ASIAN JOURNAL OF PHARMACEUTICAL AND CLINICAL RESEARCH

INNOVARE ACADEMIC SCIENCES Knowledge to Innovation

Vol 7, Issue 5, 2014

ISSN - 0974-2441

Research Article

STUDY OF CONSERVATION OF p6 REGION OF GAG POLYPROTEIN AMONG MULTIPLE STRAINS OF HUMAN IMMUNODEFICIENCY VIRUS-1 A POSSIBLE DRUG TARGET: A BIOINFORMATICS APPROACH

UNDETY BENJAMIN JASON[1]*, DANIEL ALEX ANAND[2]

[1]Department of Bioinformatics, Sathyabama University, Chennai, Tamil Nadu, India. [2]Department of Biomedical Engineering, Sathyabama University, Chennai, Tamil Nadu, India. Email: ubjason@gmail.com

Received: 01 July 2014, Revised and Accepted: 21 August 2014

ABSTRACT

Objective: The variations of p6 in the host in horizontal transmission were studied. The conserved regions are found in viral factor when analyzed with host factors. Sequence and structural analysis were carried out to understand the interaction between viral and host factors. Sequences of host factors were extracted. Study the variation and conserved regions of P6 protein of GAG gene accross horinzontal transmission.

Methods: Pairwise alignment, multiple sequence alignment, secondary structure analysis, and three-dimensional structure analysis of p6 sequence with host sequences were performed. Databases National Center for Biotechnology Information and protein data bank (PDB) were used to download sequences. Tools Swiss-PDB viewer, Needleman–Wunsch Clustal W and Dicovery Studio version 2.0, were used for the analysis. Through literature host factors TSG101, NEDD4, AIP1, and ALIX were found for p6.

Results: The percentages of conserved regions of p6 with respect to four host factors were calculated. Fragments "frs" and "gvetttppq" have 66.66% and 23.33% with respect to TSG101, "frsg"and "ppeesfrsg" are 85.15% and 46.66% with respect to NEDD4, "eptappeesf" and "idk" are 73.91% and 78.26% with respect to AIP1 and "lqsrpe" and "pqkqe" are 61.53% and 84.61% with respect to ALIX.

Conclusion: Modeled structures of the host and viral factors contain 89% of amino acids in favorable region. Whole study was concerned on finding out variation in horizontal transmission of p6 gag gene protein. Among all the conserved fragments "frsg" of NEDD4, "pqkqe" of ALIX, "idk" of AIP1 and "eptappeesf" of AIP1 are consensus fragments with the presence of hydrophobic amino acids and hence these regions are treated as active sites for viral target. Conformational analysis of host factors reached active potential at 10 µ seconds with viral factor

Keywords: Conserved, Pairwise alignment, Variation, Horizontal transmission.

INTRODUCTION

Human immunodeficiency virus (HIV)-1, which causes acquired immune deficiency syndrome, is a retrovirus in genus lentiviridae. HIV-1 is an enveloped virus, which encodes two envelope (Env) glycoproteins - the surface (SU) glycoprotein gp120 and a transmembrane glycoprotein gp41, gag has four major proteins, they are matrix, capsid (CA), nucleocapsid (NC), and p6-and the pol-encoded enzymes protease, reverse transcriptase (RT), and integrase (IN). HIV-1 also encodes two regulatory proteins, they are tat and rev and several accessory proteins, they are Vpu, Vif, Nef, and Vpr. The genome is pseudo diploid that is composed of two single strands of RNA linked in dimer. The HIV-1 infection initiates with the attachment of gp120 to target cell plasma membrane [1-4]. The principal attachment of the receptor for HIV-1 and other pri-mate lent viruses is CD4. Productive infection also requires the presence of a co-receptor; they are typically CXCR4 or CCR5. The binding of gp120 to CD4 and co-receptor initiates conformational changes in gp41, which in turn directs to fusion of the viral Env and the target cell membrane and entry of the viral core into the host cell cytoplasm. Recent evidence suggests that HIV-1 entry can also occur in a low-pH endosomal compartment after receptor-mediated endocytosis [5]. Upon entry of the virion into the cytosol, the Env glycoproteins and the lipid-associated MA protein dissociate from the incoming particle at the membrane, and the poorly understood process of uncoating is initiated. The enzymes RT and IN, together with the NC protein, remain in close association with the viral RNA as it is converted to double-stranded DNA by RT-catalyzed RT [6]. NC acts as a nucleic acid chaperone at several steps during RT to facilitate the conversion of RNA to DNA [7]. Vpr is also a component of the reverse transcription complex (RTC). The extent to which CA remains associated with the incoming RTC has been a topic of debate. However, RT and uncoating appear to be temporally linked, [8] and it is clear that some host restriction factors that block early post entry steps in the viral replication cycle target CA [9,10]. The newly RT viral DNA is translocated to the nucleus in a structure known as the preintegration complex. The nuclear import process remains incompletely understood; however, a role for CA in this process [11,12] implies that some CA protein may remain associated with the viral nucleoprotein complex as it traffics to the nuclear pore.

