Genome wide snRNP motifs and regulatory sequences in HIV1 isolates

Paushali Roy*, Protip Basu, Sayak Ganguli and Abhijit Datta
DBT-Centre for Bioinformatics, Presidency College, Kolkata, India.
*corresponding author: paushali.06@gmail.com

Abstract-The pathogenesis of HIV-1 is complex and characterized by the interplay of both viral and host factors. Within HIV 1 genome there are several snRNP motifs responsible for pre mRNA splicing and stabilization. By locating these motifs within the genome and disturbing them may result in an impaired ability of the cells to sustain HIV-1 replication. One of such regulatory sequences is riboswitches that regulate the dimerization of HIV-1 RNA, which is an essential step during packaging. The current work was undertaken to identify possible regulatory RNA motifs in the HIV1 genome from different isolates. The current work has successfully identified multiple snRNP motifs in the genome sequences of different strains of HIV-1 isolates. The identification of the multiple snRNP motifs in the genomic sequences of the various isolates lead us to believe that future studies with artificially constructed snRNPs might have the potential to inhibit HIV1 replication. Apart from containing snRNP motifs they also possess regulatory riboswitch motifs. Riboswitches bind metabolites and control the dimerization and packaging of the genome. Thus the occurrence of such motifs further strengthens the idea that apart from serving as a regulatory domain for structural constraints such motifs may also regulate genome integration and production of the necessary products by using the host transcriptional machinery. It is however beyond doubt that such sequence motifs must have originated in the RNA world as they have the power to mediate RNA induced regulation of gene expression.
Keywords: Riboswitch, snRNP motifs, HIV-1, gene regulation, RNA processing

INTRODUCTION

The human immunodeficiency virus type 1 (HIV-1) is the primary cause of the acquired immunodeficiency syndrome (AIDS), which is a slow, progressive and degenerative disease of the human immune system. The pathogenesis of HIV-1 is complex and characterized by the interplay of both viral and host factors. Despite years of intensive research and some therapeutic success, AIDS, continues to be a major health problem worldwide. It is a type of lentivirus and widely recognized as the etiologic agent of acquired immunodeficiency syndrome (aids). It is characterized by its cytopathic effect and affinity for the t4-lymphocyte. The strains of HIV-1 can be classified into three groups: the "major" group M, the "outlier" group O and the "new" group N. These three groups may represent three separate introductions of simian immunodeficiency virus into humans [1]. The enzyme reverse transcriptase (RT) is used by HIV 1 (a retrovirus) to transcribe their single-stranded RNA genome into single-stranded DNA and to subsequently construct a complementary strand of DNA, providing a DNA double helix capable of integration into host cell chromosomes. Functional HIV1-RT is a heterodimer containing subunits of 66 kDa (p66) and 51 kDa (p51) [2]. Many viral or cellular genes are involved in HIV-1 multiplication and therefore represent potential targets. Indeed, several strategies attempting to interfere with the production or function of such gene products are being tested at pre-clinical or clinical levels. Within HIV 1 genome there are several snRNPs (small nuclear ribonucleoproteins) motifs mainly responsible for pre mRNA splicing and stabilization. By locating these motifs within the genome and disturbing them may result in an impaired ability of the cells to sustain HIV-1 replication. HIV-1 pathogenesis is multifactorial and involves complex interactions between host and viral genes [3]. Several regulatory sequences that play significant role in HIV-1 infection have been so far identified. One of such regulatory sequences is riboswitches that regulate the dimerization of HIV-1 RNA, which is an essential step during packaging. Riboswitches are complex folded RNA domains that serve as receptors for specific metabolites. These domains are found in the non-coding portions of various mRNAs, where they control gene expression by harnessing allosteric structural changes that are brought about by metabolite binding. New findings indicate that riboswitches are robust genetic elements that are involved in regulating fundamental metabolic processes in many organisms. The riboswitches are made up of the three-dimensional structure of RNA, in which RNA can undergo two mutually exclusive conformations in response to an environmental signal in the form of a metabolite. Riboswitches comprise two domains: an aptamer and an expression platform. The aptamer is highly conserved even in distantly related organisms, and serves as a precise sensor for its target metabolite. The expression platform is far more variable in sequence and in structure as it can function by assuming one of many structural forms to control gene expression [4]. For all these reasons riboswitches seem to be a significant form of genetic control. The advantage of this system is that they are highly specific to their substrates. But the lack of universality is because riboswitch-mediated translational regulation is limited to mono-cistronic m-RNA. Riboswitches have been shown to function in repressing gene expression (negative regulation) in response to metabolites. If riboswitch is a primitive mode of gene regulation4 then it suggests that in the RNA world negative regulation was predominant. There are also speculations that riboswitch-mediated control mechanism might

also be playing an important role in eukaryotic gene expression. Unlike prokaryotes, eukaryotes have distinct compartments for transcription and translation and m-RNA is monocistronic. These favour post-transcriptional and translational gene regulation. A riboswitch-mediated gene regulation is expected to be acting either at post-transcriptional or translational level [5]. Each riboswitch is able to bind with high specificity their cellular target metabolite, without the involvement of a protein cofactor. Upon metabolite binding, the messenger RNA undergoes structural change that will ultimately lead to the modulation of its genetic expression. Riboswitches can alter gene expression at the level of transcription attenuation or translation initiation, and can up- or down-regulate gene expression by harnessing appropriate changes in the mRNA structure.

