# Direct and restoring mutations in the light of Mutual Information of the HIV-1 alternative splicing sites

**Rochel Pergamentz**
Genome Diversity Center, Institute of Evolution, Haifa University Mt. Carmel, Haifa 31905, Israel

## Abstract

HIV is a small retrovirus that evades human immune system. It is a mystery of HIV viruses how a virus that contained only nine proteins can capture control of its host and switch the cell to complete virus replication. Part of this mystery believed to be partial or alternative RNA splicing in the regulation of virus ability to establish an infection and reproduction. In this aspect very attractive are mathematical approach to find relevant signals between the splicing sites associated with partial RNA splicing.
We have analyzed genome nucleotide sequences mutual information pattern, and variability in the splicing elements of human immunodeficiency virus type 1 (HIV-1), applying algorithms for computing mutual information pattern on data sets of 178 nucleotide sequences. Our goal was to understand the association between different splicing elements in HIV-1 alternative splicing during the course of viral infection. Specifically we were interested to see how information theory function, mutual information, reflects amount of information between exonic splicing sites and donor /acceptor sites, that either splicing element provides about the other (information that those sites share), under normal conditions and local mutations. During course of study we revealed model for Mutual Information closeness of the compared nucleotides pair and internal logic of proximity between chosen nucleotide pairs allaying splicing sites. The pattern of revealed model is periodical with period of 16 nucleotide. Model period is disrupted under local mutation and afterwards restored with a second splicing site mutation. Results are in the parallel with laboratory results, where restored both HIV-1 replication and regulated viral splicing after second splicing site mutation was observed. Our accomplishments give new knowledge about variations in splicing sites and flow of information in the network of alternative splicing signals.

**KEYWORDS:** alternative splicing, HIV, mutual information, splicing sites, local mutation

## INTRODUCTION

The current knowledge about the HIV-1 alternative splicing mechanism has been tremendously increased; however, there is still largely missing an understanding how all splicing regulatory elements integrated, unified and coordinated.
Alternative splicing is a regulated process during gene expression that results in a single gene coding for multiple proteins. In this process, particular exons of a gene may be included within, or excluded from the final, processed messenger RNA produced from that gene The level of splicing at each of its 3'-splice sites is determined through a combination of positively acting exonic splicing enhancer (ESE) elements, negatively acting exonic and intronic splicing silencer elements (ESS and ISS elements, and the 5'-

splice sites of the regulated exons. Regulation of these splicing elements can determine the dominance of the positive or negative elements on HIV-1 alternative splicing. Both mutations of HIV-1 splicing elements and over expression or inhibition of cellular splicing factors that bind to these elements have been used to show that disruption of regulated splicing inhibits HIV-1 replication. ESSV a 24 nucleotide long hnRNP A/B-dependent exonic splicing silencer within HIV-1 exon 3, act as a repressor at 3' A2 splicing site, responsible for vpr mRNA and also inclusion of HIV-1 exon 3 (3'ss A2) (Amendt & Stoltzfus, 1995, Domsic, et al., 2003, Dowling, et al., 2008, Exline, et al., 2008, Madsen & Stoltzfus, 2005, Mandal, et al., 2008, Mandal, et al., 2009, Stoltzfus & Madsen,2006). ESSV is important in 3'ss splicing and in accumulation of wild-type levels of unspliced viral mRNA, Gag protein formation, and production of virus particles. Disruption of HIV-1 ESSV by local mutagenesis inhibits viral replication. Second site reversion during ESSV mutant lengthen culture - 3'ss A2 or 5'ss D3, bring back virus production "within transiently transfected cells to levels similar to that of pNL4-3-transfected" (Madsen & Stoltzfus, 2005). Disruption of ESSV results in an increased level of vpr mRNA and an almost complete inclusion of the noncoding exon 3, which is flanked by 3'ss A2 and 5'ss D3. Experimental results observed in the laboratory experiments have an associated parallel in mathematical function calculation results, specifically we observed restoration of periodicity in mutual information graphs when second side reverse mutation where included in study. Thus, there appears to be a finely tuned balance between ESSV inhibition and downstream 5'ss enhancement that allows appropriate level of splicing at 3'ss A2 by cross talk between the 3'ss and the downstream 5'ss across an exon (Stoltzfus & Madsen, 2006). The disruption of splicing pattern that can produce aberrant splice variants with decreased HIV infectivity and production (Tazi, et al., 2010). also may be a point for a future investigation reflected by mathematical observation.