METHODS

The sequences of p6 gag gene protein (Homo sapiens) retrieved from National Center for Biotechnology Information (NCBI) using (www.ncbi.nlm.nih.gov) is provided here in fasta format.

>gi|19172951|ref|NP_579883.1| p6 [Human immunodeficiency virus 1]

LQSRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLRSLFGNDPSSQ

Target search is carried out and found TSG101, NEDD4, AIP1 and ALIX, which are the host factors related to p6 gag gene protein. These host factors sequences are retrieved from NCBI. We retrieved 129 sequences for all the four host factors from NCBI. Each host factors have TSG101=63, NEDD4=29, AIP1=23 and ALIX=14 are retrieved from NCBI. Example of each host factor sequences is provided here in fasta format.

Pairwise sequence alignment methods are used to find the best-matching piecewise (local) or global alignments of two query sequences. Pairwise alignments can only be used between two sequences at a time. For this global alignments are done via the Needleman–Wunsch algorithm available at EBI was used. Pairwise alignment is done by giving p6 gag gene virus sequence as sequence 1 and the host factor sequence as

sequence 2 as input. Identifying all the conserved regions of the target and template with more probability is very important step in finding out the variation in horizontal transmission of HIV. For this CLUSTALW (Thompson et al. [1994]) program available at EBI was used. Multiple sequence alignment was done by the alignment of 129 sequences, which are taken from NCBI database which is well-annotated. Multiple sequence alignment is done for each host factor individually.

Secondary structure refers to highly regular local sub-structures. Two main types of secondary structure, the alpha helix and the beta strand. These secondary structures are defined by patterns of hydrogen bonds between the main-chain peptide groups. They have a regular geometry, being constrained to specific values of the dihedral angles ψ and φ on the Ramachandran plot. Both the alpha helix and the beta-sheet represent a way of saturating all the hydrogen bond donors and acceptors in the peptide backbone. Some parts of the protein are ordered, but do not form any regular structures. Secondary structure is the general three-dimensional (3D) form of local segments of biopolymers such as proteins and nucleic acids (DNA/RNA). Secondary structure can be formally defined by the hydrogen bonds of the biopolymer, as observed in an atomic-resolution structure. In proteins, the secondary structure is defined by the patterns of hydrogen bonds between backbone amide and carboxyl groups. P6 gag gene virus protein was taken from protein data bank (PDB) and is analyzed in Swiss-PDB viewer (SPDBV).

The NCBI Entrez Protein database comprises sequences taken from a variety of sources including SwissProt the Protein Information Resource the Protein Research Foundation the PDB and translations from annotated coding regions in the GenBank and RefSeq databases. Protein sequence records in Entrez have links to pre-computed protein basic local alignment search tool alignments, protein structures, conserved protein domains, nucleotide sequences, genomes and genes. All these databases are available online through the Entrez search engine.

The PDB is a repository for the 3D structural data of large biological molecules, such as proteins and nucleic acids. The PDB is a key resource in areas of structural biology, such as structural genomics. The structure files may be viewed using one of several open source computer programs. Some other free, but not open source programs include Rasmol, SPDBV and many others.

SPDBV is an application that provides a user-friendly interface allowing to analyze several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles and distances between atoms are easy to obtain thanks to the intuitive graphic and menu interface. Swiss-PDB viewer can also read electron density maps, and provides various tools to build into the density. In addition, various modeling tools are integrated and residues can be mutated.

The Needleman–Wunsch algorithm performs a global alignment on two sequences (called A and B here). It is commonly used in bioinformatics to align protein or nucleotide sequences. The Needleman–Wunsch algorithm is an example of dynamic programming. Clustal is a widely used multiple sequence alignment computer program.

RESULTS AND DISCUSSION

In the target search, it I have found TSG101, NEDD4, AIP1, and ALIX are p6 gag gene protein targets. We retrieved 129 sequences from NCBI Table 1.