The following riboswitches are known:

- *TPP riboswitch* (also THI-box) binds thiamin pyrophosphate (TPP) to regulate thiamin biosynthesis and transport, as well as transport of similar metabolites [6].
- *FMN riboswitch* (also *RFN-element*) binds flavin mononucleotide (FMN) to regulate riboflavin biosynthesis and transport [7].
- *Cobalamin riboswitch* (also *B12-element*), which binds adenosylcobalamin (the coenzyme form of vitamin B12) to regulate cobalamin biosynthesis and transport of cobalamin and similar metabolites, and other genes [8].
- *SAM riboswitches* bind S-adenosyl methionine (SAM) to regulate methionine and SAM biosynthesis and transport. Three distinct SAM riboswitches are known: *SAM-I* (originally called *S-box*), *SAM-II* and the *SAM (MK)* riboswitch. SAM-I is widespread in bacteria, but SAM-II is found only in alpha-, beta- and a few gamma-proteobacteria. SAM (MK) is found only in the order Lactobacillales. These three varieties of riboswitch have no obvious similarities in terms of sequence or structure [9].
- *Purine riboswitches* binds purines to regulate purine metabolism and transport. Different forms of the purine riboswitch can bind either guanine (a form originally known as the *G-box*) or adenine. The specificity for either guanine or adenine depends completely upon Watson-Crick interactions with a single pyrimidine in the riboswitch at position Y74. In the guanine riboswitch this residue is always a cytosine (i.e. C74), in the adenine residue it is always a uracil (i.e. U74) [10]
- *Lysine riboswitch* (also *L-box*) binds lysine to regulate lysine biosynthesis, catabolism and transport [10].
- *glmS riboswitch*, which is a ribozyme that cleaves itself when there is a sufficient concentration of glucosamine-6-phosphate [11].
- *Glycine riboswitch* binds glycine to regulate glycine metabolism genes, including the use of glycine as an energy source. As of 2007, this riboswitch is the only known natural RNA that exhibits cooperative binding, which is accomplished by two adjacent aptamer domains in the same mRNA [12].
- *Magnesium riboswitch* senses magnesium ions to regulate magnesium transport genes [13].
- *PreQ1 riboswitch* binds pre-queuosine$_1$, to regulate genes involved in the synthesis of this precursor to queuosine. The binding domain of this riboswitch is unusually small among naturally occurring riboswitches [14].
- *Adenine* riboswitch, smallest riboswitch that activates gene expression upon ligand binding [15].
- *Guanine* riboswitch [16].

Thus a likely scenario is such that there may be more than one regulatory RNA motif that may be present and be involved in the control of gene expression not only at the transcriptional level but also co – translationally. The current work was undertaken to identify possible regulatory RNA motifs in the HIV1 genome from different isolates.

MATERIALS AND METHOD

A total of 42 sequences of HIV1 genome were collected from the NCBI databases. RNA sequences of all the sequences of HIV1 genome were designed using a number of closely related software applications. General information of RNA sequences (including poly (A) signal, catalytic RNA, snRNP motifs, putative UTRs, promoter regions, AU-rich regions, etc.) were retrieved using software tool-box for analysis of regulatory RNA elements. Apart from this, regulatory sequences/riboswitches in the RNA sequences were determined using a RNA motif search program and web server which checks specific sequence elements and secondary structure, calculates and displays the energy folding of the RNA structure and runs a number of tests including this information to determine whether high-sensitivity riboswitch motifs (or variants) are present in the given RNA sequence. Batch-mode determination (all sequences input at once and separated by FASTA format) is also possible. Structures of riboswitches were

analyzed. Riboswitch sequences were obtained from the RNA sequences.

RESULTS

snRNP motifs: Among 42 sequences studied, snRNP motifs were identified in only 34 sequences. Total snRNP motifs-1194 in 34 sequences.

Riboswitch regulatory sequences: Among 42 sequences studied, riboswitches are found in 30 sequences [Table-I].

DISCUSSION:

