Proviral Insertions within the *int-2* Gene Can Generate Multiple Anomalous Transcripts but Leave the Protein-Coding Domain Intact

CLIVE DICKSON,^{1*} ROSALIND SMITH,¹ SHARON BROOKES,² AND GORDON PETERS²

Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX,¹ and Imperial Cancer Research Fund Laboratories, St. Bartholomews Hospital, Dominion House, London EC1A 7BE,² United Kingdom

Received 10 August 1989/Accepted 13 October 1989

We examined the effects of mouse mammary tumor virus integration on the multiple RNA transcripts expressed from the *int-2* proto-oncogene in virally induced breast tumors. Proviral insertion either upstream or downstream of the gene could simultaneously activate transcription from three dissimilar *int-2* promoters. In some tumors, the activating provirus lies within the transcription unit and disrupts the structures of the various RNAs. Insertions in the 5' region of the gene had complex effects depending on the orientation and position of the provirus relative to the three promoters and intron-exon boundaries. RNase protection experiments identified transcripts initiated in the viral long terminal repeat, at normal and cryptic sites in the *int-2* sequences, and from cryptic promoters in an inverted provirus. At the 3' end, insertions occurred within the untranslated trailer and provided alternative termination signals that substituted for one or both of the normal the poly(A) addition sites. However, in no instance, of the 20 tumors analyzed in detail, did a provirus perturb the presumed open reading frame of the gene. These data strongly implicate the normal product of the *int-2* gene, which is related to the fibroblast growth factor family, as a contributory factor in virally induced mammary tumors.

The *int-2* proto-oncogene is now known to encode a member of the fibroblast growth factor family of proteins, and its expression, at least in the mouse, appears to be restricted to specific embryonic cell types at various stages of development (2, 7, 26, 27). However, the initial identification and structural characterization of the gene resulted from its transcriptional activation in tumors induced by mouse mammary tumor virus (MMTV) (3, 11, 16, 17). For example, in the BR6 strain of mice, around 70% of the spontaneously arising mammary tumors are clonal proliferations of cells that have sustained a proviral insertion within about 15 kilobase pairs (kb) on either side of the *int-2* gene (16, 18, 19).

Typically, these tumors express multiple *int-2* transcripts that are superficially analogous to RNAs detected in the embryo and in cultured embryonal carcinoma (EC) cell lines, such as F9 and PCC4 (10, 24, 27). In EC lines, the expression of *int-2* increases when the cells differentiate in response to retinoic acid and dibutyryl cyclic AMP, and different cell lines contain different patterns of transcripts. They have therefore provided a convenient system in which to characterize the various species of RNA (24). By a combination of RNase protection analyses and blot hybridization with specific probes, six different classes of *int-2* mRNA have been distinguished that initiate at multiple cap sites within three separate promoter domains and terminate at either of two polyadenylation sites (10, 24).

The reasons behind this complexity are unclear, since all transcripts are predicted to encode the same protein (11, 24). One possibility would be to facilitate control of tissue-specific expression, or there could be subtle differences in the stability or translational efficiency of the various transcripts. It was therefore essential to establish whether the RNAs seen in MMTV-induced tumors are indeed analogous to their embryonic counterparts or whether they represent tumor-specific variants. Moreover, a subset of mammary

tumors express anomalous *int-2* RNAs as a consequence of proviral integration within the transcription unit (19). Here we have examined the structures of these *int-2* transcripts in a series of tumors and established that proviral insertions can influence all three promoters and alter the structures of transcripts at either the 5' or 3' end, but in no case do they disrupt the protein-coding capacity of the gene.

MATERIALS AND METHODS

Analysis of tumor DNA and RNA. The procedures used for the extraction of high-molecular-weight DNA and total RNA from MMTV-induced mammary tumors have been described previously (3, 17), and the examples analyzed here were included in the series of BR6 mammary tumors described in an earlier report (19). Poly(A)⁺ RNA was enriched by chromatography on poly(rU)-Sephrose (3). Samples (5 µg) were fractionated by electrophoresis in agarose gels containing formaldehyde (9) and transferred onto nylon filters by blotting in 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Hybridizations were performed with DNA probes that were labeled with ³²P by either nick translation or priming with random oligonucleotides. The conditions for hybridizing and washing of filters and autoradiography were as described previously (3). In some experiments, 20 µg of F9 or PCC4 cell RNAs was included as a control, and RNAs were denatured in glyoxal rather than formaldehyde (24).

