Differences between Cellular Integration Sites of Transcribed and Nontranscribed Rous Sarcoma Proviruses

VALERIE J. FINCHAM AND JOHN A. WYKE*

*Beatson Institute for Cancer Research, Cancer Research Campaign Beatson Laboratories,
Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, United Kingdom*

Received 9 July 1990/Accepted 24 September 1990

Transcribed Rous sarcoma proviruses in Rat-1 DNA tend to integrate closely 3' to C-G-rich restriction enzyme sites, and 2 of 13 such proviruses were found to have inserted at the same locus. However, most integrated proviruses were transcriptionally silent insertions at sites indistinguishable from random. We conclude that Rous sarcoma proviruses in rat cells usually fail to integrate in a favorable site for transcription, in contrast to studies on proviral integration in permissive hosts.

Proviral integration of Rous sarcoma virus (RSV) occurs with comparable efficiency in both avian and rodent cells, but in the latter only 0.1% or less of inserted proviruses are transcribed (8, 20). The low levels of proviral transcription in the cell line Rat-1 are due not to a lack of factors needed for functioning of the RSV long terminal repeat (LTR) but to an overriding effect of the site of proviral integration (1), suggesting that the predominant insertion sites in rat cells may differ from those in chick cells.

Inherent differences seem improbable if, in both hosts, proviral insertion is random or preferred targets (19) are the same, as suggested by findings that murine leukemia viruses preferentially integrate close to regions of the host genome that are DNase hypersensitive or transcribed (12, 16, 18, 22). Preferred insertions may reflect selection for proviral expression (14), but others have studied random molecular clones of proviruses, a strategy that may not, however, avoid bias against nontranscribed proviruses, since cloning of inactive RSV genomes is very inefficient in most bacterial hosts and difficult even in those lacking the *mcrA* and *mcrB* methylation restriction systems (1, 6a).

We can thus still postulate preexisting differences in the insertion sites of transcribed and nontranscribed proviruses, but it is also conceivable that in rat cells transcriptional silence is imposed on the majority after integration (24), perhaps by a mechanism that down regulates expression of foreign (26) or inappropriately transposed (13) DNA. We have examined these possibilities by comparing the cellular insertion sites of transcribed and nontranscribed proviruses in numerous RSV strain B77-infected Rat-1 clones.

Clones identified by a number alone are new isolates that are morphologically transformed and presumed to contain transcribed proviruses which, by the mapping strategy outlined in Fig. 1, are single in each clone. Clones identified by a prefix letter or letters also contain single proviruses whose transcriptional status has previously been determined by examination of proviral RNA (1, 4, 7, 8, 21, 25). Our strategy exploited the presence in RSV of CpG-containing recognition sequences for the restriction enzymes *Bss*HII, *Sac*II, and *Nar*I (Fig. 1). Sites for these enzymes in mammalian DNA are clustered in, but not confined to, the C-G-rich islands that occur, on average, about once in every 100 kb and characterize the 5' ends of housekeeping and some other

genes (11). Proximity of an inserted provirus to cellular restriction sites for one or more of these enzymes is readily detected and is presumptive evidence for integration near a C-G-rich island, a finding consistent with preferential insertion in the vicinity of transcribed DNA (12, 18). However, unlike silent RSV, many transcribed RSV proviruses in Rat-1 cells display contiguous 5' duplications of portions of the proviral genome that frequently include the *gag* region enhancer (7, 8, 10; data not shown). Such duplications may contain proviral sites for CpG-cutting enzymes (Fig. 1), which could complicate analyses of flanking sites, and we therefore present only data for clones whose proviruses lack contiguous duplications.

