Project Abstract

The Physics of Viruses: a look at genome packaging in HIV

This is an abstract of the project I will develop during the term. I have included a great deal of general information, and in italics, I have posed some tentative questions.

Gene expression in HIV is regulated by both cellular and viral controls. HIV genes can be classified as either late or early genes. There are six late genes, localized and expressed in the cytoplasm, and three late genes. HIV transcription is moderated by a single promoter (*describe the statistical mechanics associated with this promoter*). This leads to the production of a 9-kb transcript that can potentially encode all nine HIV genes. This primary transcript is about 600 bp shorter than the provirus. The primary transcript can also be spliced into over thirty different mRNA species or this can be packaged (*how much of the virus is exogenously expressed and how much is noncoding?*). Many of these are polycistronic, or they contain an open reading frame for more than one protein. These polycistronic mRNAs, however typically express a single gene product. LTR has three subregions, referred to as the U3 (unique 3' region), R (repeating central region) and U5 (unique 5' region). The LTR contains binding sites for many cellular transcription factors. Some of the most important of these are those for NF-kappa B-family of transcription factors (*How do activators affect the statistical mechanics of transcription?*). SP-1 sites are also very important for the function of the HIV promoter. Typically, activation of the LTR leads to the production of short transcripts. Some long transcripts are formed that lead to the generation of the Tat protein. The Tat protein then reacts with the TAR element to up the level of viral RNA transcription (*how can this be described mathematically?*).

For example, HIV transcription regulation can be described very similarly to that which we have done in class. In class, we have looked at the case several times in which there is a single promoter which a number of polymerases can bind to. We found the probability of binding to be:

$$P_{bound} = 1/(1 + N_{Ns} e^{\beta \Delta \varepsilon pd}/P)$$

We can also look at how transcripts are modified. As one 9-kb transcript that can potentially encode all nine HIV genes, these processes are very significant. For example, since the primary transcript is about 600 bp shorter than the provirus, we know only 600 bp is immediately noncoding. Since there is only one primer, and the virus does not produce any ribosomes or any other housekeeping elements, this degenerate code must be used for something else. We can estimate the number of polymerase molecules in a cell, and since we know the probability of finding a bound polymerase, we can estimate how often a polymerase is bound, or in other words, how often viral proteins are produced in infected cells.

Much of the information for this abstract was generalized from:

- Hope, Thomas J. and Didier Trono. *Structure, Expression, and Regulation of the HIV Genome*. HIV In Site Knowledge Base Chapter. http://hivinsite.ucsf.edu/InSite?page=kb-02-01-02#S3X. November 2000.

Other significant sources seem to be:

- HIV-1 RNA Structures Index. http://hiv-web.lanl.gov/content/hiv-db/STRUCTURE/RNA.HTML. 2001.

- Russell, Rodney S. Chen Liang and Mark A. Wainberg. Is HIV-1 RNA dimerization a prerequisite for packaging? Yes, no, probably? http://www. retrovirology.com/ content/1/1/23/abstract. 2 September 2004.

- http://www.sciencemag.org/cgi/content/summary/283/5398/8p. 1 January 1999.