## MATERIALS AND METHODS

We used the data set of 178 complete genome nucleotide sequences retrieved from the Los Alamos HIV sequences database, the infectious HIV-1 molecular clone pNL4-3 complete genome of 11 nucleotide sequences, the data set of patient PIC1365 11 time points of complete HIV-1 genome sequences, HIV-1 isolate full sequences PIC1362 patient 20 time points. All sequences in study were of type B HIV-1 isolates. Splicing sites and silencers enhancers' location referenced from authors (Amendt & Stoltzfus, 1995, Madsen & Stoltzfus, 2005, Si, et al., 1998, Swanson & Stoltzfus, 1998) research laboratory experiment results.

### MUTUAL INFORMATION CONTENT OF MULTIPLE ALIGNGMENTS

The multiple alignments were in all cases made by CLUSTALX. (Higgins , et al., 1996, Larkin, et al., 2007) and infectious molecular clone (accession number: M19921) used us master sequence. To find the relevant signal in the multiple alignments we implement in the study the probability theory and information theory Mutual Information content (Bolshoy & Volkovich, 2009, Bolshoy, et al., 2010 ), Kullback & Leibler, 1951), Peleg, et al., 2003, Peleg,et al., 2004) of retrieved sequences. This measure quantifies amount of information, or reduction of uncertainty, that knowing either variable provides about the other. In general, the Mutual information content for position *i* between *x* and *y* in the alignment may be written as:

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left( \frac{p(x,y)}{p(x)\,p(y)} \right),$$

(Wikipedia http://en.wikipedia.org/wiki/Mutual_information)
Where x and y, are nucleotides from compared splicing sites of interest. In one case we compared mutual information content between splicing silencer ESSV and splicing acceptors, donors, enhancers and silencers before mutation after mutation and after second site restoring mutation. Results were in the form of matrix with nucleotide positions in the sequence, on horizontal, and the observed mutual information through all matched sites, namely A1, A2, A3, A4c, A4a, A4b, A5, A6, A7, D2, D3, and D4 on vertical for 16 possible nucleotide combinations for each named site. In the study splicing sites compared in the region of 25 nucleotides on each site catch splicing site even it shifted.We felt that it is important to verify alignments splicing site location, considering the literature data on sites location. Alignments were checked with Web Logo3 http://weblogo.threeplusone.com/create.cgiound to be in compliance with splicand f ing site position and is represented below Fig.1.