The proviruses in transformed lines show a remarkable trend, with 11 of 13 having at least one site for *Bss*HII or *Sac*II within 2 kb of the LTR, predominantly to the 5' side of the provirus (Fig. 2). Of the two exceptions, strain BB17 has a *Nar*I site to its 5' side, while line 93 has not been tested with this enzyme. These biased insertions of transcribed proviruses contrast markedly with observations on nine clones in which the provirus was transcriptionally silent after RSV infection (Fig. 3). All but one of these clones tested lacked detectable *Bss*HII, *Sac*II, and *Nar*I sites in the vicinity of the provirus, the exception, CC103, containing a *Bss*HII site 3.3 kb 5' to the provirus. In all of these silent genomes, the more 5' of the proviral *Bss*HII sites (Fig. 1) is undetectable, presumably as a result of cytosine methylation, but the other *Bss*HII site and the *Sac*II site remain (data not shown). This permits the conclusion that cleavable sites for these enzymes in flanking cell DNA must be at least 10 kb from the provirus or they would be detected in our analyses.

Since some C-G islands might become methylated as a result of tissue culture (2) or in response to provirus insertion (13, 24), we digested genomic DNA with the enzyme *Msp*I, which cleaves the sequence CCGG even when the internal CpG is methylated. The recognition sequence for this enzyme will certainly be present in C-G-rich islands as well as elsewhere. Nonetheless, there were no *Msp*I sites closer than about 10 kb to three of the eight silent proviruses tested (in CC22, 33, and 105), making it highly improbable that they were near a C-G-rich sequence. The *Msp*I sites in the vicinity of the other five proviruses are shown in Fig. 3. To determine whether some of these sites identified presumptive but methylated C-G islands, *Hind*III-proviral junction fragments (Fig. 1) were cloned from lines CC26, CC31, and CC17. We could clone these fragments only in the bacterial

* Corresponding author.

FIG. 1. RSV strain B77 provirus, showing the positions of restriction enzyme sites used in analysis and for defining the boundaries of probes for *gag*, *src*, and the U3 and U5 portions of the LTR. Subgenomic probes were used to investigate the number of proviruses in each clone and the presence or absence of proviral duplications (8). Molecular clones of proviral junction fragments extended from the proviral *HindIII* sites to the nearest sites in flanking host DNA. Abbreviations for restriction enzymes: B, *BamHI*; Bs, *BssHII*; E, *EcoRI*; H, *HindIII*; M, *MspI* (sites in LTR only shown); N, *NarI*; S, *SacII*; X, *XhoI*; En, Position of enhancer in the *gag* region of the genome (3, 5).

host K802 (23), which lacks the *mcrA* and *mcrB* restriction systems (15), suggesting that sequences within or flanking the provirus were methylated. This suspicion was substantiated by the observation that none of the flanking cellular *MspI* sites in genomic DNA were cleavable by *HpaII* (data not shown). However, restriction enzyme analysis of cloned DNA showed no sites for rare-cutting test enzymes in the part of the flanking DNA that was cloned (Fig. 3).

To test whether a cellular response, which is not permitted in the vicinity of C-G islands, down regulates proviral transcription, we tried to examine the methylation status in uninfected Rat-1 cells of the *MspI* sites that flank the proviruses in CC17, CC26, and CC31. In the case of CC26, specific probes demonstrated that all four flanking *MspI* sites (Fig. 3) were also methylated in uninfected Rat-1 DNA. Thus, this virus has apparently integrated into a relatively methylated region of the host genome. However, our attempts to produce probes specific for the host insertion site in the molecular clones CC17 and CC31 were impeded by the presence of repetitive sequences, a phenomenon that we have never encountered in sequences flanking transcribed proviruses, further implying some qualitative difference between the two proviral classes. It is interesting that repetitive sequences of different types have been implicated in

FIG. 2. Sites for C-G-cutting restriction enzymes in the host DNA flanking transcribed proviruses. The provirus site is represented as a single LTR (■). Breaks in the flanking host DNA (line 93) indicate that no sites have been found in approximately 10 kb or more of flanking DNA. Abbreviations are as for Fig. 1.

FIG. 3. Sites for C-G-cutting restriction enzymes in the host DNA flanking silent proviruses. *, Clone produces transformed segregants upon prolonged culture (8). Dashed lines under CC17, CC26, and CC31 indicate regions of DNA that have been cloned. Other symbols and abbreviations are as for Fig. 1 and 2.

transcriptional silencing of adjacent genes (9, 17), including an endogenous mouse mammary tumor provirus (6).