Maturation of pre-mRNA in eukaryotes involves a set of factors that modify the primary transcript into an export competent mRNP (messenger ribonucleoprotein) complex, the formation of which, signals its release from the site of transcription and export into the cytoplasm for translation. These processing events include capping, splicing, 3' end cleavage and polyadenylation. Proper 3' end formation promotes transcription termination and nuclear export of the mRNA and enhances the stability and translation of the mature transcript in the cytoplasm. Recent evidence implicates the existence of extensive coupling between the gene expression machinery engaged in mRNA synthesis, processing, export and surveillance suggesting that disruption of RNA processing will compromise proper gene expression [17]. The current work has successfully identified multiple snRNP motifs in the genome sequences of different strains of HIV-1 isolates. It is known that binding of modified snRNP inhibits the polyadenylation step in 3' end formation by disrupting the protein–protein interactions in the polyadenylation machinery [18]. In contrast, interaction between snRNP and the major splice donor site downstream of the HIV-1 5' LTR poly(A) site inactivates the cleavage step in a U1 70 K dependent manner and is thought to be important for production of full-length genomic RNA [19]. Thus the identification of the multiple snRNP motifs in the genomic sequences of the various isolates lead us to believe that future studies with artificially constructed snRNPs might have the potential to inhibit HIV1 replication. Further more the snRNP motifs if found homologous with any other such motifs in the human spliceosomal machinery may also provide insight on some other possible mode of prevention of HIV1 infection. Most of the sequences under study have revealed that apart from containing multiple snRNP motifs they also possess regulatory riboswitch motifs. Riboswitches, as we know bind metabolites and control the dimerization and packaging of the genome. Thus the occurrence of such motifs further strengthens the idea that apart from serving as a regulatory domain for structural constraints such motifs may also regulate genome integration and further production of the necessary products by using the host transcriptional machinery. It is however beyond doubt that such sequence motifs must have originated in the RNA world as they have the power to mediate RNA induced regulation of gene expression.

Riboswitch Topology:

Most of the riboswitches that were obtained consist of simple stem loops and bulges. Very few are there showing branches. The details of each riboswitches are as follows:

- >gi|9629914|ref|NC_001870.1|Simian-Human immunodeficiency virus, complete genome: There are 6 bulges (one is incomplete) and 1 stem loop structure [Fig. (**1**)].
- >gi|156067835|gb|EU030417.2|HIV-1 isolate cd07_005 pol protein (pol) gene, partial cds: There are 4 bulges (one is incomplete) and 1 stem loop structure [Fig. (**2**)].
- >gi|151368121|gb|EF545108.1|HIV-1 isolate RU00051, complete genome: There are 4 bulges (one is incomplete) and 1 stem loop structure [Fig. (**3**)].
- >gi|149939408|gb|EF633445.1|HIV-1 isolate R1, complete genome: There are 3 bulges and 1 stem loop structure. The third bulge consists of 2 branches; 1st branch consists of 1 bulge and 1 stem loop and the 2nd branch consist of 2 bulges (one is incomplete) and 1 stem loop structure [Fig. (**4**)].
- >gi|149211372|gb|EF078279.2|HIV-1 isolate TW-D118, partial genome: There are 5 bulges (one is incomplete) and one stem loop structure [Fig. (**5**)].
- >gi|149211364|gb|EF078278.2|HIV-1 isolate TW-D4, partial genome: There are 5 bulges (one is incomplete) and one stem loop structure [Fig. (**6**)].
- >gi|117581798|gb|EF036534.1|HIV-1 isolate Fj065 complete genome: there are 6 bulges of which two motifs are concurrent ones. All bulges are attached by stem helices. [Fig (**7**)].
- >gi|117643940|gb|EF029066.1|HIV-1 isolate U.NL.95.H10986_D1, complete genome: There are 5 bulges and one stem loop structure [Fig. (**8**)].
- >gi|62532627|gb|AY857174.2|HIV-1 isolate 9340158 partial genome: There are 5 bulges (one is incomplete) and one stem loop structure [Fig. (**9**)].
- >gi|113171669|gb|DQ854716.1|HIV-1 isolate U61 complete genome: There are 7 bulges (one is incomplete) and one stem loop structure [Fig. (**10**)].
- >gi|94959078|gb|DQ487188.1|HIV-1 isolate WCD32P0793, complete

genome: There are 7 bulges (one is incomplete) and one stem loop structure [Fig. (**11**)].