Table 1: Host factors and the sequence count

Name of host factor	Number of sequences
TSG101	63
NEDD4	29
AIP1	23
ALIX	14

Pairwise alignments are used between two sequences at a time, and global alignments are done via the Needleman–Wunsch algorithm available at EBI was used. Pairwise alignment is done by giving p6 gag gene virus sequence as sequence 1 and the host factor sequence as sequence 2 as input. There are the examples of each pairwise alignment of the host factors with virus p6 sequence. The pairwise alignment do not show much significant variations when the p6 protein sequence is aligned pairwise with each host factor. Most probable variations are about 10-30 point mutations in the p6 gag protein with each host factor TSG101, NEDD4, AIP1 and ALIX. This indicates that the p6 Gag gene protein in HIV-1 would be a potential drug target in the gag polyprotein.

TSG101 sequence aligned with p6 gag gene

>gi|9789790|sp|Q99816.2|TS101_HUMAN RecName: Full=Tumor susceptibility gene 101 protein; Alt Name: Full=ESCRT-I complex subunit TSG101

```
1 ---------------------------------------------------- 0
MAVSESQLKKMVSKYKYRDLTVRETVNVITLYKDLKPVLDSYVFNDGSSR 50
1 ---------------------------------------------------- 0
51 ELMNLTGTIPVPYRGNTYNIPICLWLLDTYPYNPPICFVKPTSSMTIKTG 100
1 ----------------------------------------LQSRP----- 5
:.|||
101 KHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSRPISASY 150
6 ---EPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLRSLFGNDPSSQ 52
:.|.||.|:.|:.| || |.::.||.
151 PPYQATGPPNTSYMPGMPGGISP----------YP-----SGYPPNPSGY 185
53 ------------------------------------------------- 52
186 PGCPYPPGGPYPATTSSQYPSQPPVTTVGPSRDGTISEDTIRA SLISAVS 235
53 ------------------------------------------------- 52
236 DKLRWRMKEEMDRAQAELNALKRTEEDLKKGHQKLEE MVTRLDQEVAEVD 285
53 ------------------------------------------------- 52
286 KNIELLKKKDEELSSALEKMENQSENNDIDEVIIPTAPLYK QILNLYAEE 335
53 ------------------------------------------------- 52
336 NAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRAL MQKARKTAG 385
53 ----- 52
386 LSDLY 390
```

NEDD41 sequence aligned with p6 gag gene

>gi|161172341|pdb|3B7Y|B Chain B, Crystal Structure Of The C2 Domain Of The E3 Ubiquitin- Protein Ligase Nedd4

```
1 ----------------------------------LQSRPEPTA----- 9
:.:|.
1 GMATCAVEVFGLLEDEENSRIVRVRVIAGIGLAKKDILGASDPYVRVTLY 50
10 PPEESFRSGVETTTPPQKQEPIDKELYPLTSLRSLFGNDPSSQ------- 52
.|.::.|:|.| |.|.|.|.::||.|.
51 DPMNGVLTSVQTKT-------IKKSLNPKWNEEILFRVHPQQHRLLFEVF 93
53 ------------------------------------------------- 52
94 DENRLTRDDFLGQVDVPLYPLPTENPRLERPYTFKDFVLHPRSHKSRVKG 143
53 ---------- 52
144 YLRLKMTYLP 153
```

AIP1 sequence aligned with p6 gag gene.

>gi|12643797|sp|Q99961.1|SH3G1_HUMAN RecName: Full=Endophilin-A2; AltName: Full=EEN fusion partner of MLL; AltName: Full=Endophilin-2; AltName: Full=Extra eleven-nineteen leukemia fusion gene protein; Short=EEN; AltName: Full=SH3 domain protein 2B; AltName: Full=SH3 domain-containing GRB2-like protein 1

```
1 ---------------------------------------------------- 0
1 MSVAGLKKQFYKASQLVSEKVGGAEGTKLDDDFKEMEKKVDVTSKAVTEV 50
1 ---------------------------------------------------- 0
51 LARTIEYLQPNPASRAKLTMLNTVSKIRGQVKNPGYPQSEGLLGECMIRH 100
1 ---------------------------------------------------- 0
101 GKELGGESNFGDALLDAGESMKRLAEVKDSLDIEVKQNFIDPLQNLCEKD 150
1 ---------------------------------------------------- 0
151 LKEIQHHLKKLEGRRLDFDYKKKRQGKIPDEELRQALEKFEESKEVAETS 200
1 ---------------------------------------------------- 0
201 MHNLLETDIEQVSQLSALVDAQLDYHRQAVQILDELAEKLKRRMREASSR 250
1 --LQSRPEPTAP-----PEESFRSGVETTTPPQ---------KQEPI--- 31
.:.:|:|.| ||:|.|.||.|::||
251 PKREYKPKPREPFDLGEPEQS-NGGFPCTTAPKIAASSSFRSSDKPIRTP 299
32 DKELYPL--TSLRSLFGNDPSSQ---------------------------- 52
.:.:.||.|.:|::|.:.
```