Exon-specific probes. In addition to the *int-2* genomic DNA probe, designated *int-2 f* (11), which is specific for exon III of the gene, three further probes were prepared that contained sequences present exclusively in exons Ia, I, and Ib (24). These corresponded to nucleotides (nt) 590 to 1219, 1342 to 1681, and 1681 to 1815, respectively, in the genomic sequence (11).

Recombinant DNA techniques. DNA from tumors W26, E115, and D157B was digested to completion with *Eco*RI and ligated into the separated arms of λgtWES.λB, or λL47.1, as described elsewhere (3, 17). After in vitro packaging, approximately 5 × 10⁵ bacteriophage were plated on

* Corresponding author.

FIG. 1. *int-2* RNAs in mammary tumors. Samples (5 μ g) of poly(A)⁺ tumor RNA were fractionated in denaturing agarose gels, transferred to nylon membranes, and hybridized with ³²P-labeled probes specific for mouse *int-2* sequences. (a) RNA from tumor E155A was hybridized sequentially with probes specific for exons Ia, I, Ib, and III as indicated (see Materials and Methods). (b) RNAs from the indicated mammary tumors were hybridized with the *int-2* probe specific for exon III. The positions of the major classes of *int-2* RNA and their sizes in kilobase pairs are indicated on the right of each panel.

LE392 cells and screened with both MMTV long terminal repeat (LTR) and *int-2* probes. Phage containing inserts that were positive for both were plaque purified, and the respective *Eco*RI fragments were transferred into plasmid vectors for further analysis.

DNA sequence analysis. After the cloned *Eco*RI fragments were characterized by restriction enzyme mapping, selected fragments that spanned the virus-host junctions were subcloned into M13mp vectors. DNA sequences were determined in both orientations by the dideoxy-chain termination procedure. For W26, the sequence of the entire LTR was determined to verify that no additional viral information was present. A 6-base-pair (bp) duplication was noted at each end of the LTR, indicating that the site of integration was at nucleotide 1089 in the genomic sequence (11). For E115 and D157B, only the 5' junctions were recovered and characterized, but these established both the position and the orientation of the respective proviruses (indicated in Fig. 2).

RNAse protection. Samples (generally 2 μ g) of poly(A)⁺ tumor RNA or 20 μ g of total RNA from EC cells were subjected to RNAse protection analysis exactly as described by Smith et al. (24). The uniformly labeled antisense RNA probes were prepared by transcribing cloned fragments of genomic DNA with phage SP6 RNA polymerase. The probes are numbered as in the previous report (24) (with the addition of probes 1.1, 1.2, and 1.3), and the restriction enzyme sites that mark the 5' and 3' ends of the *int-2*-specific sequences in each probe are indicated in the relevant figures. The sizes of each probe refer to the *int-2*-specific sequences only (occasionally, enzyme sites in the polylinker of the transcribed plasmid were more convenient) and were calculated without considering exact cleavage positions within the recognition sites. However, when measuring the sizes of protected fragments, the exact ends were taken into account. Even so, we estimate that fragment sizes deduced experimentally are only accurate to ± 5 bp.

RESULTS

Typical and atypical *int-2* RNAs in MMTV-induced tumors. Most tumors in which the *int-2* gene is transcriptionally

activated carry an MMTV provirus in adjacent cellular DNA, rather than within the gene, in an orientation appropriate for enhancement of the normal *int-2* promoters by elements in the viral LTR (3, 18, 19). These tumors typically contain four sizes of *int-2* RNA, of approximately 2.9, 2.7, 1.8, and 1.6 kb, all of which are detectable with a probe for exon III of the gene (Fig. 1). In the present study, the tumor designated E155A was used to represent the typical case, since the activating MMTV provirus is located about 10 kb upstream of *int-2* (19). Results obtained with other typical tumors were qualitatively indistinguishable. However, by analogy with EC cells, the four bands detected on RNA blots were assumed to reflect six distinct types of RNA (10, 24). Thus, the 2.9-kb band is now understood to comprise RNAs that initiate at either the P1 or the P2 promoter. Tumor RNAs of this size contained sequences derived from either exon Ia (initiated at P1) or exon I (initiated at P2), while the 2.7-kb transcript was not detected by probes for these exons (Fig. 1a). This latter RNA presumably initiates at a third promoter, designated P3 (Fig. 2), since it was detectable with a probe specific for exon Ib (Fig. 1a) (24). A rationalization of these findings is presented in Fig. 2. The smaller 1.8- and 1.6-kb forms of *int-2* RNA that are barely visible in the reproduction of Fig. 1a have been shown to be analogs of the 2.9- and 2.7-kb classes, respectively, that terminate at the upstream poly(A) addition site (24). These smaller transcripts are consistently underrepresented in tumors compared with EC cells.