- >gi|121581791|dbj|AB289588.1|Human immunodeficiency virus 1 proviral DNA, complete genome, clone: p00JPDR2508B60: There are 7 bulges (one is incomplete) and one stem loop structure [Fig. (**12**)].
- >gi|116518650|gb|DQ886035.1|HIV-1 isolate L8157, complete genome: 5 bulges of which two are concurrent on the main stem. The main stem culminates in an incomplete bulge loop [Fig. (**13**)].
- >gi|51980229|gb|AY612637.1|HIV-1 isolate PT2695, complete genome: There are 7 bulges (one is incomplete) and one stem loop structure [Fig. (**14**)].
- >gi|119370128|gb|EF091932.1|HIV-1 isolate 110PA, complete genome: There are 7 bulges (one is incomplete) and one stem loop structure [Fig. (**15**)].
- >gi|119507379|dbj|AB287367.1|Human immunodeficiency virus 1 proviral DNA, complete genome, clone: p05JPDR7060B68: There are 6 bulges (one is incomplete) and one stem loop structure [Fig. (**16**)].
- >gi|112497920|gb|DQ676884.1|HIV-1 isolate PS3002_Day199, complete genome: There are 6 bulges (one is incomplete) and one stem loop structure [Fig. (**17**)].
- >gi|86277646|gb|DQ358812.1|HIV-1 isolate 02BR034, complete genome: There are 5 bulges (one is incomplete) and one stem loop structure [Fig. (**18**)].
- >gi|63098379|gb|DQ011175.1|HIV-1 isolate 03ZASK005B2, complete genome: There are 4 bulges and one stem loop structure [Fig. (**19**)].
- >gi|50404184|gb|AY682547.1|HIV-1 isolate 04RU128005, complete genome: 4 bulge loops terminated by incomplete bulge [Fig. (**20**)].
- >gi|29409314|gb|AY093605.1|HIV-1 isolate 96GH2911, complete genome: There are 6 bulges (one is incomplete) and one stem loop structure [Fig. (**21**)].
- >gi|18643009|gb|AY074891.1|HIV-1 isolate 00BWMO35.1, complete genome: 4 bulges with 3 stems and one concurrent bulge motif. [Fig. (**22**)]
- >gi|17902136|gb|AF423759.1|HIV-1 isolate X477, complete genome: There are 7 bulges (one is incomplete) and one stem loop structure [Fig. (**23**)].
- >gi|15788267|gb|AF408629.1|HIV-1 isolate A32879, complete genome: There are 5 bulges (one is incomplete) and one stem loop structure [Fig. (**24**)].
- >gi|7321133|emb|AJ276595.1|HIV-1 proviral gag gene, pol gene, vif gene, vpr gene, vpu gene, env gene, nef gene, tat gene and rev gene, isolate VI1035, genomic RNA: There are 5 bulges (one is incomplete) and two stem loop structures [Fig. (**25**)].
- >gi|47680175|gb|AY530889.1|HIV-2 isolate 96FR12034, complete genome: There are 5 bulges (one is incomplete) and one stem loop structure [Fig. (**26**)].
- >gi|114801456|gb|DQ853465.1|HIV-1 isolate 14301.1 complete genome: There are 6 bulges (one is incomplete) and one stem loop structure [Fig. (**27**)].
- >gi|114801377|gb|DQ853457.1|HIV-1 isolate 15389.1, complete genome: There are 6 bulges (one is incomplete) and one stem loop structure [Fig. (**28**)].
- >gi|114801193|gb|DQ853438.1|HIV-1 isolate 14300.1, complete genome: There are 7 bulges (one is incomplete) and one stem loop structure [Fig. (**29**)].
- >gi|108860394|dbj|AB231893.1|Human immunodeficiency virus 1 proviral DNA, complete genome, isolate: GHNJ175: There are 2 bulges and 1 stem loop structure. The second bulge consists of two branches; 1st branch consists of 1 bulge and 1 stem loop structure and the second branch consists of three concurrent bulges [Fig. (**30**)].

REFERENCES

[1] Nishitsuji H. , Kohara M. , Kannagi M. and Masuda T. (2006) *J Virol*, 80, 7658-7666.

[2] Jacobo-Molina A. , Ding J. , Nanni R.G. , Clark A.D. , Jr., Lu X. , Tantillo C. , Williams R.L. , Kamer G. , Ferris A.L. , Clark P. , Hizi A, Hughes S.H. and Arnold E. (1993) *Proc. Natl. Acad. Sci. USA* 90, 6320-6324.

[3] Chakraborti S. and Banerjea A.C. (2003) *Mol Ther* 7, 817-826.

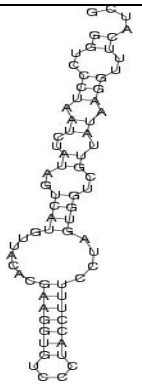
[4] Mandal M. and Breaker R. R. (2004) *Nature Reviews Molecular Cell Biology* 5, 451-463.

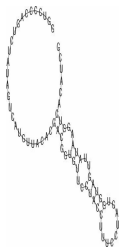
[5] Ray S. Kumar (2004) *Current Science* 87(9).

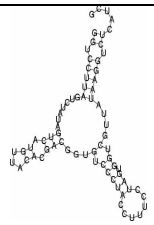
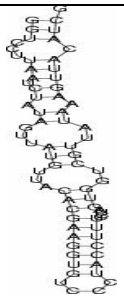
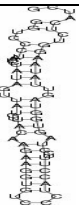
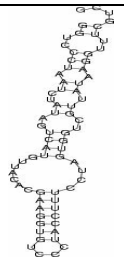
[6] Miranda-Ríos J. , Navarro M. and Soberón M. (2001) *Proc Natl Acad Sci U S A* 98(17), 9736-41.

