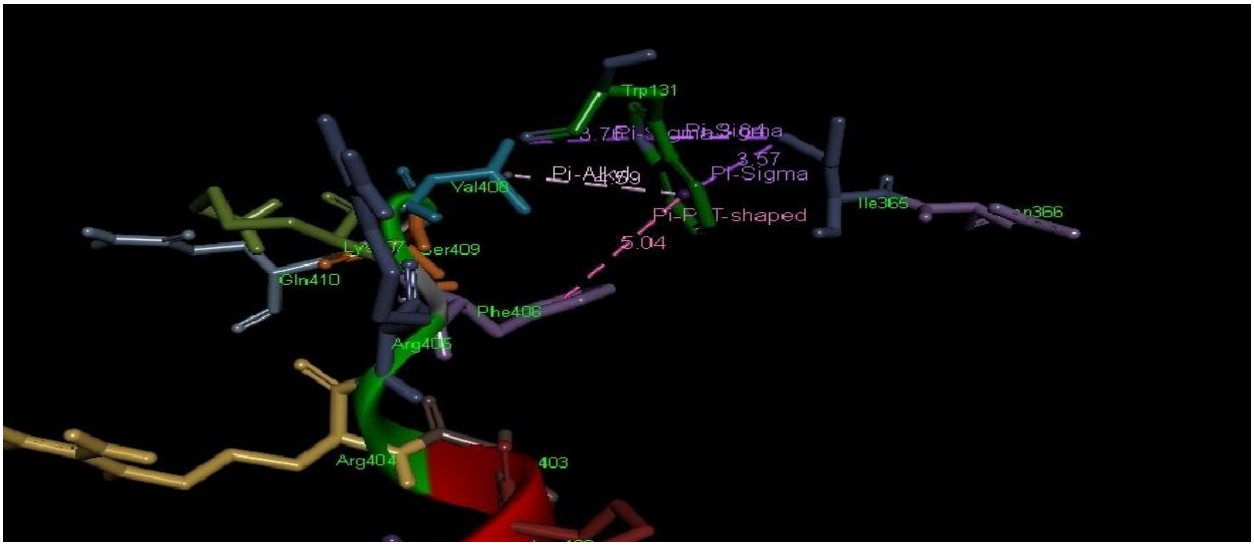

Figure7.2A: T131 (green) is interacting with Val408 (sky blue) and Phe406 (deep blue) in the form of π-π interactions

69 models were generated with wt IBD and mutant T131F CCD which starts from 0-25(max) which were separated in 4 clusters. Generation of mutation in HIV Integrase residue of T131 alters the conformation of protein structure and helps in reducing the interactions with Phe406, Val408 and Arg405.

From the result which obtained from cluspro, out of 69 models which are grouped into 4 clusters, only 3 models were showing interactions with LEDGF IBD.Those 3 interactions were also with other residues of LEDGF IBD not with Phe406, Val408 and Arg405.None of the models were showed interactions with Phe406, Val408 and Arg405.By mutating T131 residue the change of conformational state and helps in reducing the interaction with LEDGF IBD was observed in docked models as well as Discovery studio.