Although the majority of *int-2*-positive tumors conform to this typical pattern, our previous survey of spontaneous tumors in BR6 mice revealed several examples in which the patterns of *int-2* transcripts were unusual (19). Some of these atypical cases are shown in Fig. 1b. From relatively detailed mapping of the integrated proviruses in these tumors, it became clear that the anomalous transcripts arose through physical disruption of the *int-2* transcription unit and that this could occur at either the 5' or 3' end. Thus, tumors W26, E115, and D157B expressed unusual RNAs as a result of proviral insertion in the 5' region of the gene, while several others, such as D124, C46, V97A, and W106, had proviruses

FIG. 2. Proviral insertions in the promoter region of *int-2*. In the upper part of the figure, the structures of three classes of *int-2* RNA are drawn to scale (relative to the restriction map below), with exons Ia, I, and Ib represented as boxes and numbers referring to coordinates in the genomic sequence of Moore et al. (11). Shaded areas correspond to protein-coding sequences, and the multiple start sites associated with promoters P1 and P2 are depicted by the vertical lines. The positions and orientations of integrated proviruses in the numbered mammary tumors are shown as arrowheads, with their positions in the genomic sequence indicated below. The restriction map shows the sites of cleavage for the enzymes *Bam*HI (B), *Nae*I (N), *Pvu*II (P), *Sac*I (S), *Sma*I (Sm), and *Stu*I (St) that are relevant to the generation of specific antisense RNA probes. Not all sites are included. The lines below the map define the extents and positions of the genomic sequences transcribed into RNA probes, numbered as in reference 24. The length of each probe and the restriction sites that define each end are as indicated.

within the presumed 3' untranslated region, leading to extended forms of *int-2* RNA (Fig. 1b).

Locating proviruses by Southern blotting of tumor DNA was not sufficiently accurate to draw firm conclusions about their effects on the structure and coding capacity of *int-2* RNAs. Two approaches were therefore adopted to map these proviruses more precisely. First, recombinant DNA clones were isolated representing the junctions between MMTV proviral DNA and *int-2* genomic sequences. In the second approach, tumor RNAs were subjected to RNase protection analysis with a series of antisense probes derived from specific regions of the genomic DNA. Since this technique detects discontinuity between the RNA and genomic sequences, the mapping of hybrid RNAs generated by viral insertion can be used to locate the site of integration.

Characterization of virus-host junctions. Bacteriophage λ libraries were constructed containing *Eco*RI-digested DNA from tumors W26, E115, and D157B and screened with both *int-2*- and MMTV-specific probes. Tumor W26 contained a single MMTV LTR in a 5' to 3' orientation relative to *int-2*, resulting in an 11.3-kb *Eco*RI genomic DNA fragment as opposed to the 10-kb normal allele (11). With E115 and D157, only the respective 5' junction fragments were recovered. When these were characterized and the DNA sequence

was determined across the virus-cell boundaries, it was evident that the provirus in D157B was also in the same transcriptional orientation as the *int-2* gene, but apparently intact. In tumor E115, on the other hand, a complete provirus was present in an inverted orientation. These data also located the precise sites of integration in the genomic sequence described by Moore et al. (11) at nt 1089, 1195, and 1503, respectively, in tumors W26, D157B, and E115 (Fig. 2). No common features were evident in the sequences immediately adjacent to the integration sites.

The three proviruses in question lie within the promoter domains mapped in EC cells in previous studies (Fig. 2). Thus, the LTR in tumor W26 had integrated in exon Ia, between P1 and P2, the provirus in D157B lay just upstream of P2, while that in E115 split P2 and P3. We therefore attempted to reconcile these findings with the RNA structures present in the respective tumors by RNase protection.

Tumor RNAs containing exon Ia. *int-2* transcripts that initiate at promoter P1 and contain an exon Ia to Ib splice have been identified in EC cells by both RNase mapping and cDNA cloning (10, 24). However, since these transcripts are at relatively low abundance in F9 cells and are undetectable in PCC4 cells, it was important to establish whether counterparts existed in typical and atypical mammary tumors.