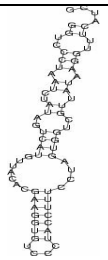
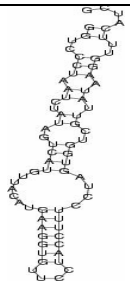
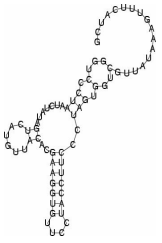
[7] Winkler W.C. , Cohen-Chalamish S. and Breaker R.R. (2002) *Proc Natl Acad Sci U S A* 99(25), 15908-13.
[8] Nahvi A. , Barrick J.E. and Breaker R.R. (2004) *Nucleic Acids Res.* 32(1), 143-50.
[9] Grundy F.J. and Henkin T.M. (1998) *Mol Microbiol* 30(4), 737-49.
[10] Mandal M. (2003) *Cell* 113, 577–586.
[11] Winkler W.C. , Nahvi A., Roth A. , Collins J.A. and Breaker R.R. (2004) *Nature* 428, 281–286.
[12] Mandal M. , Lee M. , Barrick J.E. , Weinberg Z. , Emilsson G.M. , Ruzzo W.L. and Breaker R.R. (2004) *Science* 306, 275–279.
[13] Tucker B.J. and Breaker R.R. (2005) *Curr Opin Struct Biol* 15(3), 342-8.
[14] Roth A. (2007) *Nat Struct Mol Biol* 14(4), 308-17.
[15] Kaempfer R. (2003) *EMBO reports* 4(11), 1043–1047.
[16] Batey R.T. , Gilbert S.D. and Montange R.K. (2004) *Nature* 432, 411-415.
[17] Sajic R. , Lee K. , Asai K. , Sakac D. , Branch D.R. , Upton C. and Cochrane A. (2007) *Nucleic Acids Research*, 35, 247–255.
[18] Vagner S. , Rüegsegger U. , Gunderson S.I. , Keller W. and Mattaj I.W. (2000) *RNA*, 6, 178–188.
[19] Ashe M.P. , Furger A. and Proudfoot N.J. (2000) *RNA*, 6, 170–177.

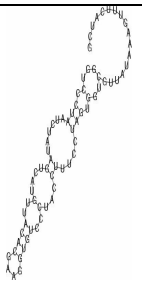
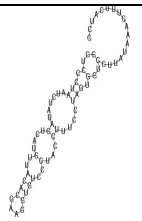
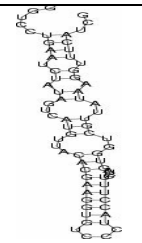
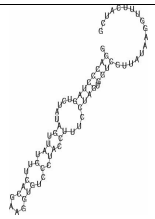
TABLE I: Identified snRNP motifs and riboswitches in HIV-1 genome

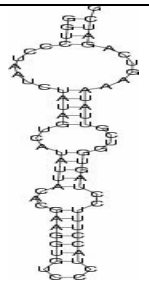
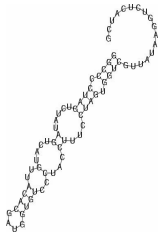
S.No.	Accession No.	Name	Country	snRNP	Riboswitch	Riboswitch Position
1.	>gi|2828036|gb|AF038398.1|AF038398	Simian-Human immunodeficiency virus strain SHIV-89.6, complete genome	NA	Start site: 581 706 749 1190 1492 1943 2198 2282 2389 2836 2997 3161 3348 3490 3583 3916 4315 4647 4663 4805 4814 4920 5099 5727 818 5872 6000 17189 7381 7446 7668 8710 9257	Not found	
2.	>gi|9629914|ref|NC_001870.1|	Simian-Human immunodeficiency virus, complete genome	NA	Start site: 581 706 749 1190 1492 1943 2198 2282 2389 2836 2997 161 348 3490 3583 3916 4315 4647 4663 4805 4814 4920 5099 727 5818 5872 6000 7189 7381 7446 7668 8710 9257 10299 10525 10722 10781 10885 11235 11540 12237 12347 12393 12996 13023 13059 13132 13345 14079 14201 14247 14321 14404 14425 15320 15429 16183 16246 16298 16612 16875 17109 17674 18148 18568 18950	 Fig. (**1**)	+ strand, position 12513 And + strand, position 12606
3.	>gi|156067837|gb|EU030418.2|	HIV-1 isolate vl05_153 pol protein (pol) gene, partial cds	India	NONE	Not found	
4.	>gi|156067835|gb|EU030417.2|	HIV-1 isolate cd07_005 pol protein (pol) gene, partial cds	India	Start site: 184 438 473 550 596	 Fig. (**2**)	+ strand, position 716
5.	>gi|155964970|gb|EU008325.1|	HIV-1 isolate patient D clone M envelope glycoprotein (env) gene, partial cds	Belgium	NONE	Not found	
6.	.>gi|155964942|gb|EU008311.1|	HIV-1 isolate patient D clone L envelope glycoprotein (env) gene, partial cds	Belgium	NONE	Not found	
7.	>gi|9628880|ref|NC_001722.1|	Human immunodeficiency virus 2, complete genome	NA	Start site: 275 693 1273 1316 1338 1757 1862 1973 3186 3436 3547 3597 3664 3761 4090 4204 4409 4915 5037 5083 5250	Not found	