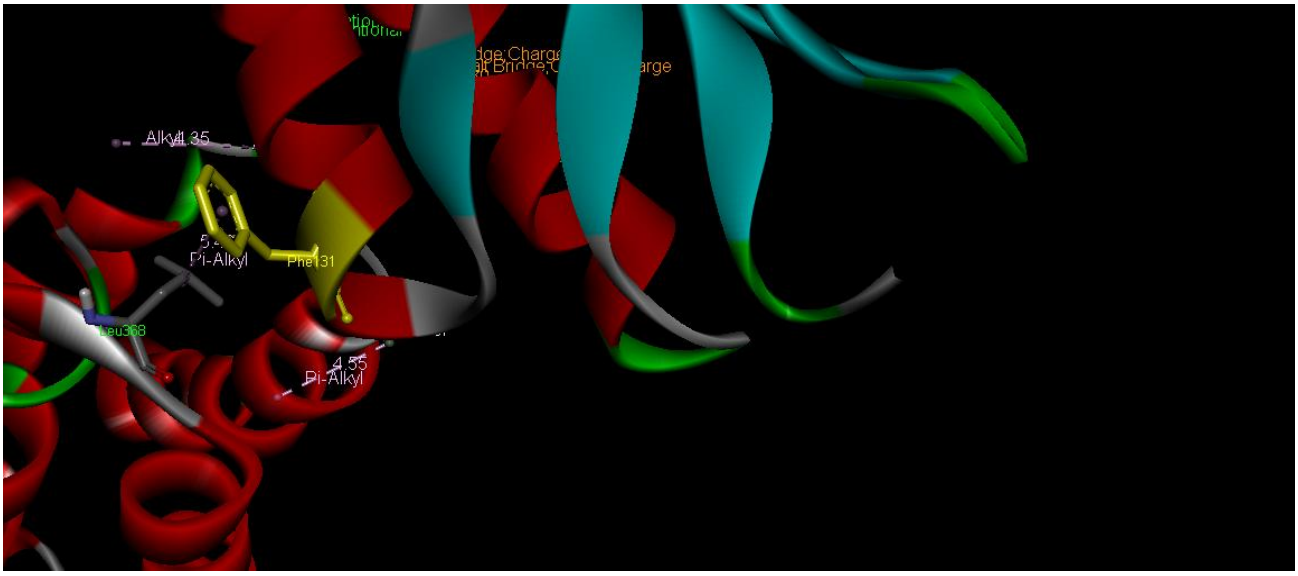

Figure7.2B: Phe131 (Yellow) is interacting with Leu368 (Ash) in the form of π-π interactions

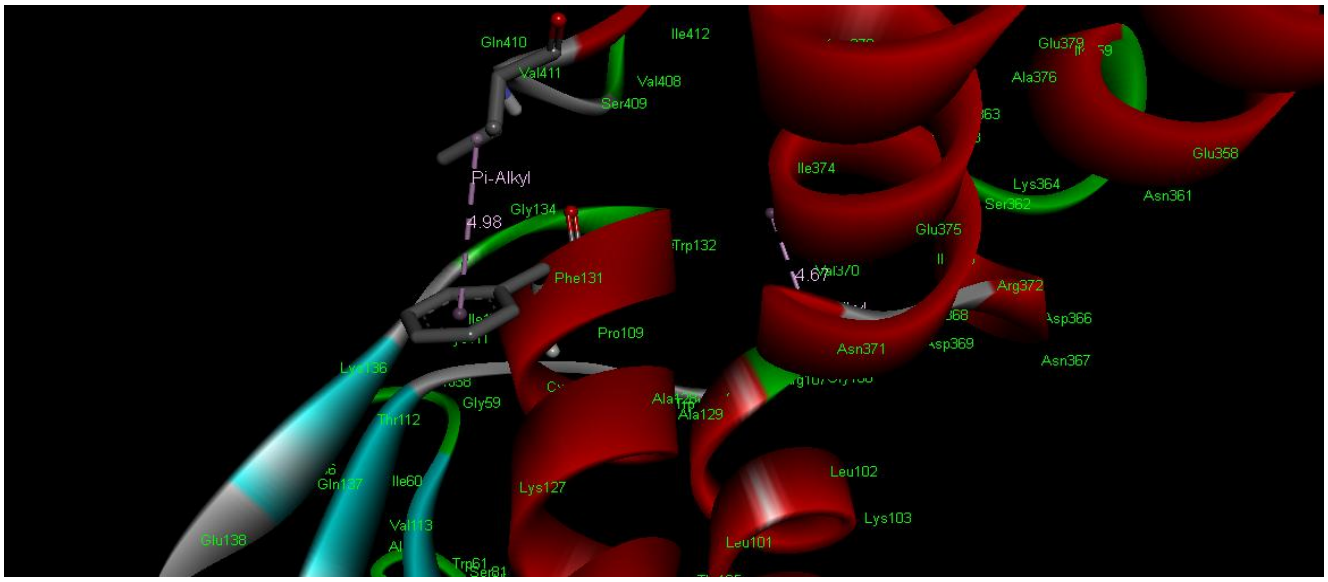

Figure7.2C: Phe131 (Ash) is interacting with Val411 (Ash) in the form of π-π interactions