300 SRSMPPLDQPSCKALYDFEPENDGELGFHEGDVITLTNQIDENWYEGMLD 349
53 ---------- 52
350 GQSGFFPLSYVEVLVPLPQ 368

ALIX sequence aligned with p6 gag gene

>gi|35210502|dbj|BAC87888.1| Snf7 homologue associated with Alix 3 [Homo sapiens]
1 ---------- 0
1 MSKLGKFFKGGGSSKSRAAPSPQEALVRLRETEEMLGKKQEYLENRIQRE 50
1 ---------- 0
51 IALAKKHGTQNKRAALQALKRKKRFEKQLTQIDGTLSTIEFQREALENSH 100
1 ---------- 0
101 TNTEVLRNMGFAAKAMKSVHENMDLNKIDDLMQEITEQQDIAQEISEAFS 150
1 ----------LQSRPEPTAPPEESFR 16
|.:.|.::.|.:.:.:
151 QRVGFGDDFDEDELMAELEELEQEELNKKMTNIRLPNVPSSSLPAQPNRK 200
17 SGVETT------TPPQKQEPIDKELYPLTSLRSLFGNDPSSQ 52
.|:::|.|:.|.|.::|.::
201 PGMSSTARRSRAASSQRAEEEDDDIKQLAAWAT--------- 233

Identifying all the conserved regions of the target and template with more probability is very important step in finding out the variation in horizontal transmission of HIV. For this Clustal W available at EBI was used. Multiple sequence alignment was done by the alignment of 129 sequences which are taken from NCBI.

