JOURNAL OF VIROLOGY, Nov. 1986, p. 683–692
0022-538X/86/110683-10$02.00/0

Vol. 60, No. 2

Acceptor Sites for Retroviral Integrations Map Near DNase I-Hypersensitive Sites in Chromatin

S. VIJAYA, D. L. STEFFEN, AND H. L. ROBINSON*

Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts 01545

Received 16 June 1986/Accepted 3 July 1986

Seven cellular loci with acceptor sites for retroviral integrations have been mapped for the presence of DNase I-hypersensitive sites in chromatin. Integrations in three of these loci, chicken c-*erb*B, rat c-*myc*, and a rat locus, *dsi*-1, had been selected for in retrovirus-induced tumors. Of the remaining four, two, designated *dsi*-3 and *dsi*-4, harbored acceptor sites for apparently unselected integrations of Moloney murine leukemia virus in a Moloney murine leukemia virus-induced thymoma, and two, designated C and F, harbored unselected acceptor sites for Moloney murine leukemia virus integrations in a rat fibroblast cell line. Each acceptor site mapped to within 500 base pairs of a DNase I-hypersensitive site. In the analyses of the unselected integrations, six hypersensitive sites were observed in 39 kilobases of DNA. The four acceptor sites in this DNA were localized between 0.05 and 0.43 kilobases of a hypersensitive site. The probability of this close association occurring by chance was calculated to be extremely low. Hypersensitive sites were mapped in cells representing the lineage in which integration had occurred as well as in an unrelated lineage. In six of the seven acceptor loci hypersensitive sites could not be detected in the unrelated lineage. Our results indicate that retroviruses preferentially integrate close to DNase I-hypersensitive sites and that many of these sites are expressed in some but not all cells.

A central feature of the retrovirus lifecycle is the ordered integration of the DNA form of the viral genome into the host genome. Integration is accomplished by recombination between viral long terminal repeat and host sequences. An early step in integration, endonucleolytic cleavage of long terminal repeat and host sequences, appears to be catalyzed by the product of the viral *pol* gene. This is evidenced by in vitro cleavage of the preintegration form of viral DNA near the junction of the long terminal repeats by purified reverse transcriptase (8). Recombination between viral and host sequences occurs at a specific position within the long terminal repeat, generally 2 base pairs (bp) internal to the termini (14, 18, 26). In contrast, host acceptor sites for integrations appear to be at least locally nonspecific, with analyses of host-virus junctions failing to reveal a consensus acceptor sequence (for review, see reference 32). Host sequences are duplicated at the site of integration, with the number of duplicated bases being characteristic of the virus (6, 12, 18, 25). The replication of chromosomal DNA appears to be a necessary requirement for integration, with proviruses undergoing insertion into newly synthesised DNA (33).

Despite the apparent absence of sequence specificity for acceptor sites, chromatin structure may determine regional specificity of integration. DNase I and S1 digestions of chromatin reveal short regions of DNA which are hypersensitive to digestion (for reviews, see references 19, 35). DNase I-hypersensitive sites are thought to represent regions of chromatin which may be preferentially available for the entry of proteins that effect the replication, transcription, and rearrangement of DNA (19, 29, 40). These regions may also be preferentially available for the integration of retroviral DNA. In bursal lymphomas, avian leukosis virus integrations map near each of the five major hypersensitive sites which are immediately upstream of the coding sequences for c-*myc* (23, 24). The integration of a Moloney murine leukemia virus (MoMLV) provirus in the alpha 1 collagen gene has also been mapped near a DNase I-hypersensitive site (4).

To determine whether the occurrence of retroviral integrations near DNase I-hypersensitive sites is a general phenomenon, we have mapped a series of acceptor sites for their proximity to DNase I-hypersensitive sites. Three of the loci were targets for insertions that had been selected for in retrovirus-induced tumors (20, 27; this paper), while four were targets for apparently unselected integrations (this paper; 13). Our results indicate that both selected and unselected integrations occur near DNase I-hypersensitive sites.

MATERIALS AND METHODS

Cell lines. The RCC2-1M cell lines was established from a MoMLV-induced rat thymoma. This cell line harbors more than 10 copies of MoMLV proviral DNA. None of the insertions were in the *dsi*-1, *dsi*-3, *dsi*-4, or c-*myc* locus. RCC2-1M cells were grown in RPMI 1640 medium supplemented with 10% heat-inactivated fetal calf serum and 5×10^{-5} M 2-mercaptoethanol. NRK, a continuous line of rat kidney fibroblast cells, was grown in Dulbecco modified Eagle medium supplemented with 10% calf serum. HD3 is an avian erythroblast cell line transformed by avian erythroblastosis virus strain R (AEV-R) (3). HD3 cells, a kind gift of T. Graf and H. Beug, were grown in Dulbecco modified Eagle medium supplemented with 8% calf serum and 2% heat-inactivated chicken serum.

Tissues. Erythroblastosis was induced in birds by injecting 1-week-old chickens intravenously with 5×10^4 U of AEV-R. Onset of erythroblastosis was monitored by determining the number of erythroblasts in peripheral blood. Bone marrow from femur and tibia of leukemic birds was collected; the nuclei were isolated and digested with DNase I as described below. Samples of DNase I-digested DNA from 14-day embryo erythrocytes, anemic adult erythrocytes, thymus, and bursa were kindly provided by Mark Groudine.

* Corresponding author.

Purification and DNase I digestion of nuclei. Nuclei were isolated from subconfluent cells. Pelleted cells were washed with ice-cold phosphate-buffered saline and suspended in reticulocyte standard buffer (10 mM Tris, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$), and the plasma membranes were lysed by addition of 0.5% Nonidet P-40. The nuclei were pelleted by centrifugation in an IEC centrifuge for 5 min at 1,500 rpm. Nuclei from tissues were isolated by first homogenizing the tissues in a Dounce homogenizer followed by sedimentation for 5 min at 2,000 rpm. The pellet was suspended in reticulocyte standard buffer and treated with 0.5% Nonidet P-40. The lysed nuclei were pelleted through a 1.7 M sucrose cushion for 15 min at 13,000 rpm in an HB-4 rotor of a Sorvall refrigerated centrifuge.

The nuclear pellet was suspended in reticulocyte standard buffer to give approximately 1 mg of DNA per ml. The nuclear suspension was allowed to stand for 3 min to allow partially disrupted nuclei to aggregate and settle. The supernatant nuclei were transferred to a fresh tube, and 200-μl aliquots were distributed into Eppendorf vials. The nuclei in the control tube were immediately lysed with DNase I stop buffer (20 mM Tris, pH 7.4, 1% sodium dodecyl sulfate, 0.6 M NaCl, 20 mM EDTA), while the nuclei in the other tubes received concentrations of DNase I (Worthington Diagnostics) ranging from 0.4 to 25.6 μg/ml in a constant volume of 10 μl. The digestion was carried out for 10 min at 37°C and the reactions were terminated by addition of 200 μl of stop buffer. The samples were then digested with 200 μg of proteinase K for 4 h, and DNA from the samples was isolated by extraction with phenol and chloroform followed by precipitation with alcohol as described before (20).