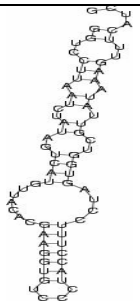
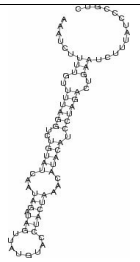
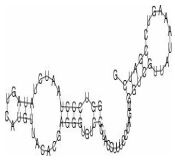
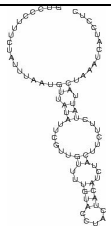
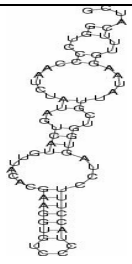
				5270 5405 5645 5866 6435 6488 6667 6992 7013 7092 7597 7971 8061 8135 8149 8663 8951 9033 9455 9779 10197		
8.	.>gi|151368121|gb|EF545108.1|	HIV-1 isolate RU00051, complete genome	Russia	Start site: 55 108 183 282 477 949 986 1785 1978 2134 2308 2735 2762 2798 2871 3084 3256 3816 3938 3984 4058 4162 4897 5054 5163 5536 6028 6080 6190 6345 6545 6579 6602 6708 6809 7381 7869 8019 8164 8286 8464	 Fig. (**3**)	+ strand, position 2254 And + strand, position 2347
9.	>gi|149939408|gb|EF633445.1|	HIV-1 isolate R1 , complete genome	South Africa	Start site: 97 325 522 581 685 1035 1293 1778 1788 2037 2070 2623 2796 2823 2859 2921 2932 3145 3879 4001 4047 4121 4204 4225 4732 4965 5122 5231 5455 5602 5915 6023 6167 6399 6624 6689 6914 6929 7956 8379	 Fig. (**4**)	+ strand, position 2313 And + strand, position 2406
10.	>gi|149211372|gb|EF078279.2|	HIV-1 isolate TW-D118, partial genome	Taiwan	Start site: 12 87 186 374 433 537 889 1186 1895 1972 2018 2621 2821 3057 3166 3308 3517 3635 3733 4799 5218	 Fig. (**5**)	+ strand, position 2231
11.	>gi|155964862|gb|EU008271.1|	HIV-1 isolate patient D clone P envelope glycoprotein (env) gene, partial cds	Belgium	NONE	Not found	
12.	>gi|155676107|gb|EU079115.1|	HIV-1 isolate 0063513 protease (pol) gene, partial cds	China	NONE	Not found	
13.	>gi|155624970|gb|EU045534.1|	HIV-1 isolate GDLEVV_189 pol protein (pol) gene, partial cds	Mexico	NONE	Not found	

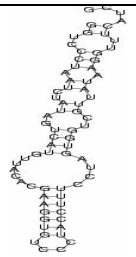
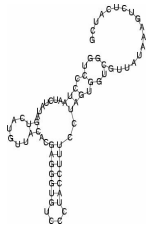
14.	>gi|149211364|gb|EF078278.2|	HIV-1 isolate TW-D4, partial genome	Taiwan	Start site: 12 87 186 374 433 537 889 1138 1186 1895 1972 2018 2621 2684 2746 2757 2970 3307 3411 4072 4308 4417 4556 5326 5848 5966 6064 7130 7550	 Fig. (**6**)	+ strand, position 2231
15.	>gi|30793424|dbj|AB085203.1|	Human immunodeficiency virus 1 genes for pol protein (p6pol region) and gag protein (p6gag region), partial cds, isolate: patient #18	NA	NONE	Not found	
16.	>gi|117581798|gb|EF036534.1|	HIV-1 isolate Fj065 complete genome	China	Start site: 319 820 888 1191 1405 1669 1704 1804 2003 2700 2856 3459 3486 3522 3595 3808 3980 4542 4664 4710 4784 4867 4888 5187 5625 5782 5891 6033 6162 6819 7084 7344 7536 7967 8181 8578 8873 9000 9382	 Fig. (**7**)	+ strand, position 2976, + strand, position 3069 And - strand, position 7554
17.	>gi|117643940|gb|EF029066.1|	HIV-1 isolate U.NL.95.H10986_D1, complete genome	Netherlands	Start site: 777 876 1064 1123 1227 1542 1579 1879 2576 2686 2732 2936 3335 3362 3398 3684 4418 4540 4586 4660 4764 4999 5658 5909 6219 6566 6698 6936 6963 7226 7463 8075 8148 8508 8574 8949 9132	 Fig. (**8**)	+ strand, position 2852 And + strand, position 2945
18.	>gi|62532627|gb|AY857174.2|	HIV-1 isolate 9340158 partial genome	Australia	Start site: 132 492 551 655 706 964 1000 1714 1968 2080 2126 2300 3358 3404 3476 3573 4399 4504 5024	 Fig. (**9**)	+ strand, position 2246
19.	>gi|113171669|gb|DQ854716.1|	HIV-1 isolate U61 complete genome	Spain	Start site: 756 882 981 1178 1237 1341 1996 2434 2693 2803 2849 3452 3479 3515 3588 3801 3954 4535 4657 4703 4777 4860 4881 5388 5621 5778 5887 6638 6722 7286 7351 7579 7628		+ strand, position 2969 **And** + strand, position 3062