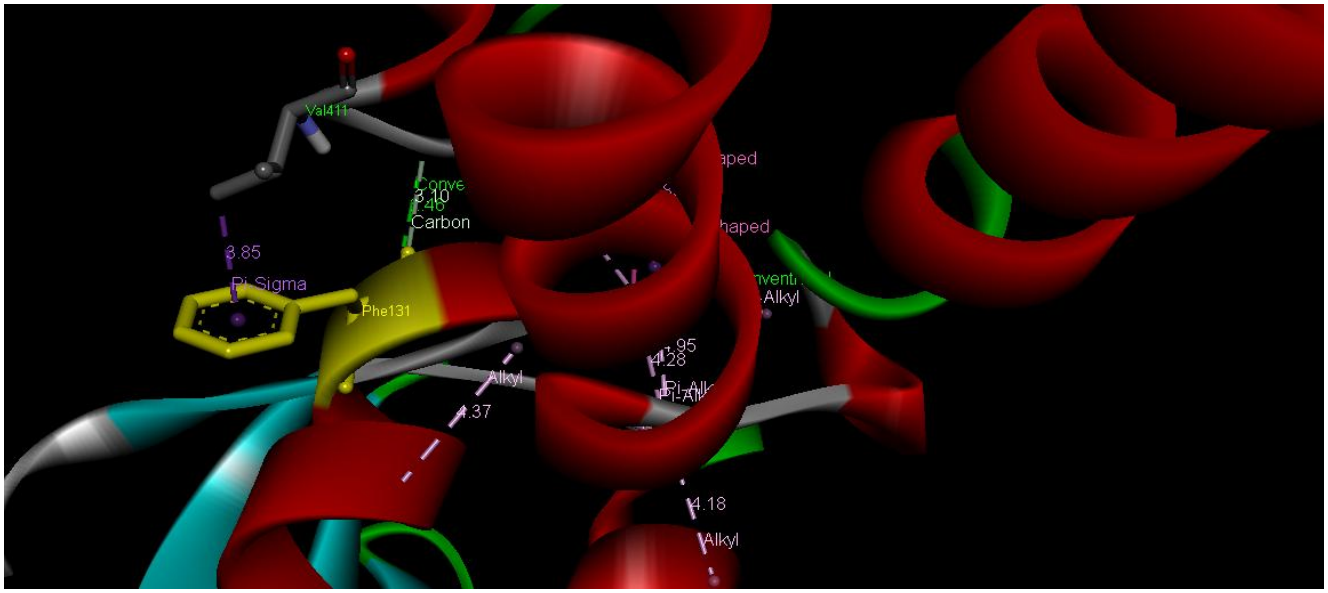

Figure7.2D: Phe131 (Yellow) is interacting with Val411 (Ash) in the form of π-π interactions

7.3 Energy calculations

Energy calculations were calculated based on given formula.

$$E=0.40Erep + -0.40Eatt + 600Eelec + 1.00EDARS$$

Wt LEDGF IBD and Wt HIV Integrase CCD scores

Cluster	Members	Representative	Weighted Score
0	288	Center	-666.1
0	288	Lowest Energy	-696.3
1	285	Center	-669.1
1	285	Lowest Energy	-715.3
2	107	Center	-682
2	107	Lowest Energy	-708.6
3	76	Center	-582.8
3	76	Lowest Energy	-655.3
4	68	Center	-587.3
4	68	Lowest Energy	-660.2
5	37	Center	-596.6
5	37	Lowest Energy	-624.8
6	24	Center	-645.4
6	24	Lowest Energy	-645.4
7	21	Center	-578.8
7	21	Lowest Energy	-667.9
8	18	Center	-584.4
8	18	Lowest Energy	-625.3
9	18	Center	-619.1
9	18	Lowest Energy	-619.1