Chromosomal sequences with acceptor sites for integrations. Five proviruses with flanking host sequences were molecularly cloned from RT10-2, a MoMLV-induced thymoma which had five integrated proviruses. One of these integrations was in the c-*myc* locus (27; D. L. Steffen and E. Q. Nacar, manuscript in preparation). The remaining proviruses were in loci designated *dsi*-1 to -4. Loci *dsi*-1 and *dsi*-2 were each isolated as two *Hin*dIII fragments in lambda Charon 4A. Locus *dsi*-3 was isolated as a single *Eco*RI fragment in lambda EMBL 4. Locus *dsi*-4 was isolated both as an *Eco*RI fragment in lambda Charon 4A and as two *Hin*dIII fragments, the one on the 3′ end of the virus inserted into lambda Charon 21A and the one on the 5′ end inserted into lambda Charon 4A. Insertion of *Hin*dIII fragments into lambda Charon 4A was accomplished by using oligonucleotide restriction site adapters obtained from Collaborative Research. The *Hin*dIII clones were associated with the appropriate integration sites by hybridizing unique rat sequence probes from each clone to *Eco*RI-digested RT10-2 DNA. *dsi*-2 was not used in this study since the insertion in *dsi*-2 appears to be associated with a rearrangement in flanking host sequences.

Two proviruses with flanking host sequences were molecularly cloned (13) from MoMLV-infected NRK cells. These clones were designated lambda 836 and lambda 21. The rat loci that harbor these two proviruses are called C and F, respectively. Lambda c-*erb*1, a clone containing the 5′ end of the c-*erb*B gene of chickens (34), was a kind gift of B. Vennstrom. The mouse *myc* clone S107 was provided by P. Leder (15).

Molecularly cloned DNAs were tested for the presence of repetitive sequences by hybridizing blots of restricted DNA with nick-translated rat or chicken DNA (specific activity, $>10^7$ cpm/μg of DNA). Fragments free of repeat sequences which could be used as probes for one end of a fragment being analyzed for hypersensitive sites were subcloned into pUC12 or pUC18. Use of hybridization probes for one end of a fragment facilitates the positioning of sites within the fragment (38). For c-*myc* and *dsi*-3, the positions of repeat sequences and restriction sites necessitated the use of probes that did not hybridize with the immediate ends of the fragments being tested for hypersensitive sites. In these two cases the positions of hypersensitive sites were confirmed by additional digestions. Restriction endonuclease maps of chicken c-*erb*B sequences and rat c-*myc* sequences are according to the transcriptional orientation of c-*erb*B and c-*myc*, respectively. Restriction endonuclease maps of the other loci are oriented according to the transcriptional sense of integrated proviruses.

Southern blots. Restriction enzyme digestions, separation of digested DNA on agarose gels, transfer to nitrocellulose filters (Schleicher & Schuell BA 85), and hybridization to nick-translated probes were as described previously (20).

Sizing of fragments. Fragment sizes were routinely estimated from the positions of fragments relative to those of *Hin*dIII-digested lambda DNA. In some blots ^{32}P-end-labeled 1-kilobase (kb) ladder fragments (Bethesda Research Laboratories) were used in sizing. Where possible, the positions of cellular fragments that had been determined from molecularly cloned DNAs were used in the construction of sizing curves. The sizes of fragments used to position hypersensitive sites are averages of independent analyses of different DNase I-treated DNAs. The sizing of these fragments, which are typically represented by broad and fuzzy bands, was done by measuring to a point about one-third of the way from the bottom of the band. Independent sizings were always within 10% of each other (a difference which would represent a 1- to 2-mm error in measurement on an autoradiograph of a typical Southern blot). Thus the position of a hypersensitive site represented by a 10-kb fragment is accurate to within ±500 bp, whereas the position of a hypersensitive site represented by a 1-kb fragment is accurate to within ±50 bp. The positioning of proviral insertions was facilitated by the sizing of a series of junction fragments with defined ends in proviral sequences. The positions of one insertion each in rat c-*myc* and *dsi*-1 as well as the positions of each of the unselected insertions were determined with molecularly cloned DNA. We consider the positions of these insertions to be accurate to within 100 to 200 bp. Positioning of hypersensitive sites and integration sites was carried out independently so as not to bias the determinations.

Statistical analysis. Statistical analysis of the data in Table 1 was performed by Michael Sutherland, University of Massachusetts, Amherst. For each of the four unselected integrations, it was assumed that the positions of proviral integration and DNase I-hypersensitive sites were uniformly distributed over the region of DNA analyzed. Given that assumption, it was then possible to calculate the probability that the distance between these sites would be less than or equal to what was observed (37). Having computed each of these four probabilites, the probability of the aggregate data being due to chance was the product of the four. The sensitivity of this calculation to errors in the measurement of DNA fragment size was determined by increasing the distance between the integration and hypersensitive sites and repeating the calculation.

RESULTS

Erythroblastosis-inducing insertions in the chicken epidermal growth factor receptor gene. Insertions of avian leukosis