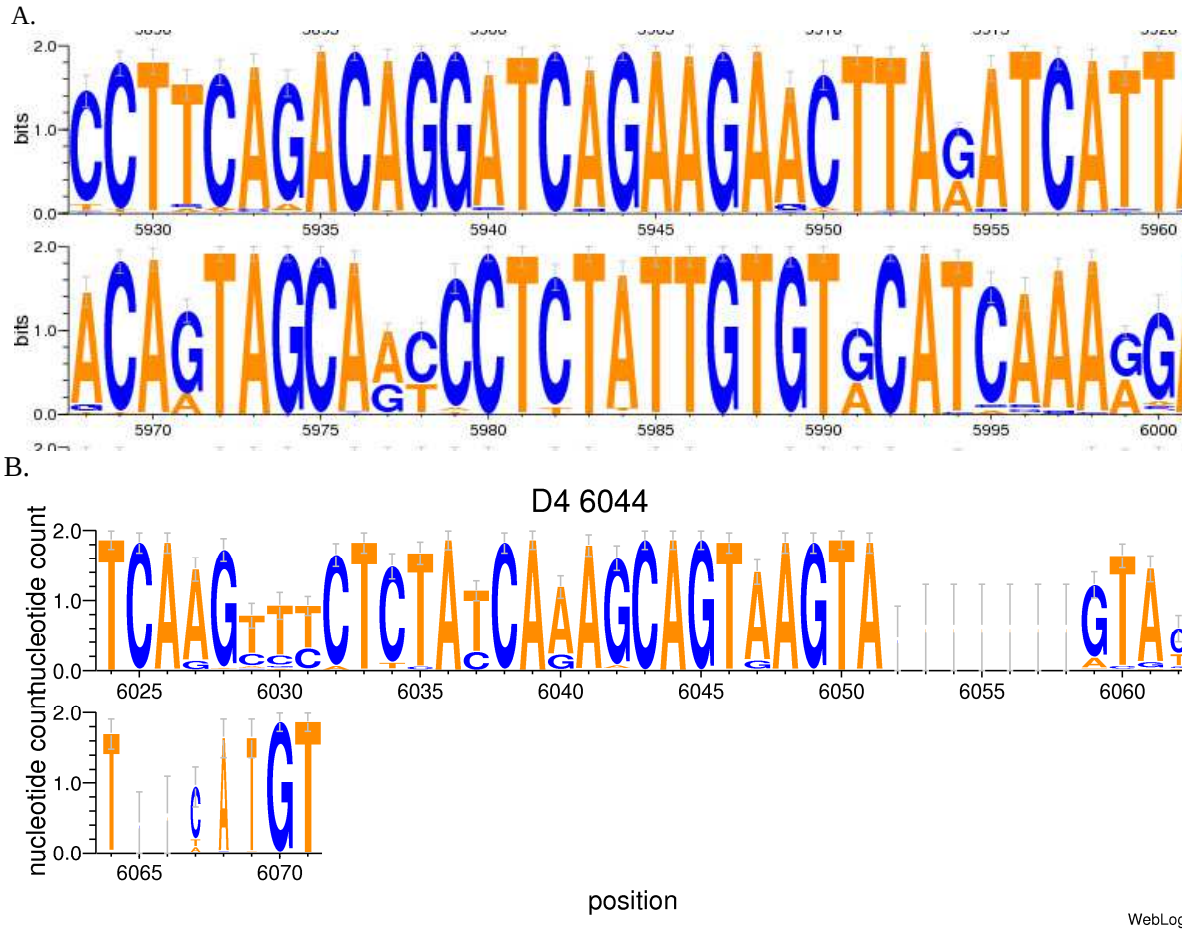



Fig.1. WebLogo on Los Alamos database nucleotide sequences on horizontal: nucleotide position on vertical: nucleotide count. a). Weblogo on 178 nucleotide sequences. Splicing acceptor site A5 canonical dinucleotide AG genome position 5976. AG dinucleotide of the splicing site shift to position 5974. b) Weblogo on 11 molecular clone pNL4-3 nucleotide sequences. Splicing donor site D4 consensus GT dinucleotide genome position 6045.

**RESULTS**

In our work we study mutual information pattern between HIV splicing silencers / enhancers and HIV Acceptor and Donor site elements. We found against the background of the revealed model, that the closeness of the nucleotide pair and internal logic of proximity in pattern of Mutual Information is periodical with expected period of 16 nucleotide of similar chosen pair. The pattern is persistent between splicing enhancers and silencers and cis elements studied. Procedure (order) correctly matches togetherness of elected pair. Results for Mutual Information of ESE nucleotide content compared to nucleotide content of acceptor or donor sites of HIV-1 Type B of 178 sequences retrieved from Los Alamos HIV database in position 8421, displays closeness of the pair with 16 nucleotides period (Fig. 2)