TSG101 CLUSTALW
gi|239787092|ref|NP_077285.3| -NPMSRDPGSGGWEEAPRAAAA LCTLYHEAGQRLRRLQDQLAARDALIAR 49
gi|20385504|gb|AAM21315.1|AF37 -AFMSRDPGSGGWEEAPRAAAA LCTLYHEAGQRLRRLQDQLAARDALIAR 49
gi|119602922|gb|EAW82516.1| EAWMSRDPGSGGWEEAPRAAAA LCTLYHEAGQRLRRLQDQLAARDALIAR 50
gi|119602920|gb|EAW82514.1| EAWMSRDPGSGGWEEAPRAAAA LCTLYHEAGQRLRRLQDQLAARDALIAR 50
gi|123234960|emb|CAM28233.1| CAMMSRDPGSGGWEEAPRAAA ALCTLYHEAGQRLRRLQDQLAARDALIAR 50
gi|123233229|emb|CAM28252.1| CAMMSRDPGSGGWEEAPRAAA ALCTLYHEAGQRLRRLQDQLAARDALIAR 50
gi|13445188|emb|CAC34835.1| CACMSRDPGSGGWEEAPRAAAAL CTLYHEAGQRLRRLQDQLAARDALIAR 50
gi|31455243|gb|AAH02740.2| AAHMSRDPGSGGWEEAPRAAAAL CTLYHEAGQRLRRLQDQLAARDALIAR 50
gi|239787092|ref|NP_077285.3| LRANP---------- LQSRPEPTAPPEESFRSGVE 74
gi|20385504|gb|AAM21315.1|AF37 LRANP---------- LQSRPEPTAPPEESFRSGVE 74
gi|119602922|gb|EAW82516.1| LRANP---------- LQSRPEPTAPPEESFRSGVE 75
gi|119602920|gb|EAW82514.1| LRANP---------- LQSRPEPTAPPEESFRSGVE 75
gi|123234960|emb|CAM28233.1| LRANP---------- LQSRPEPTAPPEESFRSGVE 75
gi|123233229|emb|CAM28252.1| LRANP---------- LQSRPEPTAPPEESFRSGVE 75
gi|13445188|emb|CAC34835.1| LRANP---------- LQSRPEPTAPPEESFRSGVE 75
gi|31455243|gb|AAH02740.2| LRANP---------- LQSRPEPTAPPEESFRSGVE 75
gi|239787092|ref|NP_077285.3| TTTPP-NPRLAALEG-DAAPSLVDALLEQVARFRE QLRRQEGGAAEAQMR 122
gi|20385504|gb|AAM21315.1|AF37 TTTPP-AFRLAALEG-DAAPSL VDALLEQVARFREQLRR QEGGAAEAQMR 122
gi|119602922|gb|EAW82516.1| TTTPPEAWRLAALEG-DAAPSLVD ALLEQVARFREQLRRQE GGAAEAQMR 124
gi|119602920|gb|EAW82514.1| TTTPPEAWRLAALEG-DAAPSLVDA LLEQVARFREQLRRQE GGAAEAQMR 124
gi|123234960|emb|CAM28233.1| TTTPPCAMRLAALEG-DAAPSLV DALLEQVARFREQLRR QEGGAAEAQMR 124
gi|123233229|emb|CAM28252.1| TTTPPCAMRLAALEG-DAAPSLVD ALLEQVARFREQLRRQE GGAAEAQMR 124
gi|13445188|emb|CAC34835.1| TTTPPCACRLAALEG-DAAPSLVDAL LEQVARFREQLRRQEG GAAEAQMR 124
gi|31455243|gb|AAH02740.2| TTTPPAAHRLAALEG-DAAPSLVDAL LEQVARFREQLRRQEGG AAEAQMR 124
gi|239787092|ref|NP_077285.3| QEIERLT--NPQKQEPIDKELYPLTSLRSLFGNDPS SQ---------- 158
gi|20385504|gb|AAM21315.1|AF37 QEIERLT--NPQKQEPIDKELYPLTSLRSLFGND PSSQ---------- 158
gi|119602922|gb|EAW82516.1| QEIERLT--NPQKQEPIDKELYPLTSLRSLFGNDPS SQ---------- 160
gi|119602920|gb|EAW82514.1| QEIERLT--NPQKQEPIDKELYPLTSLRSLFGNDPS SQ---------- 160
gi|123234960|emb|CAM28233.1| QEIERLT--NPQKQEPIDKELYPLTSLRSLFGNDPS SQ---------- 160
gi|123233229|emb|CAM28252.1| QEIERLT--NPQKQEPIDKELYPLTSLRSLFGNDPS SQ---------- 160
gi|13445188|emb|CAC34835.1| QEIERLT--NPQKQEPIDKELYPLTSLRSLFGNDPS SQ---------- 160
gi|31455243|gb|AAH02740.2| QEIERLT--NPQKQEPIDKELYPLTSLRSLFGNDPS SQ---------- 160
gi|239787092|ref|NP_077285.3| ----------NPERLEEKEREMQQLL S---------- ---------- 175
gi|20385504|gb|AAM21315.1|AF37 ----------AFERLEEKEREM QQLLS---------- 175
gi|119602922|gb|EAW82516.1| ----------EAWERLEEKERE MQQLLS---------- 178
gi|119602920|gb|EAW82514.1| ----------EAWERLEEKEREM QQLLS---------- 178
gi|123234960|emb|CAM28233.1| ----------CAMERLEEKERE MQQLLS---------- 178
gi|123233229|emb|CAM28252.1| ----------CAMERLEEKERE MQQLLS---------- 178
gi|13445188|emb|CAC34835.1| ----------CACERLEEKER EM QQLLS---------- 178
gi|31455243|gb|AAH02740.2| ----------AAHERLEEKERE MQQLLS---------- 178
gi|239787092|ref|NP_077285.3| ----------QPQH EREKEVV LLRRSMA 193
gi|20385504|gb|AAM21315.1|AF37 ----------QPQ HEREK EVVLLRRSMA 193
gi|119602922|gb|EAW82516.1| ----------QPQHE REKEVVLLRRS MA 196
gi|119602920|gb|EAW82514.1| ----------QPQH EREKEVVLLR RSMA 196
gi|123234960|emb|CAM28233.1| ----------Q P QHEREKEVVLLR RSMA 196
gi|123233229|emb|CAM28252.1| ----------QP QHEREKEVVLL RRSMA 196
gi|13445188|emb|CAC34835.1| ----------QPQ HEREKEV VLLR RSMA 196
gi|31455243|gb|AAH02740.2| ----------QPQHE REKEV VLLR RSMA 196

TSG101Host sequences of 63 are downloaded from the database, and when aligned with p6 gag protein using Clustal W. The multiple sequence alignment has proved the previous r esults and their assu mption are correct. Most of the regions studied are conserved according to the alignment attac hed above.