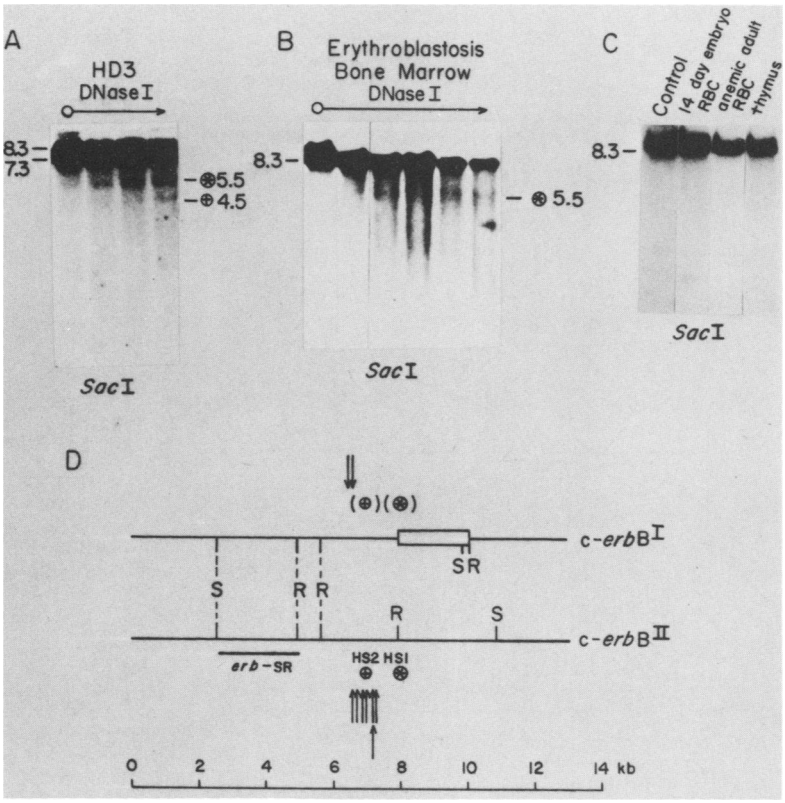

FIG. 1. DNase I-hypersensitive sites near selected integrations in the chicken c-*erb*B gene. (A–C) Tests for hypersensitive sites in HD3 cells (A), in a primary case of erythroblastosis (B), and in several normal tissues (C). Autoradiographs are of Southern blots of *Sac*I-digested DNAs hybridized with *erb*-SR sequences. In each panel, the extreme left lane is DNA that was not digested with DNase I. In (A) and (B), the amount of DNase I used for the generation of hypersensitive sites increases as the panels progress from left to right. In (C), each of the digestions represents a point from a series of digestions. These points displayed clear hypersensitive sites in c-*myc*. (D) Restriction endonuclease map of the region of the chicken epidermal growth factor receptor gene that is a target for erythroblastosis-inducing insertions. The open box in c-*erb*B^{I} is a sequence unique to the B^{I} allele. The vertical arrows denote the positions of erythroblastosis-inducing proviral insertions (20). The closed bar represents the *erb*-SR probe. RBC, Erythrocytes; R, *Eco*RI; S, *Sac*I. Symbols: *, prominent hypersensitive site; +, less prominent hypersensitive site. Numbers give DNA fragment sizes in kilobases.

virus into the c-*erb*B region of the gene for the epidermal growth factor receptor (7, 17, 21, 31) cause erythroblastosis in line 15_1 chickens (11, 20, 22). These insertions occur in a <1.5-kb region immediately upstream of sequences encoding the transmembrane domain of the receptor (20, 22). DNase I-hypersensitive sites were mapped in this region in a line of erythroblasts (HD3) transformed by AEV-R. This cell line contained two normal alleles of c-*erb*B, c-*erb*B^{I} and c-*erb*B^{II} (11, 20).

Hypersensitive sites were identifed by hybridizing Southern blots of *Sac*I-digested DNAs with a probe for the 5′ ends of the fragments that are acceptors for erythroblastosis-inducing insertions (Fig. 1D). This probe, erb-SR, detects chromosomal, but not viral, *erb*B sequences. Southern blots of HD3 chromatin that had not been digested with DNase I displayed only the 7.3- and 8.3-kb *Sac*I fragments of c-*erb*B^{I} and c-*erb*B^{II}, respectively, while those of chromatin that had been digested with increasing concentrations of DNase I revealed decreasing amounts of the 7.3- and 8.3-kb fragments and increasing amounts of a prominent 5.5-kb and a less prominent 4.5-kb fragment (Fig. 1A). Based on the sizes of these two novel fragments, candidate hypersensitive sites were positioned 5.5 (HS1) and 4.5 (HS2) kb from the 5′ ends of the acceptor *Sac*I fragments in c-*erb*B^{I} and c-*erb*B^{II} (Fig. 1D). Verification that both HS1 and HS2 occur in c-*erb*B^{II} was obtained by analyzing bone marrow from chickens with AEV-R-induced erythroblastosis. These chickens were homozygous for c-*erb*B^{II}. Both HS1 and HS2 were detected in leukemic marrows (Fig. 1B), with the less prominent HS2 being observed in some but not all marrows. Each of seven erythroblastosis-inducing avian leukosis virus insertions in c-*erb*B^{II} (20) mapped within 500 nucleotides from HS2 (Fig. 1D).

HS1 and HS2 were not detected in DNase I-treated chromatin of 10-day-old chicken embryos, bursa or thymus of young chicks, and erythrocytes from 14-day embryos or anemic adults (Fig. 1C and data not shown). Thus, it appears that the c-*erb*B-hypersensitive sites are not expressed in all cells. The significance of these hypesensitive sites in different cell types is currently under investigation.

Thymoma-associated insertion in rat c-*myc*. From 10 to 20% of thymic lymphomas induced in rats in MoMLV have MoMLV integrations near the 5′ end of the rat c-*myc* gene (27, 30). The normal as well as target c-*myc* alleles from one of these thymomas were molecularly cloned and mapped for restriction endonuclease sites (Steffen and Nacar, in preparation). These analyses as well as Southern blot analyses of thymoma DNAs positioned four thymoma-associated insertions 5′ to the coding sequences for c-*myc*.

Chromatin of a line of T cells (RCC2-1M) established from a MoMLV-induced thymoma was analyzed for the presence of hypersensitive sites near the 5′ end of c-*myc*. This line

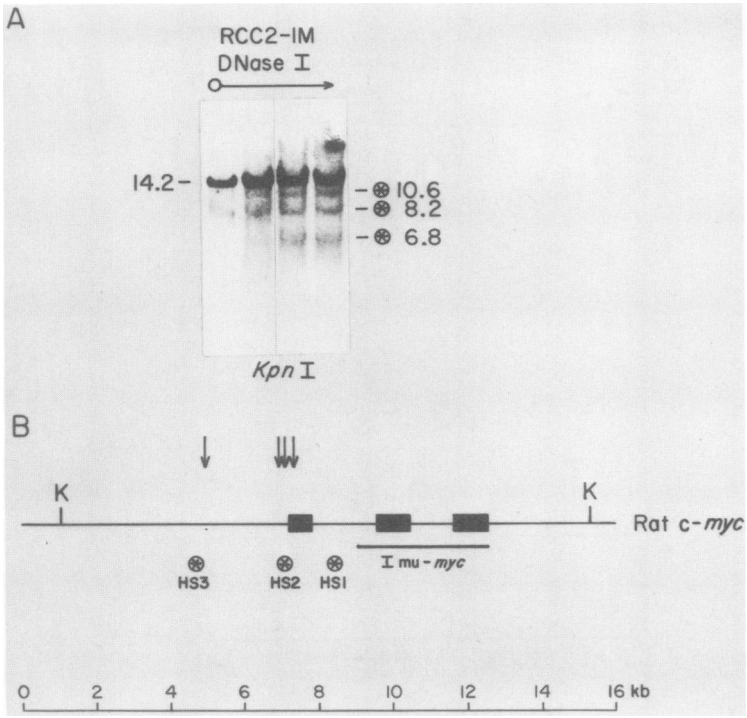

FIG. 2. DNase I-hypersensitive sites near selected integrations in the rat c-*myc* gene. (A) Autoradiograph of a blot of *Kpn*I-digested DNA isolated from DNase I-digested nuclei from RCC2-1M cells hybridized with the probe I mu-*myc*. (B) Restriction endonuclease map of the rat c-*myc* gene. Designations are as in Fig. 1. K, *Kpn*I.

does not carry an insertion in c-*myc*. Southern blots of *Kpn*I-digested DNAs were hybridized with a probe (I mu-*myc*) for unique sequences towards the 3′ end of c-*myc* (Fig. 2B). DNA from chromatin that had not been digested with DNase I gave rise to a single 14.2-kb fragment, whereas DNA from chromatin digested with increasing concentrations of DNase I revealed additional 10.6-, 8.2-, and 6.8-kb fragments (Fig. 2A). The sizes of these additional fragments positioned three hypersensitive sites 5′ to the coding sequences of c-*myc*. HS1, represented by the 6.8-kb fragment, mapped within the first intron; HS2, represented by the 8.2-kb fragment, mapped to immediately 5′ of the noncoding exon; and HS3, represented by the 10.6-kb fragment, mapped in upstream sequences. The positions of these sites were confirmed by the analysis of *Eco*RI-digested DNAs (27).