				8624 9044	Fig. (**10**)	
20.	>gi|94959078|gb|DQ487188.1|	HIV-1 isolate WCD32P0793, complete genome	USA	Start site: 297 720 779 882 1232 1537 2249 2359 2405 3008 3035 3071 3144 3357 4091 4213 4259 4333 4416 4437 4944 5098 5236 5334 5443 6333 6365 6386 6859 6901 6924 7161 7176 8227 8646	 Fig. (**11**)	+ strand, position 2525 And + strand, position 2618
21.	>gi|125661184|gb|EF363123.1|	HIV-1 clone ES1-20, complete genome	USA	Start site: 753 979 1176 1235 1339 1689 1758 1906 1937 1994 2691 2801 2847 3450 3477 3513 3586 3799 4533 4655 4701 4775 4879 5543 5888 6651 6763 6854 7272 7339 7583 8148 8832 9042	Not found	
22.	>gi|121581791|dbj|AB289588.1|	Human immunodeficiency virus 1 proviral DNA, complete genome, clone: p00JPDR2508B60	NA	Start site: 976 1179 1238 1342 1692 1947 1997 2161 2709 2742 2819 3221 3468 3495 3531 3604 3817 4551 4673 4719 4793 4876 4899 5404 5555 5693 5791 5900 6345 6748 7257 7376 7410 7440 7556 8128 8667 9036	 Fig. (**12**)	+ strand, position 2985 And + strand, position 3078
23.	>gi|116518650|gb|DQ886035.1|	HIV-1 isolate L8157, complete genome	USA	Start site: 204 401 460 564 914 983 1162 1219 1928 2038 2687 2714 2812 2823 3036 3770 3892 3938 4012 4095 4118 4657 4774 5010 5119 5788 5936 6515 6580 6799 7429 7835 8267 8474	 Fig. (**13**)	+ strand, position 2204 And + strand, position 2297

24.	>gi|51980229|gb|AY612637.1|	HIV-1 isolate PT2695, complete genome	Portugal	Start site: 703 930 1124 1183 1287 1602 1639 1939 2374 2633 2666 2743 2789 3154 3455 3741 3956 4034 4597 4643 4717 4800 4821 5558 5715 5824 5966 6068 6263 6710 6778 6839 7358 7568 8089 8634 8930 8977 9057	 Fig. (**14**)	+ strand, position 2909
25.	>gi|119370128|gb|EF091932.1|	HIV-1 isolate 110PA, complete genome	Brazil	Start site: 259 624 683 787 1391 1439 1864 2233 2882 2909 2945 3007 3018 3231 4087 4133 4207 4311 5043 5200 5308 5449 5633 6078 6281 6803 6860 7042 7607 8022 8375 8836	 Fig. (**15**)	+ strand, position 2399 And + strand, position 2492
26.	>gi|119507379|dbj|AB287367.1|	Human immunodeficiency virus 1 proviral DNA, complete genome, clone: p05JPDR7060B68	NA	Start site: 769 827 902 1201 1260 1364 1679 1714 2019 2716 2826 2872 3538 3611 3824 3996 4558 4680 4726 4800 4883 4904 5798 6655 6776 7090 7350 7557 8608 8818 8940 9064	 Fig. (**16**)	+ strand, position 2992 And + strand, position 3085
27.	>gi|112497920|gb|DQ676884.1|	HIV-1 isolate PS3002_Day199, complete genome	Australia	Start site: 16 384 443 546 893 1137 1192 1888 1997 2670 2706 2779 2992 3143 3283 3843 3889 3963 4046 4067 4321 4570 4720 4954 5060 5999 6452 6465 6514 6733 7307 7779 8183	 Fig. (**17**)	+ strand, position 2162 And + strand, position 2254
28.	>gi|86277646|gb|DQ358812.1|	HIV-1 isolate 02BR034, complete genome	Brazil	Start site: 337 436 633 692 796 1148 1401 1451 2178 2288 2937 2964 3000 3073 3286 3451 3579 4020 4142 4188 4262 4345 4366 4873 5025 5260 5369 5511 5603 5790 6063 6166 6269 6330 6575 6852 6909 7044 8110 8176 8433 8480 8560 8940	 Fig. (**18**)	+ strand, position 2547