Wt LEDGF IBD and Mutant HIV Integrase CCD scores

Cluster	Members	Representative	Weighted Score
0	207	Center	-751.5
0	207	Lowest Energy	-751.5
1	145	Center	-607.8
1	145	Lowest Energy	-646.3
2	116	Center	-658
2	116	Lowest Energy	-658
3	76	Center	-620.1
3	76	Lowest Energy	-677.4
4	72	Center	-587
4	72	Lowest Energy	-664.9
5	62	Center	-636.1
5	62	Lowest Energy	-657
6	61	Center	-639.8
6	61	Lowest Energy	-664.4
7	49	Center	-622.8
7	49	Lowest Energy	-673.8
8	34	Center	-654.6
8	34	Lowest Energy	-654.6
9	27	Center	-632.2
9	27	Lowest Energy	-640
10	25	Center	-648.7
10	25	Lowest Energy	-648.7
11	20	Center	-596.1
11	20	Lowest Energy	-626.7
12	19	Center	-651.9
12	19	Lowest Energy	-651.9
13	18	Center	-664.2
13	18	Lowest Energy	-664.2
14	14	Center	-619.8
14	14	Lowest Energy	-642.5
15	12	Center	-606.1
15	12	Lowest Energy	-635.8
16	10	Center	-588.1
16	10	Lowest Energy	-617.7
17	6	Center	-650.6
17	6	Lowest Energy	-650.6

8. References

Part I

DESHAIES, R. J. & C. A. JOAZEIRO (2009). "RING domain E3 ubiquitin ligases." Annu Revn Biochem **78**: 399-434.

DYE, B. T. & B. A. SCHULMAN (2007). "Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins." Annu Rev Biophys Biomol Struct **36**: 131-50.

HICKE, L. (2001). "Protein regulation by monoubiquitin." Nat Rev Mol Cell Biol **2**(3): 195-201.

HOCHSTRASSER, M. (2009). "Origin and function of ubiquitin-like proteins." Nature **458**(7237):422-9.

HOFMANN RM, PICKART CM (1999) No canonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. Cell96:64653

HUANG, A., R. N. DE JONG, ET AL. (2009). "E2-c-Cbl recognition is necessary but not sufficient for ubiquitination activity." J Mol Biol **385**(2): 507-19.

LI, W., M. H. BENGTSON, ET AL. (2008). "Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling." PLoS ONE **3**(1): e1487

LORICK, K. L., J. P. JENSEN, ET AL. (1999). "RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination." Proc Natl Acad Sci U S A **96**(20): 11364-9.

MCKENNA S, SPYRACOPOULOS L, MORAES T ET AL (2001) No covalent interaction between ubiquitin and the human DNA repair protein Mms2 is required for Ubc13- mediated polyubiquitination. J Biol Chem. 276:40120–40126

MCKENNA, S., T. MORAES, ET AL. (2003). "An NMR-based model of the ubiquitin-bound human ubiquitin conjugation complex Mms2.Ubc13. The structural basis for lysine 63 chain catalysis." J Biol Chem **278**(15): 13151-8.

MORAES TF, EDWARDS RA, MCKENNA S ET AL (2001) Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2-hUbc13. Nat Struct Biol 8:669–673

NIJMAN, S. M., M. P. LUNA-VARGAS, ET AL. (2005). "A genomic and functional inventory of Deubiquitinating enzymes." Cell **123**(5): 773-86.

OZKAN, E., H. YU, ET AL. (2005). "Mechanistic insight into the allosteric activation of aubiquitin conjugating enzyme by RING-type ubiquitin ligases." Proc Natl Acad Sci U S A **102**5218890-5

PICKART, C. M. (2001). "Mechanisms underlying ubiquitination." Annu Rev Biochem **70**: 50333.

PICKART, C. M. AND M. J. EDDINS (2004). "Ubiquitin: structures, functions, mechanisms."Biochim Biophys Acta **1695**(1-3): 55-72.

VERDECIA, M. A., C. A. JOAZEIRO, ET AL. (2003). "Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase." Mol Cell **11**(1): 249-

WU, P. Y., M. HANLON, ET AL. (2003). "A conserved catalytic residue in the ubiquitin conjugating enzyme family." EMBO J **22**(19): 5241-50.

Part II

1. ANDARKE, SKALKA AM (1995) Multimerization determinants reside in both the catalytic core and C terminus of avian sarcoma virus integrase. J Biol Chem 270: 29299-29306.

2. BARRÉ-SINOUSSI, F., et al., Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science, 1983. **220**(4599): p. 868-71.