The thymoma-associated insertions in three tumors mapped approximately 50, 100, and 300 bp from HS2, while the insertion in a fourth mapped about 300 bp from HS3. The rat c-*myc*-hypersensitive sites were also present in chromatin from a line of fibroblasts established from a normal rat kidney (NRK cells).

Thymoma-associated insertion in rat *dsi*-1. Recently we have observed that 3 of 24 MoMLV-induced thymomas have insertions in a locus designated *dsi*-1 (unpublished data). The restriction map and the location of highly repeated sequences in *dsi*-1 were deduced from molecularly cloned fragments which contained the insertion in *dsi*-1 as well as flanking 5′ (lambda-5′*ds*1) and 3′ (lambda-3′*ds*1) sequences. The positions of three apparently complete proviruses in *dsi*-1 were determined from the sizes of diagnostic restriction enzyme fragments in the molecularly cloned DNA as well as in Southern blots of tumor DNAs. DNase I-digested chromatin from the T-cell line RCC2-1M was analyzed for hypersensitive sites in *dsi*-1 by hybridizing Southern blots of *Eco*RI- plus *Hin*dIII-digested DNA with a probe for unique sequences (*ds*1-SR) at the 3′ end of the fragment that was the acceptor for the three insertions (Fig. 3D). RCC2-1M cells do not have a proviral insertion in *dsi*-1. A cluster of four prominent hypersensitive sites was observed (Fig. 3A). The three lymphoma-associated insertions mapped within 300 bp of two of the hypersensitive sites in this cluster (Fig. 3D). Hypersensitive sites further downstream of the lymphoma-associated insertions were mapped by using Southern blots of *Bgl*II-digested DNAs. These experiments revealed two hypersensitive sites about 4 kb downstream of the cluster in which proviral insertions had occurred (Fig. 3B). Interestingly, DNase I digestion of the chromatin of NRK cells did not reveal hypersensitive sites in the *Eco*RI to *Hin*dIII fragment (Fig. 3C), indicating that the hypersensitive sites near this insertion are not expressed in all cells.

Origin of unselected insertions. To determine whether unselected insertions also occur near DNase I-hypersensitive sites, sequences flanking two apparently unselected insertions in a MoMLV-induced thymoma as well as sequences flanking two unselected insertions in a MoMLV-infected NRK cell were mapped for their proximity to DNase I-hypersensitive sites. The two loci with apparently unselected proviruses in the MoMLV-induced thymoma are designated *dsi*-3 and *dsi*-4. There are two lines of evidence that suggest that the insertions in *dsi*-3 and *dsi*-4 were unselected. The first was obtained by testing for insertions in these loci in 24 MoMLV-induced thymomas. Unlike *dsi*-1, which had undergone insertions in two additional tumors, none of these 24 thymomas revealed an insertion in *dsi*-3 or *dsi*-4. Second, the insertions in *dsi*-3 and *dsi*-4 were cloned from a tumor RT10-2, which contained selected integrations in the proto-oncogene c-*myc* and in *dsi*-1. Since the vast majority of MoMLV insertions do not alter the growth potential of cells, the likelihood that every single one of the

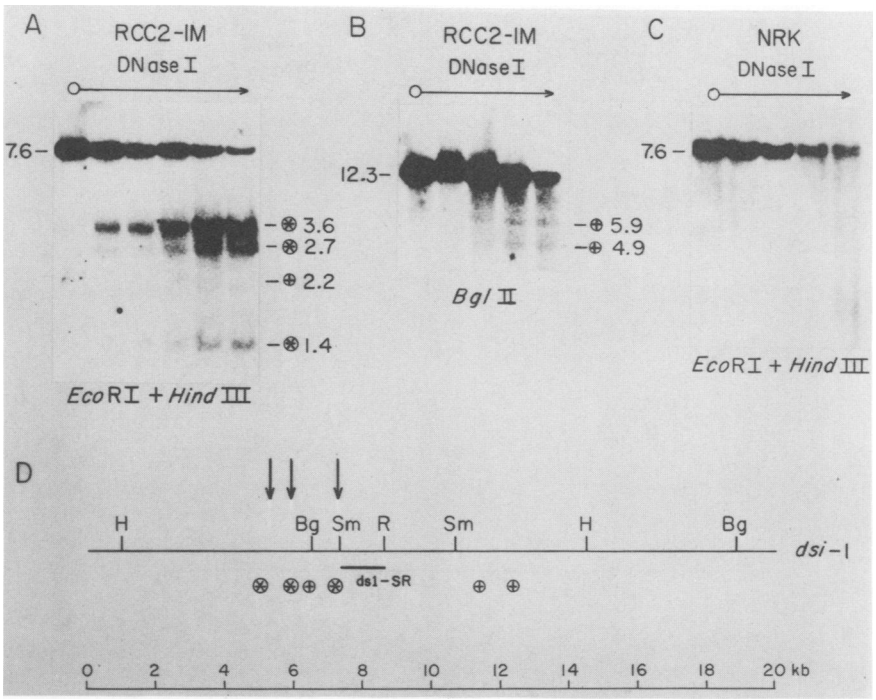

FIG. 3. Hypersensitive sites near the selected integrations in locus *dsi*-1. RCC2-1M nuclei were DNase I digested, and Soutern blots of *Eco*RI plus *Hin*dIII (A) or *Bgl*II (B) fragments were hybridized with probe *ds*1-SR. (C) Same analysis as in (A) for DNase I-digested DNA of NRK cells. (D) Restriction endonuclease map of *dsi*-1. Designations are as in Fig. 1. Sm, *Sma*I.

insertions in tumor RT10-2 was tumor inducing is highly unlikely. However, even if one of these two loci later turned out to be selected, this would not negate the significance of our results (see section on statistical analysis below).

The loci which harbor the unselected insertions in NRK cells are designated C and F (13). These insertions were molecularly cloned from a line of NRK cells (designated NRK-5) whose founder was cloned immediately following infection with MoMLV at a multiplicity of infection of ~1 (28). Since MoMLV infections do not alter the plating efficiency or the phenotype of NRK cells, and since, in particular, the growth properties of NRK-5 cells are similar to NRK cells, it is highly unlikely that insertions in C and F represent a growth-altering mutation in a proto-oncogene. The provirus at C is not expressed in NRK-5 cells, whereas the provirus at F presumably is expressed (13).

Molecular clones of DNA flanking each of these insertions were mapped for the presence of restriction endonuclease sites and repeated DNA sequences. Appropriate restriction endonuclease sites and unique sequence probes were then used to demonstrate that the insertions were not associated with discernible rearrangements in host sequences.