29.	>gi|63098379|gb|DQ011175.1|	HIV-1 isolate 03ZASK005B2, complete genome	South Africa	Start site: 241 316 603 662 766 1118 1415 2098 2133 2210 2256 2859 2886 2922 2984 2995 3208 3942 3967 4064 4110 4184 4267 4288 5025 5291 5433 6282 6804 6861 6925 7677 8503	 Fig. (**19**)	+ strand, position 2376
30.	>gi|50404184|gb|AY682547.1|	HIV-1 isolate 04RU128005, complete genome	Russia	Start site: 236 292 464 660 718 821 1146 1167 1469 2159 2269 2314 2488 2741 2914 2941 3049 3262 3434 4116 4162 4236 4319 4340 4846 5028 5073 5229 5338 6146 6531 6680 6701 6762 6987 8022 8440 8822	 Fig. (**20**)	+ strand, position 2527
31.	gi|37935919|gb|AY169810.1|	HIV-1 isolate 96CMABB637, complete genome	Cameroon	Start site: 358 417 456 642 806 1123 1256 1680 1973 2059 2080 2525 2794 2924 2951 3003 3151 3799 4005 4247 4353 4646 5225 5241 5843 6041 6122 6195 6321 6560 6779 7229 7495 7798 8286 8403 8412 8954 8986	Not found	
32.	>gi|29409314|gb|AY093605.1|	HIV-1 isolate 96GH2911, complete genome		Start site: 387 446 864 901 1195 1621 1793 1880 2036 2639 2702 2775 2988 3153 3844 3890 3964 4047 4068 4575 4686 4726 4962 5123 5216 5318 5838 5987 6300 6728 6743 7079 7761 7827 8058 8181	 Fig. (**21**)	+ strand, position 2156 And + strand, position 2249
33.	>gi|18643009|gb|AY074891.1|	HIV-1 isolate 00BWMO35.1, complete genome	Botswana	Start site: 256 355 540 599 703 1055 1307 1355 2038 2150 2196 2400 2799 2826 2862 2924 2935 3148 3441 3882 4004 4050 4105 4188 4219 5005 5103 5212 5354 5518 5890 5960 6054 6095 6112 6350 6574 6631 6847 7456 7910 8206 8333	 Fig. (**22**)	+ strand, position 2316

34.	>gi|17902136|gb|AF423759.1|	HIV-1 isolate X477, complete genome	Spain	Start site: 202 428 570 621 680 784 1099 1136 1436 1871 2130 2163 2240 2490 2952 3238 4093 4139 4213 4296 4317 4824 5053 5210 5319 5461 5563 5758 6013 6079 6139 6477 6497 6689 6754 6981 8037 8331 8378	 Fig. (**23**)	+ strand, position 2406
35.	>gi|15788267|gb|AF408629.1|	HIV-1 isolate A32879, complete genome	Argentina	Start site: 320 546 743 802 906 1258 1558 2255 2365 2411 3014 3041 3077 3150 3363 3528 4158 4219 4339 4443 4953 5104 5340 5449 6341 6644 6924 6981 7146 8170 8236 8466 8513 8590 8973	 Fig. (**24**)	+ strand, position 2624
36.	>gi|7321133|emb|AJ276595.1|	HIV-1 proviral gag gene, pol gene, vif gene, vpr gene, vpu gene, env gene, nef gene, tat gene and rev gene, isolate VI1035, genomic RNA	NA	Start site: 296 484 543 999 1248 1296 2017 2127 2776 2803 2901 2912 3125 3297 3859 3981 4027 4101 4184 4215 4470 4897 5099 5208 5350 5406 5545 6008 6166 6428 6685 6907 7516 8266 8772	 Fig. (**25**)	+ strand, position 2293
37.	>gi|110347746|gb|DQ672626.1|	HIV-1 clone 5 nonfunctional envelope glycoprotein (env) gene, complete sequence	Italy	NONE	Not found	
38.	>gi|47680175|gb|AY530889.1|	HIV-2 isolate 96FR12034, complete genome	France	Start site: 190 283 752 795 1349 1697 1868 2248 2449 2470 2667 2884 2951 3404 4174 4515 4561 4741 4977 5341 5894 5981 6446 7056 7562 7632 7898 8023 8155 8935 9128	 Fig. (**26**)	+ strand, position 8823
39.	>gi|114801456|gb|DQ853465.1|	HIV-1 isolate 14301.1 complete genome	USA	Start site: 215 266 341 646 705 809 1124 1159 1461 2158 2191 2268 2917 2944 2980 3053 3266 4000 4122 4168 4242 4325 4346 5240 5349 6100 6273 6538 6730 6798 7035 7050 8143		+ strand, position 2434 And + strand, position 2527

				8287	Fig. (**27**)	
40.	>gi|114801377|gb|DQ853457.1|	HIV-1 isolate 15389.1, complete genome	USA	Start site: 215 266 342 637 696 800 1150 1452 1604 2149 2182 2259 2305 2908 2935 2971 3044 3257 3991 4113 4159 4233 4316 4337 5231 5340 6094 6264 6529 6559 6721 6788 7003 8111	 Fig. (**28**)	+ strand, position 2425 And + strand, position 2518
41	>gi|114801193|gb|DQ853438.1|	HIV-1 isolate 14300.1, complete genome	USA	Start site: 215 266 341 646 705 809 1124 1159 1461 2167 2200 2277 2323 2926 2953 2989 3062 3275 4008 4130 4176 4250 4333 4354 5249 5358 6109 6279 6544 6736 6804 7041 7056 7560 8149 8293	 Fig. (**29**)	+ strand, position 2443 And + strand, position 2536
42.	>gi|108860394|dbj|AB231893.1|	Human immunodeficiency virus 1 proviral DNA, complete genome, isolate: GHNJ175	NA	Start site: 1212 1271 1375 1690 1825 2030 2727 2837 3087 3549 3835 4000 4050 4128 4691 4737 4811 4894 4915 5422 5918 6060 6162 6879 7372 7404 7602 7617 7648 8173 8250 8668 8964	 Fig. (**30**)	+ strand, position 3003 And + strand, position 3096