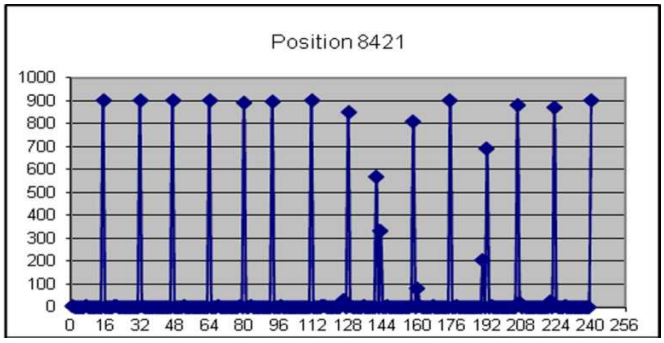



Fig. 2. Mutual Information between ESE and acceptor / donor sites of HIV-1 Type B of 178 sequencesat position 8421

The same pattern observed in the infectious HIV-1 molecular clone pNL4-3 (Fig. 3.)

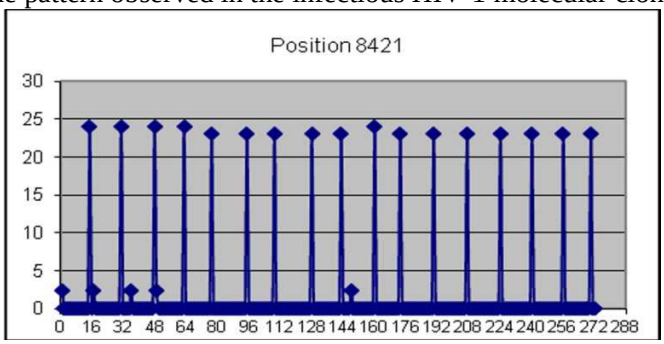



Fig.3. Graph displayed Mutual Information pattern between exonic splicing enhancer ESE nucleotide content versus nucleotide content of A1, A2, A3, A4c, A4a, A4b, A5, A6, A7, D2, D3, D4 acceptor, donor and splicing silencers sites of HIV-1 molecular clone pNL4-3.

Calculated mutual information between psi mutant and donor or acceptor sites of HIV-1 molecular clone pNL4-3 has period, which is disturbed with experiment mutation. In this study we used local mutations data from the publication of (Mougel, et al., 2007) (Fig 4).

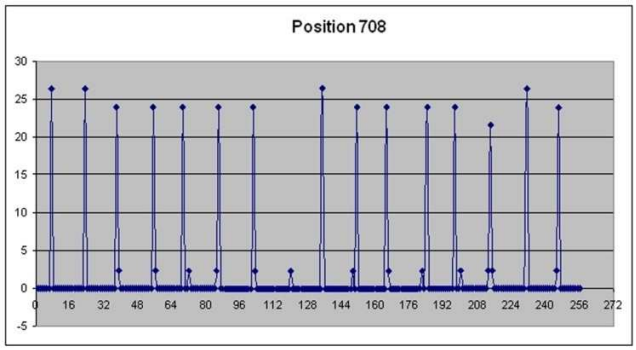
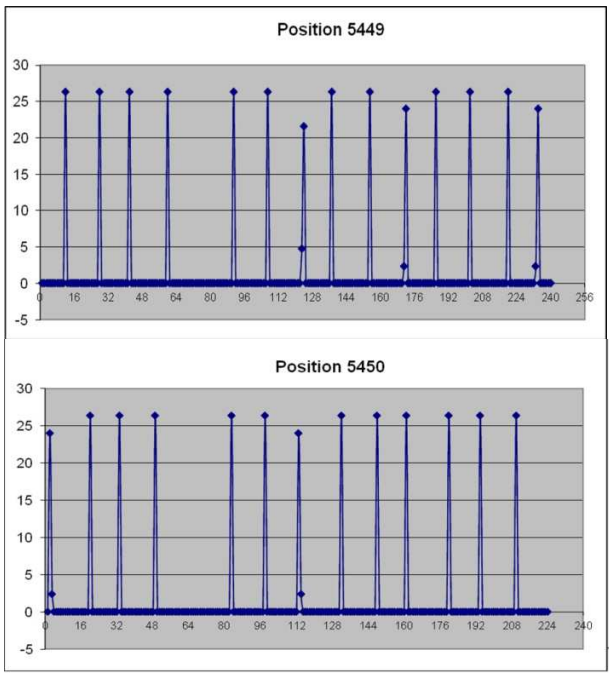



Fig.4 Mutual information between psi mutant and donor and acceptor sites of HIV-1 molecular clone pNL4-3Period is disturbed with local mutation in psi region.