We have previously shown that when only 1 in 2×10^3 infected Rat-1 cells becomes transformed by RSV, 35 of 95 randomly cloned infected cells harbor stable transcriptionally silent proviruses (8). Thus, only about 0.1% of integrated proviral genomes are transcribed, and it is only this population that shows the biased insertion sites highlighted in other studies (Fig. 2; 12, 16, 18, 22). Given the expected distribution of C-G-rich regions in the host genome, the integration patterns of the vast silent majority (Fig. 3) are consistent with insertion at random. If, however, integration within a few kilobases of a C-G-rich island is enough to ensure expression, then far more than 0.1% of randomly integrated RSV proviruses in Rat-1 cells should be transcribed. A significant minority is thus likely to be silent for reasons other than distance from a C-G-rich island. Evidence for a favored subset of C-G-rich regions in Rat-1 cells comes from the finding that the active proviruses in two separate clones had inserted in the same orientation in a unique site within 2 kb of one another (Fig. 4). This site contains a characteristic C-G-rich island (Fig. 4; sequence data not shown), but since there are more than 10^4 such islands in the haploid genome, its proximity to two proviruses is a remarkable coincidence.

Our data thus suggest that there is a hierarchy of RSV insertion sites in Rat-1 cells, but we do not know its basis. Investigation of a preferred insertion site (Fig. 4) may reveal elements that promote proviral expression. On the other hand, although we have failed to find evidence for site-specific down regulation of proviral transcription, such ideas could be discarded only by laborious examination of numerous silent proviruses.

Moreover, these ambiguities apply only to a minority of inserted proviruses. It is clear that the insertion sites of the silent majority are inherently different from those reported in the other studies. The mechanism by which a provirus inserts in a particular region of chromatin is thus influenced by host cell factors that are species or cell type specific, and

FIG. 4. Cellular insertion sites of the proviruses in clones BB16 and B31. A partial restriction map of the insertion sites is shown, the enzymes below the line being those that have C-G-rich restriction sites. The location and orientations of the two proviruses are shown at the top; the dashed line at the bottom denotes the region that has been sequenced. The divisions on this line are each of 200 bp. Abbreviation for restriction enzymes: Na, *NarI*; Bg, *BglII*; Ea, *EagI*; K, *KpnI*; Ne, *NaeI*; Sm, *SmaI*; St, *StuI*. Other abbreviations are as for Fig. 1.

this phenomenon could have profound implications for retrovirus biology and pathogenesis.

We are grateful to Bob Brown for suggesting the use of restriction enzymes with C-G-containing recognition sequences, for telling us of his unpublished data, and for commenting on the manuscript.

This work was supported by the Cancer Research Campaign.