FIG. 3. RNase protection analysis of exon Ia. Approximately 2 μ g of poly(A)⁺ tumor RNA from the indicated tumors or 20 μ g of total RNA from either F9 or PCC4 cells that had been treated with retinoic acid and dibutyryl cyclic AMP was subjected to RNase protection analysis as described previously (24). (a) Results obtained with the probe designated 12 in Fig. 2. Numbers to the left of the panel identify the positions and sizes (nucleotides) of the RNase-resistant products. The sizes were estimated relative to a series of plasmid DNA fragments of known length (*Hinf*I-digested pAT153) and a standard DNA sequence ladder, indicated as lane M. Numbers on the right identify the positions and sizes of the standard markers. (b) Probe 13 was used to delineate the multiple start sites associated with initiation at promoter P1. The arrowheads on the right identify products corresponding to the major cap sites observed in F9 cells. Numbers on the left identify the sizes (in nucleotides) and positions of DNA markers calibrated with a sequence ladder.

The expression of exon Ia was first examined by protection of the 184-nucleotide antisense RNA transcribed from the genomic DNA segment designated as probe 12 (Fig. 2). Assuming that the structure of the transcript was as depicted in Fig. 2 (subsequently confirmed with multiple probe combinations), probe 12 should have detected discontinuity at the splice donor site at the 3' end of exon Ia. Tumor E155A, in which the provirus lies distant from the *int-2* gene, contained RNAs that include exon Ia and splice at the predicted position, yielding a protected fragment of 138 nt (Fig. 3a). The 184-nt fragment corresponded to full protection of the probe and may have been indicative of unspliced forms in the preparation of total RNA used in these studies; the control RNA from rat 208F cells and F9 cell RNA did not yield this fully protected product, indicating that conditions were adequate for complete digestion of the probe.

A different result was obtained with RNA from tumor W26, in which the solo LTR lies within exon Ia. Two protected fragments were observed that were each about 50 nt smaller than those seen in the typical tumor sample. Other data (Fig. 4) confirm that the larger of the two (~135 nt) was derived from unspliced RNA and represented the distance between the LTR and the end of the probe (calculated to be 137 nucleotides), while the ~90-nt fragment reflected the distance between the LTR and the end of exon Ia (calculated as 91 nucleotides). It is not clear why this latter fragment resolves into multiple species, 1 base apart, but the effect is consistently seen with probes that span this discontinuity (Fig. 4a and other data not shown). Tumor E115, in which the provirus lies downstream of the splice donor site, did not show evidence of transcripts that contain exon Ia (data not shown).

FIG. 4. RNase protection analysis of exons I and Ib. (a) RNase protection was performed on the indicated RNA samples with the 1,225-nucleotide *PvuII-SacI* probe designated 9 in Fig. 2. Numbers on the left of the figure identify the positions and sizes of relevant RNase-resistant fragments, calculated relative to the standard markers and sequence ladder on the right (lane M). Discontinuities characteristic of the multiple starts in P2 and P1 (24) are indicated by the lines, and other discontinuities are identified that correspond to the splice donor in exon Ia (SD), the splice acceptor in exon Ib (SA), and initiation at P3 (10). The 321-nt fragment maps the position of provirus integration in tumor E115. (b) A similar analysis was performed with probe 11. Since this probe extends from the *NaeI* site at nucleotide 1681, shorter analogs are produced for each of the discontinuities shown in panel a. The positions of the multiple start sites associated with the P2 promoter are indicated by arrowheads, while the 496- and 188-nt fragments correspond to the sites of proviral integration in tumors D157B and E115, respectively. The arrowhead adjacent to the ~500-nt fragment in the W26 lane identifies the novel P2 initiation site characteristic of this tumor, while the larger band in this track derives from unspliced RNA initiated in the LTR. The sizes of fragments (in nucleotides) were deduced directly from the sequence ladder (lane M).

Although these results implied splicing from the 3' end of exon Ia, in at least a proportion of the tumor RNAs, it was of interest to examine the 5' end of such transcripts, to ascertain whether they initiated at the same cap sites as in F9 cells. The analyses were therefore repeated with probe 13, which extends upstream of promoter P1. In E155A, a series of protected fragments were observed that were indicative of multiple start sites within P1, each fragment being equivalent to a complete exon Ia (Fig. 3b). Most of these 5' ends were analogous to those seen in F9 cells, but there were also additional discontinuities reflected in fragments of between 280 and 450 nt. It is not clear whether these represent novel, tumor-specific initiation sites or simply an increase in the efficiencies with which normal cap sites are employed. Neither the W26 nor E115 tumor RNAs gave signals for initiation at P1 (data not shown).