Unselected insertions map near DNase I-hypersensitive sites. Cells representing the lineages in which the unselected insertions had occurred were analyzed for DNase I-hypersensitive sites near the sites of insertion. The RCC2-1M line of T cells was used to represent the T cell in which the insertions in the thymoma RT10-2 had occurred since the subclass of T cells in the thymus that gave rise to the thymoma is not known. RCC2-1M cells do not contain an insertion in *dsi*-3, *dsi*-4, C, or F. NRK cells were used to represent the lineage in which the insertions in an NRK cell had occurred.

A single hypersensitive site was observed in *dsi*-3 in the T-cell line RCC2-1M. This site was revealed by hybridizing Southern blots of *Bgl*II-digested DNA with probe *ds*3-XR for unique sequences towards the 3′ end of the 12.8-kb acceptor *Bgl*II fragment (Fig. 4C). Blots of DNase I-digested DNA revealed the 12.8-kb acceptor fragment as well as a novel 7.5-kb fragment (Fig. 4A) which positioned a hypersensitive site approximately 150 bp from the insertion (Fig. 4C). This hypersensitive site was detected in blots of two independent preparations of DNase I-digested DNA. DNase I-treated chromatin from NRK cells did not reveal the presence of this or other hypersensitive sites in *dsi*-3 (Fig. 4B). The position of this hypersensitive site was confirmed by using an *Eco*RI digestion.

Analysis of *dsi*-4 showed the presence of two distinct hypersensitive sites in DNase I-digested chromatin from the T-cell line RCC2-1M (Fig. 5A and C). The provirus in *dsi*-4 in the tumor RT10-2 had integrated approximately 400 bp away from the hypersensitive site represented by the novel 3.4-kb fragment (Fig. 5A and C). The sizing of the fragment that positioned this site ranged from 3.4 to 3.6 kb in independent analyses of independent preparations of DNase I-treated DNA. Neither of the two hypersensitive sites in *dsi*-4 was detected in the chromatin of NRK cells (Fig. 5B). Thus, hypersensitive sites in *dsi*-3 and *dsi*-4 are present in a line of T cells but not in fibroblasts. The same preparation of NRK cells revealed the presence of the hypersensitive sites in c-*myc*, as well as in loci C and F (see below), indicating that the failure to see the hypersensitive sites in *dsi*-3 and *dsi*-4 was not due to technical problems related to the preparation of chromatin from these cells.

Analyses for hypersensitive sites in rat locus C (an acceptor for an unselected integration in NRK cells) revealed a single hypersensitive site in the chromatin of NRK cells (Fig. 6A). This hypersensitive site mapped ~50 bp from the integration in C (Fig. 6C). Two independent analyses of independent preparations of DNase I-treated chromatin po-

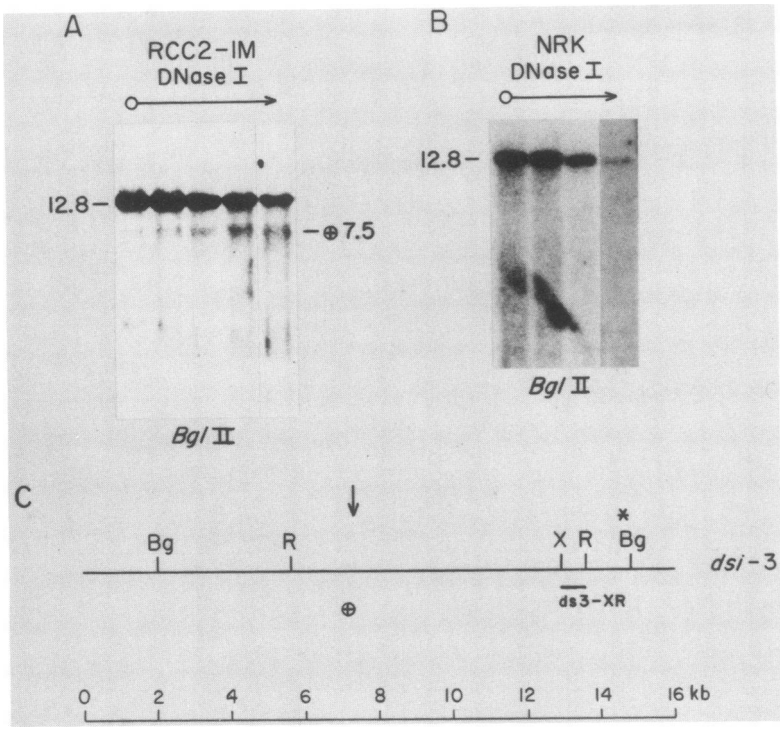

FIG. 4. Hypersensitive sites near the unselected integration in locus *dsi*-3. Autoradiographs of *Bgl*II-digested DNA from DNase I-treated nuclei of RCC2-1M (A) or NRK (B) cells hybridized with probe *ds*3-XR. (C) Restriction endonuclease map of *dsi*-3. Designations are as in Fig. 1. X, *Xho*I.

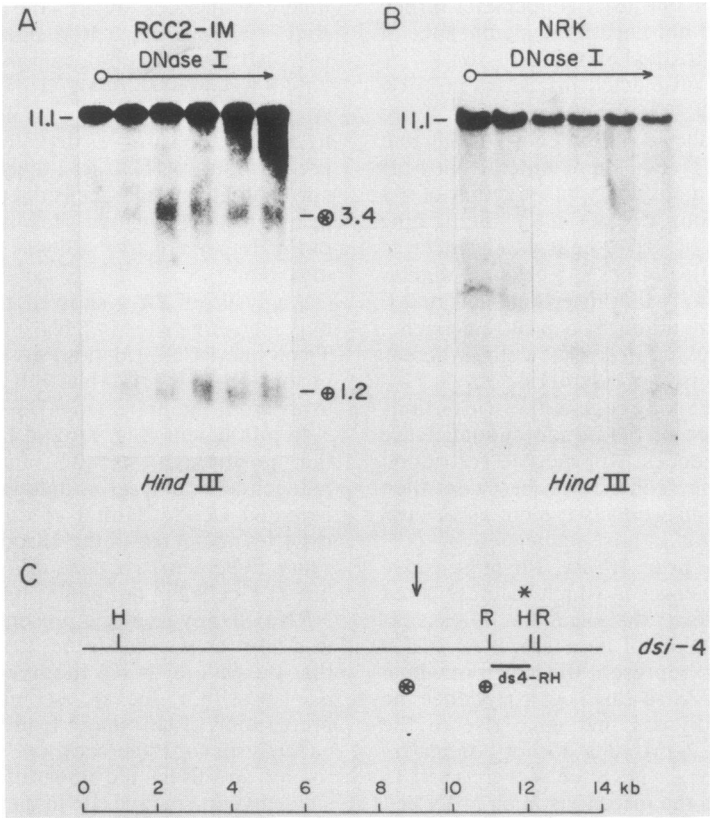

FIG. 5. Hypersensitive sites near the unselected integration in locus *dsi*-4. Autoradiographs of *Hin*dIII-digested DNA from DNase I-treated nuclei of RCC2-1M (A) or NRK (B) cells hybridized with probe *ds*4-RH. (C) Restriction endonuclease map of *dsi*-4. Designations are as in Fig. 1. H, *Hin*dIII.