3. BENKIRANE, M., et al., Mechanism of transdominant inhibition of CCR5-mediated HIV-1 infection by ccr5delta32. J Biol Chem, 1997. **272**(49): p. 30603-6.

4.BRENKE R, HALL DR, CHUANG G-Y, COMEAU SR, BOHNUUD T, BEGLOV D, SCHUELER-FURMAN O, VAJDA S, KOZAKOV D. 2012. Application of asymmetric statistical potentials to antibody-protein docking. *Bioinformatics*. 28(20):2608-2614

5. Brown Po COFFIN JM, HUGHES SH, VARMUS HE, editors. Retroviruses. Plainview, NY: Cold Spring Harbor Laboratory. pp. 161-204.

6. CHEN JC, KRUCINSKI J, MIERCKE LJ, FINER-MOORE JS, TANG AH, et al 2000 Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. Proc Natl Acad Sci USA 97: 8233-8238.

7. CHEREPANOV, P., et al., Identification of an evolutionarily conserved domain in Human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. J Biol Chem, 2004. **279**(47): p. 48883-

8. CHEREPANOV, P., et al., HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem, 2003. **278**(1): p. 372-81.

9. COMEAU SR, GATCHELL DW, VAJDA S, CAMACHO CJ. CLUSPRO: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*. 2004 Jan

10. COMEAU SR, GATCHELL DW, VAJDA S, CAMACHO CJ. CLUSPRO: a fully automated algorithm for protein-protein docking *Nucleic Acids Research*. 2004 Jul 1

11. DE RIJCK, J., et al., Over-expression of the LEDGF/p75 Integrase Binding Domain Inhibits HIV Replication. J Virol, 2006.

12. Gallo, R.C., et al., Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science, 1984. **224**(4648): p. 500-3.

13. GE, Y.Z., et al., Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. J Biol Chem, 2004. **279**(24): p. 25447-54.

14. JENKINS TM, ENGELMAN A, GHIRLANDO R, CRAIGIE R (1996) A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. J Biol Chem 271: 7712-7718

15. KOZAKOV D, BEGLOV D, BOHNUUD T, MOTTARELLA S, XIA B, HALL DR, VAJDA, S. How good is automated protein docking? *Proteins: Structure, Function, and Bioinformatics*, 2013 Aug

16. KOZAKOV D, BRENKE R, COMEAU SR, VAJDA S. PIPER: An FFT-based protein docking program with pair wise potentials. *Proteins*. 2006 Aug 24

17. LEVY, J.A., ET al., Isolation of lymphocytopathic retroviruses from San Francisco Patients with AIDS. Science, 1984. **225**(4664): p. 840-2.

18. LI X, KRISHNAN L, CHEREPANOV P, ENGELMAN A (2011) Structural biology of retroviral DNA integration. Virology 411: 194-205.

19. LLANO, M., et al., LEDGF/p75 Determines Cellular Trafficking of Diverse Lentiviral but Not Murine Oncoretroviral Integrase Proteins and Is a Component Of Functional Lentiviral Preintegration Complexes. J Virol, 2004. **78**(17): p. 9524-37.

20. MAERTENS, G., et al., LEDGF/p75 is essential for nuclear and chromosomal

Targeting of HIV-1 integrase in human cells. J Biol Chem, 2003. **278**(35): p. 33528-39.

21. MEADOWS AND GERVAY-HAGUE, ChemMedChem, 2006, 1, 16 – 29

22. OLIVIER DELELIS*1, KEVIN CARAYON1, ALI SAÏB2, ERIC DEPREZ1 AND JEAN- FRANÇOIS Mouscadet Retrovirology 2008, **5**:114 doi: 10.1186/1742-4690-5-114

23. STEC, I., et al., the PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? FEBS Lett, 2000. **473**(1): p. 1-5.

24.WANG JY,LING H,YANG W,CRAIGE R (2001) structure of a two domain fragment of HIV-1 integrase: implications for domain organization in the intact protein Embo J 20: 7333-7343

25. WISKERCHEN, M. AND C. CHENG-MAYER, HIV-1 Nef association with cellular serine kinase correlates with enhanced virion infectivity and efficient proviral DNA synthesis. Virology, 1996. **224**(1): p. 292-301.