Tumor RNAs containing exon Ib. Previous studies on EC cells established that *int-2* transcripts that include exon Ia are spliced such that the donor site (nucleotide 1175) is linked to the consensus acceptor site (nucleotide 1661) that marks the 5' end of exon Ib (10, 24) (Fig. 2). However, we now recognize that other transcripts initiate at a unique site very close to the 5' end of exon Ib, at what was originally interpreted as an alternative splice acceptor (nucleotide 1667). Although separated by only 6 bases, the two discontinuities, attributable to either spliced RNA initiated at P1 or de novo initiation at promoter P3, can be readily distinguished. For example, when probe 9 (Fig. 2) was used in RNase protection experiments with tumor E155A RNA, two protected fragments of 147 and 141 nt were observed at approximately equivalent intensities (Fig. 4a). This would imply that transcription initiation at promoters P1 and P3 occurred with roughly equal efficiency in typical mammary tumors, in some contrast to the situation in F9 cells where P3, reflected in the 141-nt product, was clearly dominant (24). It also contrasts with the result obtained from tumor W26 in which the converse was true (Fig. 4a). Here, the major RNA species initiate in the MMTV LTR, and about 70% of these transcripts undergo a splice at the end of exon Ia (91-nt fragment in Fig. 3a). The recipient of this splice is presumably the 5' end of exon Ib, as reflected in the 147-nt protected fragment, providing an explanation for the strength of this signal relative to the fragment generated by initiation at P3 (Fig. 4a).

With tumor E115, the signal indicative of the splice acceptor site was extremely faint (Fig. 4a). Since there was no evidence for RNA splicing from exon Ia in this tumor, it is possible that minor transcripts initiating at a cryptic promoter(s) within the MMTV provirus become spliced to this site. However, the signal corresponding to initiation at P3 was also unexpectedly weak, despite the fact that the MMTV provirus in E115 is presumably in an ideal position to enhance transcription from this promoter. Either the level of transcription from P3 is inherently low in this tumor, for reasons that are not clear, or the signals derived from P3 were masked by more abundant transcripts initiated elsewhere. The most likely origin of these RNAs would be initiation from cryptic promoters on the negative strand of the MMTV provirus, but this has not been formally proved.

Tumor RNAs containing exon I. The major protected fragment generated from E115 RNA was approximately 321 nt and corresponded to discontinuity at the site of proviral integration. This mapped about 100 nucleotides downstream of the P2 domain, but proviral insertion appeared to prevent initiation at this promoter. However, in the typical tumor, E155A, initiation within P2 could also be visualized with

probe 9, as evidenced by a heterogeneous set of bands of between 450 and 600 nt. These are very similar in distribution to the analogous start sites observed at much lower levels in PCC4 cells (Fig. 4a). The data are complicated by the fact that probe 9 can also detect heterogeneous starts within P1 (inactive in PCC4 cells), as represented by the characteristic series of protected fragments between 215 and 265 nt. These latter fragments were not observed with RNA from W26 and E115, confirming that P1 does not function in these tumors (Fig. 4a). However, with W26 RNA, there was evidence for novel initiation sites in the upstream region of P2 (fragments of ~650 to 750 nt) and a protected fragment of around 91 nt was again observed (Fig. 4a), representing the distance between the MMTV LTR and the splice donor site in exon Ia, as in Fig. 3a.

These conclusions were supported and clearer indications for initiation in P2 were obtained by using probe 11 (Fig. 4b). This probe extended from the *NaeI* site at 1681 (Fig. 2) and therefore yielded shorter analogs of each of the fragments generated with probe 9. Only those over about 150 nt in size are shown in the gel depicted in Fig. 4b. Transcription initiation at P2 was reflected in protected fragments of between 308 and 443 nt, with the major start site (361-nt product) centered on nucleotide 1320. Thus, typical mammary tumors such as E155A express RNAs that are almost exact counterparts of those initiated at P2 in PCC4 cells. In W26, on the other hand, a different set of presumptive cap sites were again evident with probe 11, the most prominent of which (~500-nt fragment) mapped about 150 nucleotides upstream of the major start normally associated with P2. Since a fragment of similar size was also observed with E155A RNA, albeit at much lower intensity (Fig. 4b), it may correspond to a normally rare cap site whose usage is greatly enhanced by the nearby insertion of the MMTV LTR. The larger product from W26 RNA would be consistent with unspliced RNA since it represents the distance between the LTR (1089) and the *NaeI* site (1681) that marks the end of probe 11.