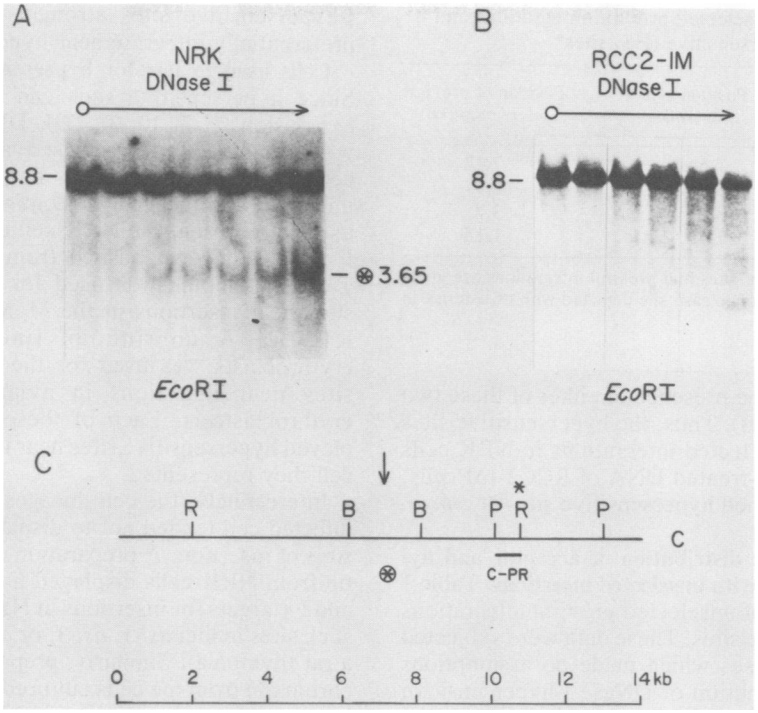

FIG. 6. Hypersensitive sites near the unselected integration in locus C. Autoradiographs of *Eco*RI-digested DNA from DNase I-treated nuclei of NRK (A) or RCC2-1M (B) cells hybridized with probe C-PR. (C) Restriction endonuclease map of C. Designations are as in Fig. 1.

sitioned this hypersensitive site within a 250-bp range. Once again, this hypersensitive site could not be detected in the chromatin of RCC2-1M cells (Fig. 6B).

The restriction map of locus F is shown in Fig. 7C. Southern blots of *Sac*I-digested chromatin from NRK cells revealed novel 2.2- and 1.45-kb fragments in DNase I-digested DNA (Fig. 6A). The acceptor site for the insertion in NRK cells mapped approximately 330 bp from the hypersensitive site revealed by the novel 1.45-kb fragment (Fig. 7C). Analyses of independent preparations of DNase I-treated chromatin positioned this hypersensitive site within a 50-bp sequence. DNase I-digested chromatin from RCC2-

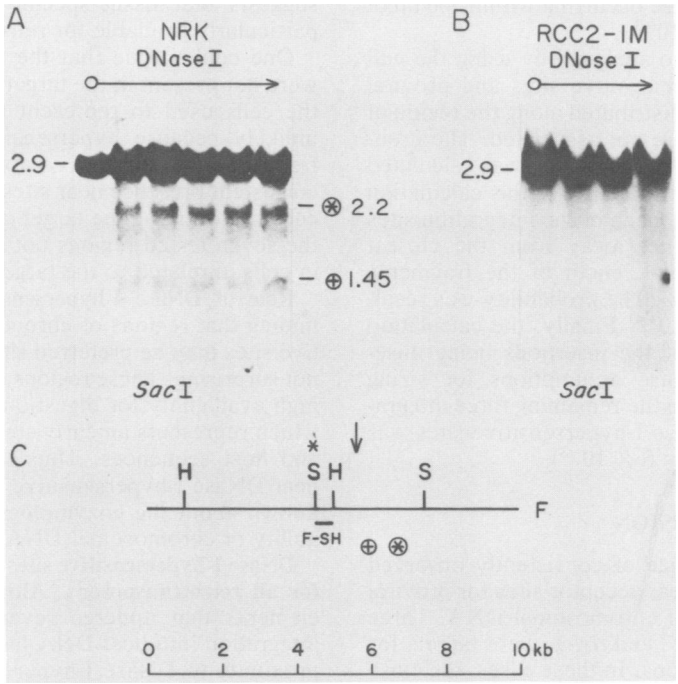

FIG. 7. Hypersensitive sites near the unselected integration in locus F. Autoradiographs of *Sac*I-digested DNA from DNase I-treated nuclei of NRK (A) or RCC2-1M (B) cells hybridized with probe F-SH. (C) Restriction endonuclease map of F. Designations are as in Fig. 1. S, *Sac*I.

TABLE 1. Positions of unselected proviral integrations and DNase I-hypersensitive (HS) sites[a]

Locus	Total length of DNA analyzed (kb)	Position of HS site(s) (kb)	Position of proviral integration (kb)
dsi-3	12.8	7.54	7.42
dsi-4	11.1	3.4, 1.2	2.97
C	8.8	3.65	3.7
F	6.4	1.48, 2.2	1.15

[a] The positions of hypersensitive sites and proviral integrations are given with respect to the restriction endonuclease site depicted with an asterisk in Fig. 4 to 7.

1M T cells did not reveal the presence of either of these two hypersensitive sites (Fig. 7B). Thus, the hypersensitive sites that map close to two unselected integrations in NRK cells were absent from DNase I-treated DNA of RCC2-1M cells, although these cells contained hypersensitive sites at c-*myc*, *dsi*-1, *dsi*-3, and *dsi*-4.

Statistical analysis of the distribution of acceptor and hypersensitive sites in regions with unselected insertions. Table 1 summarizes the positions of unselected proviral integrations and DNase I-hypersensitive sites. These data were subjected to several statistical analyses which made no assumptions about the frequency distribution of DNase I-hypersensitive sites in the rat genome, but rather asked whether the positions of DNase I-hypersensitive sites and proviral integration sites within the 6.4- to 12.8-kb DNA analyzed for each locus were likely to have occurred by chance. Initially we asked what portion of the total DNA analyzed in Table 1 was near a hypersensitive site and whether four insertions would have occurred in this portion of the total DNA by chance. The calculation indicated that the probability that each of the four unselected proviruses would have inserted within 500 bp of a hypersensitive site was 9×10^{-4}. Even if one of these four insertions turned out to be selected, the probability of the remaining three occurring within 500 bp of a hypersensitive site was 5×10^{-3}.

The data in Table 1 were also analyzed by using the null hypothesis that DNase I-hypersensitive sites and proviral integration sites are uniformly distributed along the region of DNA analyzed and therefore are not associated. The probability of the observed distribution occurring was calculated to be 2×10^{-6}. To determine how sensitive this calculation was to inaccuracies in mapping, each of the integration sites was then moved 500 bp farther away from the closest hypersensitive site and the total length of the fragments analyzed was decreased by 10%. The probability was recalculated and found to be 5×10^{-4}. Finally, the calculation was repeated for only three of the insertions being unselected and using the worst-case assumptions for sizing errors. The association between the remaining three integrations and the positions of DNase I-hypersensitive sites was still statistically significant ($P = 6 \times 10^{-3}$).