Study of nucleotide sequences patient PIC1365with mutation in ESSV exonic splicing silencer consequently confirms that initially observed pattern is disturbed by mutation. Fig. 5(a) displays study results for position 5449 and 5450. Thereafter second site mutation corrected back mutual information pattern in our calculation and in the parallel restored virus multiplication impaired by the first mutation Fig. 5 (B) at the same positions. Local mutations and experiment results in the study from publications of Joshua M. Madsen and C. Martin Stoltzfus (Madsen & Stoltzfus, 2005).

**A**



**B**

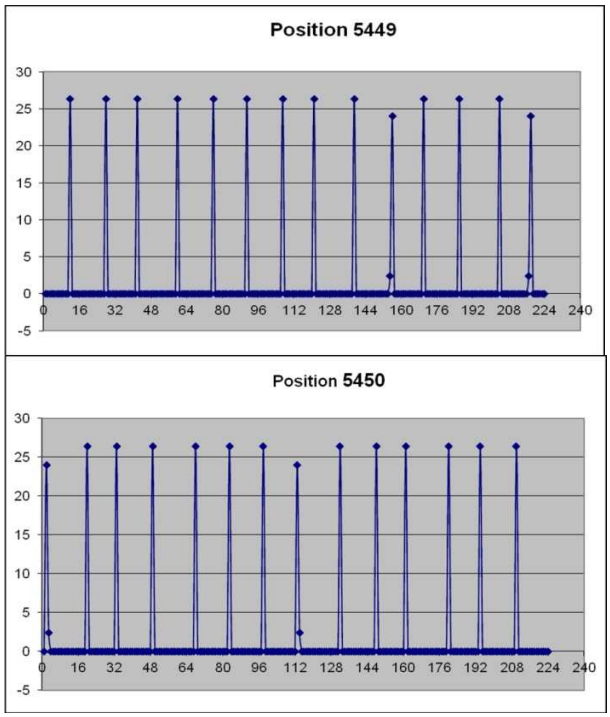



Fig. 5 (A).Initial pattern disturbed by mutation.Study results at positions 5449 and 5450for nucleotide sequences of patient PIC1365with mutation in ESSV exonic splicing silencer. (B)Impaired by the first mutation mutual information pattern is corrected back by second site mutation. Computation results for positions5449 and 5450 nucleotide sequence in study of the patient PIC1365 **Los Alamos database.**

## DISCUSSION

Mutual Information of exonic splicing silencer vs. acceptor/donor sites under against background of developed model displayed the closeness of nucleotide pair with 16 nucleotide period. Model period is disrupted under local mutation as showed in the results for ESSV exonic splicing silencer mutant NEVM1 and restoration of the pattern in the second-site reversions mutation and this observation is in the parallel with observed by Madsen J. and M. Stoltzfus ( Madsen & Stoltzfus, 2005) restored both HIV-1 replication and regulated viral splicing.

## ACKNOWLEDGMENTS

We thank Valery Kirzhner and Alexander Bolshoy for critical evaluation of the manuscript, Department of Molecular Evolution, Haifa University, Israel for support.

## REFERENCES

A**mendt, B.A., Si, Z.H. and Stoltzfus, C.M. (1995)**. Presence of exon splicing silencers within human immunodeficiency virus type-1 tat exon-2 and tat-rev exon-3: evidence for inhibition mediated by cellular factors. Mol. Cell. Biol. 1995, **15(8)**, 4606.

**Bolshoy A. , Volkovich Z. (2009).**Whole-genome prokaryotic clustering based on gene lengths, Discrete Applied Mathematics **157**, 2370-2377

**Bolshoy, A., Volkovich, Z., Kirzhner, V. and Barzily, Z. (2010).**Genome Clustering: from Linguistic Models to Classification of Genetic Texts. Springer-Verlag, Berlin Heidelberg.