Secondary structure is the general 3D form of local segments of bi opolymers such as prote ins and nucleic acids (DNA/RNA). Secondary structure can be formally defined by the h ydrogen bonds of the biopolymer as observed in an atomic-resolution structure. In proteins, the secondary str ucture is defined by the patterns of hydrogen bonds between backbone amide and carboxyl groups. The conserved fragments, which have the hydrophobic amino acids are identified in the virus sequence using SPDBV.

Hydro means water, phobic means to hate and hence hydrophobic are those substances that do not dissolve in water. Hydrophobic amino acids are found in the active sites for the targets to be bounded. The active site is part of an enzyme where substrates bind and undergo a chemical reaction. The majority of enzymes are proteins. The active site of the enzyme is usually found in a cleft or pocket that is lined by amino acid residues (or nucleotides in ribozymes) that participates in recognition of the substrate. Residues that directly participate in the catalytic reaction mechanism are called active site residues.

Here the SPDBV results are given in which the fragments with hydrophobic amino acids are identified with different colors for each host factors.

The secondary structure analysis is performed by SOPMA and Net Surf P.

Secondary structure results, SOPMA

Alpha helix (Hh)	: 8 is 15.38%
3_{10} helix (Gg)	: 0 is 0.00%
Pi helix (Ii)	: 0 is 0.00%
Beta bridge (Bb)	: 0 is 0.00%
Extended strand (Ee)	: 1 is 1.92%
Beta turn (Tt)	: 1 is 1.92%
Bend region (Ss)	: 0 is 0.00%
Random coil (Cc)	: 42 is 80.77%
Ambigous states (?)	: 0 is 0.00%
Other states	: 0 is 0.00%

By Net Surf, P server predicted the SU accessibility and secondary structure of amino acids in the amino acid sequence. The method also simultaneously predicts the reliability for each prediction, in the form of a Z-score. The Z-score is related to the SU prediction.

CONCLUSIONS

The variation in horizontal transmission of p6 gag gene protein across the population is generated by identifying host factors for p6 gag gene protein, they are TSG101, NEDD4, AIP1 and ALIX. These host factor sequences are taken using NCBI and are aligned with the virus sequence. The pair-wise alignment is carried done by Needleman–Wunsch algorithm. The multiple sequence alignment is generated by taking the aligned sequences and is carried out using Clustal W through which conserved regions are identified to study the variation in horizontal transmission. The secondary structure of p6 gag gene protein is taken to find out the active site and identifying the most consensus fragment sequence, which are having hydrophobic amino acids. Whole study was concerned on finding out the variation in horizontal transmission of p6 gag gene protein.

REFERENCES

1. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: Roles in viral entry, tropism, and disease. Annu Rev Immunol 1999;17:657-700.
2. Doms RW. Beyond receptor expression: The influence of receptor conformation, density, and affinity in HIV-1 infection. Virology 2000;276(2):229-37.
3. Freed EO, Martin MA. HIVs and their replication. In: Knipe DM, Howley PM, editors. Fields Virology. Philadelphia: Lippincott, Williams & Wilkins; 2007. p. 2107-85.
4. Melikyan GB. Common principles and intermediates of viral protein-mediated fusion: The HIV-1 paradigm. Retrovirology 2008;5:111.
5. Miyauchi K, Kim Y, Latinovic O, Morozov V, Melikyan GB. HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. Cell 2009;137(3):433-44.
6. Telesnitsky A, Goff SP. Reverse transcriptase and the generation of retroviral DNA. In: Coffin J, Varmus H, Hughes S, editors. Retroviruses. Cold Spring Harbor New York: Cold Spring Harbor Laboratory; 1997. p. 121-60.
7. Levin JG, Guo J, Rouzina I, Musier-Forsyth K. Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: Critical role in reverse transcription and molecular mechanism. Prog Nucleic Acid Res Mol Biol 2005;80:217-86.
8. Hulme AE, Perez O, Hope TJ. Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. Proc Natl Acad Sci U S A 2011;108(24):9975-80.
9. Song B. TRIM5alpha. Curr Top Microbiol Immunol 2009;339:47-66.
10. Towers GJ. The control of viral infection by tripartite motif proteins and cyclophilin A. Retrovirology 2007;4:40.
11. Dismuke DJ, Aiken C. Evidence for a functional link between uncoating of the human immunodeficiency virus type 1 core and nuclear import of the viral preintegration complex. J Virol 2006;80(8):3712-20.
12. Yamashita M, Emerman M. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. J Virol 2004;78(11):5670-8.