DISCUSSION

Our results show the presence of consistently observed DNase I-hypersensitive sites near acceptor sites for proviral integrations in seven regions of chromosomal DNA. Three of these regions, c-*erb*B, c-*myc*, and *dsi*-1, were targets for multiple tumor-inducing insertions. In these cases, the close proximity of hypersensitive and acceptor sites could be required for tumor induction. However, the demonstration that the four unselected insertions also occur near DNase I-hypersensitive sites strongly suggests that retroviruses preferentially integrate near hypersensitive sites.

Cells used to test for hypersensitive sites near insertions. Since hypersensitive sites can be tissue specific in their expression (1, 2, 10, 36, 39), DNase I-hypersensitive sites were examined in the closest available representative of the cell in which the insertions had occurred as well as in an unrelated cell lineage. The continuous line of NRK cells was used to represent the NRK cell that had insertions in C and F. A cell line established from another MoMLV-induced T-cell lymphoma was used for analysis of hypersensitive sites near insertions in the MoMLV-induced rat thymoma RT10-2. A continuous line of AEV-transformed erythroblasts was used for the analysis of hypersensitive sites near insertions in avian leukosis virus-induced erythroblastosis. Each of these lineages consistently displayed hypersensitive sites near the insertions in the infected cell they represented.

Interestingly, the cell lineages that were unrelated to the infected cell tended not to display hypersensitive sites near sites of insertion. A preparation of DNase I-treated chromatin from NRK cells displayed hypersensitive sites in loci C and F (targets for insertions in NRK cells) but did not display such sites in loci *dsi*-1, *dsi*-3, or *dsi*-4 (targets for insertion in a rat thymoma). Similarly, preparations of DNase I-treated chromatin from the cells cultured from a thymoma contained hypersensitive sites in *dsi*-1, *dsi*-3, and *dsi*-4 but not in C and F. The hypersensitive sites in the chicken c-*erb*B gene were detected only in AEV-R-transformed erythroblasts. These sites were not detected in the DNase I-treated chromatin of 10-day embryos, thymus, or bursa or in erythrocytes from 14-day embryos or anemic adults which displayed clear hypersensitive sites in c-*myc*. Of the seven regions analyzed, only one, c-*myc*, displayed hypersensitive sites in lineages that were both related and unrelated to the target cell. The frequent occurrence of insertions near hypersensitive sites expressed in some but not all cells is provocative in that it suggests that tissue-specific hypersensitive sites may be particularly available for retroviral integrations.

One could argue that the sites observed near insertions were not present in the target cell but occurred by chance in the cells used to represent the target cell. We think this unlikely because hypersensitive sites are not sufficiently frequent in cellular DNA for chance to account for their consistent presence near sites of insertion in chromatin from cells representing the target cell. This is evidenced by six of the seven tested regions not displaying hypersensitive sites in cells unrelated to the target cell.

Role of DNase I-hypersensitive sites in integration. The finding that regions of chromatin near DNase I-hypersensitive sites may be preferred sites for retroviral integrations is not surprising. These regions are identified by virture of their high availability for digestion by endonucleases, a reaction which represents an early step in the recombination of viral and host sequences. Thus the occurrence of integrations near DNase I-hypersensitive sites is consistent with what is known about the enzymology of integration and the availability of chromosomal DNA for such enzymatic reactions.

DNase I-hypersensitive sites as preferred sites of integration for all retrotransposons. Although acceptor sites for other elements that undergo reverse transcription and ordered integration into host DNA have not been analyzed for their proximity to DNase I-hypersensitive sites, the positions of selected as well as unselected insertions of these elements in known genes are consistent with insertions preferentially occurring close to DNase I-hypersensitive sites. For exam-

ple, hypersensitive sites frequently occur at the 5′ ends of genes or at sequences involved in chromosomal rearrangements. Insertions of intracisternal type A particles have been mapped near the 5′ end of the proto-oncogene c-*mos* (5) and near the site in the mouse kappe gene which undergoes rearrangements (16). Also, insertions of the Ty elements of yeasts occur at the 5′ ends of the *LYS2*, *ADH2*, *CYC7*, *CAR1*, and *HIS3* genes and near the 5′ ends of several tRNA genes (9). Most of the above insertions were selected for the alteration of a phenotype. In certain cases, the selection might have enriched for insertions at the 5′ ends of genes by requiring that the insertion not disrupt a coding sequence. However, Ty insertions into *LYS2* and *HIS4* were selected for the inactivation of a gene product which should not have discriminated against recombinations throughout the body of the gene. In conclusion, we suggest that all retrotransposons will preferentially integrate near DNase I-hypersensitive sites.

ACKNOWLEDGMENTS

We thank M. Groudine for advice on the preparation of DNase I-digested DNA and for some of the DNase I-digested DNAs. We also acknowledge D. Ross, J. Mielcarz, and E. Nacar who helped in the isolation of the MoMLV provirus clones, P. Leder for supplying us with the clone containing mouse c-*myc* sequences, and B. Vennstrom for the lambda clone c-*erb*1.

This work was supported by Public Health Service grants CA 27223 and CA 23086 (to H.L.R.) and CA 30674 (to D.L.S.) from the National Cancer Institute, a research grant (to D.L.S.) from the American Cancer Society, Massachusetts Division, and Cancer Center core grant CA 12708 awarded to the Worcester Foundation for Experimental Biology.