REFERENCES

1. Akroyd, J., V. J. Fincham, A. R. Green, P. Levantis, S. Searle, and J. A. Wyke. 1987. Transcription of Rous sarcoma proviruses in rat cells is determined by chromosomal position effects that fluctuate and can operate over long distances. *Oncogene* 1:347-354.
2. Antequera, F., J. Boyes, and A. Bird. 1990. High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell* 62:503-514.
3. Arrigo, S., M. Yun, and K. Beemon. 1987. *cis*-Acting regulatory elements within *gag* genes of avian retroviruses. *Mol. Cell. Biol.* 7:388-397.
4. Chiswell, D. J., P. J. Enrietto, S. Evans, K. Quade, and J. A. Wyke. 1982. Molecular mechanisms involved in morphological variation of avian sarcoma virus-infected rat cells. *Virology* 116:428-440.
5. Cullen, B. R., R. Kyle, and G. Ju. 1985. Functional analysis of the transcription control region located within the avian retroviral long terminal repeat. *Mol. Cell. Biol.* 5:438-447.
6. Fanning, T. G., D. W. Morris, R. D. Cardiff, and H. D. Bradshaw, Jr. 1985. Characterization of an endogenous retrovirus-repetitive DNA chimera in the mouse genome. *J. Virol.* 53:998-1000.
- 6a. Fincham, V. J., and D. A. F. Gillespie. Unpublished data.
7. Gillespie, D. A. F., K. A. Hart, and J. A. Wyke. 1985. Rearrangements of viral and cellular DNA are often associated with expression of Rous sarcoma virus in rat cells. *Cell* 41:279-287.
8. Green, A. R., S. Searle, D. A. F. Gillespie, M. Bissell, and J. A. Wyke. 1986. Expression of integrated Rous sarcoma viruses: DNA rearrangements 5' to the provirus are common in transformed rat cells but not seen in infected but untransformed cells. *EMBO J.* 5:707-711.
9. Laimins, L., M. Holmgren-Konig, and G. Khoury. 1986. Transcriptional "silencer" element in rat repetitive sequences associated with the rat insulin 1 gene locus. *Proc. Natl. Acad. Sci. USA* 83:3151-3155.
10. Levantis, P., D. A. F. Gillespie, K. Hart, M. J. Bissell, and J. A. Wyke. 1986. Control of expression of an integrated Rous sarcoma provirus in rat cells: role of 5' genomic duplications reveals unexpected patterns of gene transcription and its regulation. *J. Virol.* 57:907-916.
11. Lindsay, S., and A. P. Bird. 1987. Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature (London)* 327:336-338.
12. Mooslehner, K., U. Karls, and K. Harbers. 1990. Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA regions. *J. Virol.* 64:3056-3058.
13. Nur, I., E. Pascale, and A. V. Furano. 1988. The left end of rat L1 (L1Rn, long interspersed repeat) DNA which is a CpG island can function as a promoter. *Nucleic Acids Res.* 16:9233-9251.
14. Peckham, I., S. Sobel, J. Comer, R. Jaenisch, and E. Barklis. 1989. Retrovirus activation in embryonal carcinoma cells by cellular promoters. *Genes Dev.* 3:2062-2071.
15. Raleigh, E. A., and G. Wilson. 1986. *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl. Acad. Sci. USA* 83:9070-9074.
16. Rohdewold, H., H. Weiher, W. Reik, R. Jaenisch, and M. Breindl. 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J. Virol.* 61:336-343.
17. Saffer, J. D., and S. J. Thurston. 1989. A negative regulatory element with properties similar to those of enhancers is contained within an *Alu* sequence. *Mol. Cell. Biol.* 9:355-364.
18. Scherdin, U., K. Rhodes, and M. Breindl. 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* 64:907-912.
19. Shih, C.-C., J. P. Stoye, and J. M. Coffin. 1988. Highly preferred targets for retrovirus integration. *Cell* 53:531-537.
20. Varmus, H. E., R. V. Guntaka, C.-T. Deng, and J. M. Bishop. 1975. Synthesis, structure and function of avian sarcoma virus-specific DNA in permissive and non-permissive cells. Cold Spring Harbor Symp. Quant. Biol. 39:987-996.
21. Varmus, H. E., N. Quintrell, and J. Wyke. 1981. Revertants of an ASV-transformed rat cell line have lost the complete provirus or sustained mutations in *src*. *Virology* 108:25-46.
22. Vijaya, S., D. L. Steffen, and H. L. Robinson. 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J. Virol.* 60:683-692.
23. Wood, W. B. 1966. Host specificity of DNA produced by *Escherichia coli*: bacterial mutations affecting the restriction and modification of DNA. *J. Mol. Biol.* 16:118-133.
24. Wyke, J. A., J. Akroyd, D. A. F. Gillespie, A. R. Green, and C. J. Poole. 1989. Proviral position effects: possible probes for genes that suppress transcription. CIBA Found. Symp. 142: 117-130.
25. Wyke, J. A., and K. Quade. 1980. Infection of rat cells by avian sarcoma virus: factors affecting transformation and subsequent reversion. *Virology* 106:217-233.
26. zur Hausen, H. 1987. Papillomaviruses in human cancer. *Cancer* 59:1692-1696.