

Specificity of Retroviral DNA Integration

John M. Coffin Department of Molecular Biology and Microbiology, Tufts University, Boston MA

Integration of viral into host DNA is the defining event in retrovirus infection. Retrovirus vectors for gene therapy, which rely on the ability of the virus to efficiently stably integrate their DNA, pose a risk of oncogenesis due to effects of the integrated provirus on host cell gene expression. These effects have recently come dramatically into view with the finding of cases of leukemia in children with X-linked SCID successfully treated with a murine leukemia virus (MLV) vector expressing the γ chain of the IL-7 receptor. In these cases, the vector has been found to be integrated near or within the LMO2 gene, a known protooncogene, in associations very much like those seen in leukemias induced by MLV or avian retroviruses. This unfortunate outcome is likely to be due to the interaction of several factors including numbers of independently transduced cells introduced into the patient, interaction of overexpressed LMO2 with the introduced γ chain, and targeting of integration to specific regions or sequences.

Recent genome-wide screening studies from a number of laboratories have implied that at least some retroviruses have regional preferences (strong, but not absolute) for integration. MLV, for example, appears to prefer to integrate its DNA near the start sites of transcription of active genes. To examine possible sequence specificity of integration targeting, we have applied a bioinformatics approach to address the same question. From each published study, we reconstructed the 1 kb integration target sequence for each event for each virus. These sequences were aligned and the frequency of each base at each position (offset) relative to the integration site was determined and compared (using a X^2 test) with the frequency expected for a random base. Results of this analysis revealed a consensus sequence, which, while not strong (the most favored base is present at only about twice the random frequency), has extremely strong statistical support. Similarly, strongly supported consensus sequences were also obtained for MLV and ALV integration targets. However, the sequences obtained were completely unrelated to one another, as were the offsets where preferred bases were located. Others have reported similar results. A striking characteristic of all the consensus sequences is their symmetry,

such that the preferred sequence read from the integration site on one strand is the perfect complement of the sequence read in from the other site in the opposite direction after correcting for the 4-6 base offset between integrations on the two strands. This symmetry includes not only the preferred sequence, but also the level of preference and its statistical support). These results suggest the possibility that integration may not be concerted; rather that the complex at each end of the viral DNA may recognize an asymmetric sequence (or the structure derived from it), make the first joint, and the other end simply follow along. If the integrase-DNA complex at either end is equally able of initiating the label, then sequences of integration joints will give the symmetrical results observed.

For a number of years, our laboratory has taken another approach to analysis of integration targeting. As a model, we have been using ALV, an Alpharetrovirus, which (in other laboratories) has been found to have little, if any preference for integration into or near genes. To study integration preferences, we have developed a method for analyzing the specificity of integration. For this purpose, we used a PCR assay, in which one primer is complementary to sequences within the LTR, the other to a selected region of the target DNA. From analysis of reactions where the target was plasmid DNA, we could conclude that: (1) Integration can occur at a very large number of sites in any given fragment of DNA, but there is a characteristic and unique pattern of strong and weak sites varying in efficiency by a factor of more than 200. This pattern must be determined by local features of DNA structure, not directly by sequence or by the overall structure of the target DNA. Methylation of CpG dinucleotides does not inhibit integration, but can actually enhance it, most likely a consequence of altered structure.

We then extended the use of this analytical tool to the much more difficult case of integration in infected cells, improving the assay to detect single integration events in as much as 10 μ g of cell DNA, and locate them precisely relative to the target primer. We determined and compare the frequency and distribution of integration sites in regions of cell DNA that were chosen randomly, or were selected based on having been identified as integration targets in a previous experiment. No significant difference was observed in the distribution of target sites among all these regions, strongly implying that integration events are not confined to a small (e.g., transcriptionally active) fraction of cell DNA, but, rather,

seem to be more or less randomly distributed among all regions, a result consistent (for ALV) with more recent analyses from other groups.

We then extended this analysis to a strongly inducible, natural gene, present in single copy. For this purpose, we have chosen the metallothionein (MT) gene. Metallothionein is a protein expressed at high levels in the presence of certain kinds of stress, including divalent ions, oxidative stress, and endotoxin. It has a fairly small regulatory region that has been very well characterized for both the mouse and chicken genes, and we have cloned and sequenced the quail counterpart, and found it to be present in single copy. We have used this sequence to develop assays for MT expression. For this purpose, we are using real-time PCR, with primers designed to detect either spliced mRNA or unspliced primary transcripts. In this way, we can readily detect and quantitate (at least relatively), metallothionein mRNA levels before and after induction with 100 μM Zn^{++} . This treatment had no effect of viral DNA synthesis and integration. Since it has been reported that metal induction can also alter stability of metallothionein mRNA, we also examined induction of unspliced precursor RNA, a more direct measure of RNA synthesis. The results, show that induction of expression of MT was very rapid, with unspliced RNA levels increasing 10-fold within 15 min. and 100-fold in an hour. These high levels of induced expression were sustained for at least 4 days, well past the time frame of the integration experiment.

We then performed an experiment where we infected cells with ALV and incubated them in medium with and without 100 μM Zn^{++} . Cell DNA was prepared about 1 day after infection, and analyzed by PCR for the distribution of integration events in the MT coding region. No product was observed with uninfected cell DNA, while each reaction with DNA from infected, uninduced cells yielded about 1 band. Remarkably, induction of MT expression with zinc led to a strong reduction in the number of integrants into the MT, with only 2 bands visible in the 8 lanes. This result has proven to be reproducible in 3 experiments totaling over 70 analyses..

Thus, we can conclude that transcriptional activity is not necessarily attractive for retroviral DNA integration, and can even be inimical to it. The inhibition of integration may be a result of blockade by transcription complexes, of chromatin remodeling, of DNA melting consequent to transcription, or some combination of these effects.