LITERATURE CITED

1. **Barsh, G. S., C. L. Roush, and R. E. Gelinas.** 1984. DNA and chromatin structure of the human alpha 1(1) collagen gene. J. Biol. Chem. **259**:14906–14913.
2. **Becker, P., R. Renkawitz, and G. Schutz.** 1984. Tissue-specific DNase I hypersensitive sites in the 5′-flanking sequences of the tryptophan oxygenase and the tyrosine aminotransferase genes. EMBO J. **3**:2015–2020.
3. **Beug, H., A. Kirchbach, G. Doederlein, J.-F. Conscience, and T. Graf.** 1979. Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation. Cell **18**:375–390.
4. **Breindl, M., K. Harbers, and R. Jaenisch.** 1984. Retrovirus induced lethal mutation in collagen 1 gene of mice is associated with an altered chromatin structure. Cell **38**:9–16.
5. **Canaani, E., O. Dreazen, A. Klar, G. Rechavi, D. Ram, J. B. Cohen, and D. Givol.** 1983. Activation of the c-*mos* oncogene in a mouse plasmacytoma by insertion of an endogenous intracisternal A-particle genome. Proc. Natl. Acad. Sci. USA **80**:7118–7122.
6. **Dhar, R., W. I. McClements, L. W. Enquist, and G. F. Vande Woude.** 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. USA **77**:3937–3941.
7. **Downward, J., Y. Yarden, E. Mayes, G. Scrace, H. Totty, P. Stockwell, A. Ullrich, J. Schlessinger, and M. D. Waterfield.** 1984. Close similarity of epidermal growth factor receptor and v-*erb*B oncogene protein sequences. Nature (London) **307**:521–527.
8. **Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. deHaseth, A. M. Skalka, and J. Leis.** 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses: nucleotide sequences required for site-specific cleavage. J. Virol. **56**:589–599.
9. **Eibel, H., and P. Philippsen.** 1984. Preferential integration of yeast transposable element Ty into a promoter region. Nature (London) **307**:386–389.
10. **Fritton, H. P., T. Igo-Kemenes, J. Nowock, U. Strech-Jurk, M. Theisen, and A. E. Sippel.** 1984. Alternative sets of DNase I-hypersensitive sites characterise the various functional states of the chicken lysozyme gene. Nature (London) **311**:163–165.
11. **Fung, Y. K., W. G. Lewis, L. B. Crittenden, and H. J. Kung.** 1983. Activation of the cellular oncogene c-*erb*B by LTR insertion: molecular basis for induction of erythroblastosis by avian leukosis virus. Cell **33**:357–368.
12. **Hishinuma, F., P. J. DeBona, S. Astrin, and A. M. Skalka.** 1981. Nucleotide sequence of acceptor site and termini of integrated avian endogenous provirus ev 1: integration creates a 6 bp repeat of host DNA. Cell **23**:155–164.
13. **Hoffmann, J. W., D. L. Steffen, J. Gusella, C. Tabin, S. Bird, D. Cowing, and R. Weinberg.** 1982. DNA methylation affecting the expression of murine leukemia proviruses. J. Virol. **44**:144–157.
14. **Ju, G., and A. M. Skalka.** 1980. Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. Cell **2**:379–386.
15. **Kirch, I. R., J. V. Ravetch, S.-P. Kwan, E. E. Max, R. L. Ney, and P. Leder.** 1981. Multiple immunoglobulin switch region homologies outside the heavy chain constant region locus. Nature (London) **293**:585–587.
16. **Kuff, E. L., A. Feenstra, K. Lueders, L. Smith, R. Hawley, N. Hozumi, and M. Shulman.** 1983. Intracisternal A particle genes as movable elements in the mouse genome. Proc. Natl. Acad. Sci. USA **80**:1992–1996.
17. **Lin, C. R., W. S. Chen, W. Kruiger, L. S. Stolarsky, W. Weber, R. M. Evans, I. M. Verma, G. N. Gill, and M. G. Rosenfeld.** 1984. Expression cloning of human EGF receptor complementary DNA: gene amplification and three related messenger RNA products in A431 cells. Science **224**:843–848.
18. **Majors, J. E., and H. E. Varmus.** 1981. Nucleotide sequences at host-proviral junctions for mouse mammary tumour virus. Nature (London) **289**:253–258.
19. **Mathis, D., P. Oudet, and P. Chambon.** 1980. Structure of transcribing chromatin. Progr. Nucleic Acid Res. Mol. Biol. **24**:1–55.
20. **Miles, B. D., and H. L. Robinson.** 1985. High-frequency transduction of c-*erb*B in avian leukosis virus-induced erythroblastosis. J. Virol. **54**:295–303.
21. **Nilsen, T. W., P. A. Maroney, R. G. Goodwin, F. M. Rottman, L. L. Crittenden, M. A. Raines, and H.-J. Kung.** 1985. c-*erb*B activation in ALV-induced erythroblastosis: novel RNA processing and promoter insertion result in expression of an amino truncated EGF receptor. Cell **41**:719–726.
22. **Raines, M. A., W. G. Lewis, L. B. Crittenden, and H.-J. Kung.** 1985. c-*erb*B activation in avian leukosis virus-induced erythroblastosis: clustered integration sites and the arrangement of provirus in the c-*erb*B alleles. Proc. Natl. Acad. Sci. USA **82**:2287–2291.
23. **Robinson, H. L., and G. Gagnon.** 1986. Patterns of proviral insertion and deletion in avian leukosis virus-induced lymphomas. J. Virol. **57**:28–36.
24. **Schubach, W., and M. Groudine.** 1984. Alteration of c-*myc* chromatin structure by avian leukosis virus integration. Nature (London) **307**:702–708.
25. **Shimotohno, K., S. Mizutani, and H. M. Temin.** 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. Nature (London) **285**:550–554.
26. **Shoemaker, C., J. Hoffman, S. Goff, and D. Baltimore.** 1981. Intramolecular integration within Moloney murine leukemia virus DNA. J. Virol. **40**:164–172.
27. **Steffen, D.** 1984. Proviruses are adjacent to c-*myc* in some murine leukemia virus induced lymphomas. Proc. Natl. Acad. Sci. USA **81**:2097–2101.
28. **Steffen, D., and R. A. Weinberg.** 1978. The integrated genome of murine leukemia virus. Cell **15**:1003–1010.
29. **Storb, U., B. Arp, and R. Wilson.** 1981. The switch region associated with immunoglobulin C_u genes is DNase I hypersensitive in T lymphocytes. Nature (London) **294**:90–92.
30. **Tsichlis, P. N., P. G. Strauss, and M. A. Lohse.** 1985. Concerted

DNA rearrangments in Moloney murine leukemia virus-induced thymomas: a potential synergistic relationship in oncogenesis. J. Virol. **56**:258–267.

31. **Ullrich, A., L. Coussens, J. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. Mayes, N. Whittle, M. D. Waterfield, and P. H. Seeburg.** 1984. Human epidermal growth factor receptor cDNA sequences and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature (London) **309**:418–425.
32. **Varmus, H. E.** 1982. Form and function of retroviral proviruses. Science **216**:812–820.
33. **Varmus, H. E., T. Padgett, S. Heasley, G. Simon, and M. Bishop.** 1977. Cellular functions are required for the synthesis and integration of avian sarcoma virus-specific DNA. Cell **11**:307–319.
34. **Vennstrom, B., and J. M. Bishop.** 1982. Isolation and characterization of chicken DNA homologous to the two putative oncogenes of avian erythroblastosis virus. Cell **28**:135–143.
35. **Weintraub, H.** 1985. Assembly and propagation of repressed and derepressed chromosomal states. Cell **42**:705–711.
36. **Weintraub, H., A. Larsen, and M. Groudine.** 1981. α-Globin gene switching during the development of chicken embryos: expression and chromosome structure. Cell **24**:333–344.
37. **Wilks, S.** 1962. Mathematical statistics, p. 248. John Wiley & Sons, Inc., New York.
38. **Wu, C.** 1980. The 5′ ends of Drosophila heat shock genes in chromatin are sensitive to DNase I. Nature (London) **286**:854–860.
39. **Wu, C., and W. Gilbert.** 1981. Tissue-specific exposure of chromatin structure at the 5′ terminus of the rat preproinsulin II gene. Proc. Natl. Acad. Sci. USA **78**:1577–1580.
40. **Yancopoulos, G., T. K. Blackwell, H. Suh, L. Hood, and F. W. Alt.** 1986. Introduced T cell receptor variable region gene segments in pre-B cells: evidence that B and T cells use a common recombinase. Cell **44**:251–259.