**Domsic, J.K., Wang, Y.B., Mayeda, A., Krainer, A.R. and Stoltzfus, C.M. (2003).** Human immunodeficiency virus type 1 hnRNP A/B-Dep endent exonic splicing silencer ESSV antagonizes binding of U2AF65 to viral polypyrimidine tracts, Mol Cell Biol.**23**, 8762-8772.

**Dowling, D., Nasr-Esfahani, S., Tan, C.H., O'Brien, K., Howard, J.L., Jans, D.A., Purcell, D.F.J., Stoltzfus, C.M. and Sonza, S. (2008).** HIV-I infection induces changes in expression of cellular splicing factors that regulate alternative viral splicing and virus production in macrophages, Retrovirology**5**:18.

**Exline, C.M., Feng, Z. and Stoltzfus, C.M. (2008)**. Negative and positive mRNA splicing elements act competitively to regulate human immunodeficiency virus type 1 Vif gene expression, *J Virol.*, **82**, 3921-3931.

**Higgins D., Thompson J., and Gibson T. (1996).** Using CLUSTAL for multiplesequence alignments. Methods Enzymol. **266**, 383 –402.

**Kullback, S. and Leibler, R.A. (1951).** On information and sufficiency.Ann. Math. Stat.**22**, 79-86.

**Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A. & other authors. (2007).** Clustal W and clustal X version 2.0, Bioinformatics. **23**, 2947-2948.

**Madsen, J.M., Stoltzfus, C.M. (2005).** An Exonic Splicing Silencer Downstream of the 3' Splice Site A2 Required for Efficient Human Immunodeficiency Virus Type 1 Replication.J Virol. **79**, 10478–10486 ).

**Mandal, D., Feng, Z.H. and Stoltzfus, C.M. (2008).** Gag-processing defect of human immunodeficiency virus type 1 integrase E246 and G247 mutants is caused by activation of an overlapping 5 ' splice site. J Virol.**82**, 1600-1604.

**Mandal, D., Exline, C.M., Feng, Z.H. and Stoltzfus, C.M. (2009)**. Regulation of vif mRNA Splicing by Human Immunodeficiency Virus Type 1 Requires 5 ' Splice Site D2 and an Exonic Splicing Enhancer To Counteract Cellular Restriction Factor APOBEC3G. J. Virol. **83**, 6067-6078.

**Mougel, M., Marquet, R., Houzet, L., Paillart, J.C., Smagulova, F., Maurel, S., Morichaud, Z. (2007).** HIV controls the elective packaging of genomic, spiced viral and cellular RNA intovirions through different mechanism. Nucleic Acids Res. **35(8)**, 2695-2704

**Peleg, O., Trifonov, E.N. and Bolshoy, A. (2003).** Hidden messages in the nef gene of human immunodeficiency virus Type 1 suggest a novel RNA secondary structure, Nucleic Acids Res.**31**, 4192-4200.

**Peleg, O., Kirzhner, V., Trifonov, E.N. and Bolshoy, A. (2004)**. Overlapping messages and survivability, J. Mol. Evol.**59**, 520-527.

**Si, Z.H., Rauch, D. and Stoltzfus, C.M.(1998)**. The exon splicing silencer in human immunodeficiency virus type 1 tat exon 3 is bipartite and acts early in spliceosome assembly. Mol Cell Biol.**18**, 5404-5413.

**Stoltzfus C. Martin and Madsen Joshua, M.(2006).** Role of Viral Splicing Elements and Cellular RNA Binding Proteins in Regulation of HIV-1 Alternative RNA Splicing.Curr HIV Res. **4**, 43-55

**Swanson, A.K. and Stoltzfus, C.M. (1998)**. Overlapping cis sites used for splicing of HIV-1 env/nef and **rev** mRNAs. J. Biol. Chem.**273**, 34551-34557.

**Tazi, J., Bakkour, N., Marchand, V., Ayadi, L., Aboufirassi A., Branlant, C. (2010).** Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. FEBS Journal **277**, 867–876