Probe 11 also revealed discontinuities that confirmed the sites of proviral integration in tumors E115 (188-nt fragment) and D157B (496-nt fragment) relative to the *NaeI* site. For D157B, there was insufficient tumor RNA to perform a comprehensive analysis, but the results shown in Fig. 4b suggested that the major transcripts in this tumor were initiated in the viral LTR, upstream of and substituting for P2.

Tumor RNAs with abnormal 3' ends. Although proviral insertions tend to be more common 5' to the *int-2* gene, there are nevertheless several tumors in which proviruses have been mapped at the 3' end of the gene. In 5 of the 20 *int-2*-positive BR6 tumors, the provirus was inserted within the transcription unit, in the same transcriptional orientation. Figure 1b showed that these tumors contained atypical *int-2* RNAs, presumably owing to perturbation of the normal termination signals for transcription. Indeed, some of these RNAs were shown by sandwich blotting to be hybrid transcripts comprising both *int-2* and MMTV LTR sequences (22; data not shown), suggesting that they terminate in the viral LTR. However, since the *int-2* gene has two alternative polyadenylation signals, it was of interest to determine the precise location of the various integrated proviruses and to assess whether any of the insertions affected the protein-coding domain.

Rather than construct recombinant phage libraries for all these tumors, we located the proviruses by RNase mapping. Although this method is less precise than cloning and se-

FIG. 5. Proviral insertions in the 3' untranslated region of *int-2*. The 3' end of the *int-2* gene is drawn to scale relative to the restriction map, with exon III depicted as boxes in which protein-coding sequences are shaded. The exon terminates at two alternative polyadenylation signals at nucleotides 7508 and 6404 in the genomic sequence (11). Arrowheads at the top of the figure identify the positions and orientations of proviruses in the indicated tumors. The restriction map shows the cleavage sites for the enzymes *AccI* (Ac), *EcoRI* (E), *HindIII* (H), *PstI* (P), and *PvuII* (Pv) that were used in generating the numbered antisense probes shown beneath the map. Only relevant sites are shown.

quencing of virus-host junctions, we estimated that the positions deduced from the sizes of protected fragments were accurate to within about 5 bases. The results of these analyses are summarized in Fig. 5, and data supporting the conclusions are presented in Fig. 6. Briefly, a series of antisense probes were prepared, designated 1, 1.1, 1.2, and 1.3 (24) (Fig. 5), that were capable of detecting discontinuity between tumor-specific RNA and the genomic sequences encompassing exon III. Wherever possible, each discontinuity and therefore site of integration was mapped with at least two specific probes. For example, the provirus in tumor D124 was located at nucleotide 6203 (± 5) by using probes 1.2 and 1.3 (24). With the former, extending to the *AccI* site (nucleotide 5997), a protected fragment of 206 nt was obtained (data not shown), while the latter, extending to the *PstI* site (5896), yielded a 302-nt fragment (Fig. 6a), suggesting that integration had occurred at these distances from the respective restriction sites. By similar reasoning, the provirus in tumor Z79 was mapped to nucleotide 6260, approximately 359 nt from the *PstI* site (Fig. 6a). Representative data for the other tumors and the respective probes used are illustrated in Fig. 6b and c.

DISCUSSION

Multiple *int-2* promoters respond to proviral activation. Previous studies on the expression of the *int-2* gene in mouse embryos and EC cells have indicated an inherent transcriptional complexity for which there is as yet no clear explanation (7, 10, 24, 26, 27). Much of the complexity arises through the differential usage of three distinct promoter regions in which RNA transcripts initiate at multiple sites (10, 24). However, all the transcripts characterized thus far would be predicted to encode the same 27-kilodalton pro-

tein, now recognized as a member of the fibroblast growth factor family (2, 4, 11). One possibility is that differential promoter usage reflects a need to respond to factors governing gene expression in very specific cell types and at highly restricted times during development.

In MMTV-induced mammary tumors, expression of *int-2* is activated in a cell type in which the gene is normally silent, at least as judged by steady-state levels of RNA (3, 7). It was therefore important to establish whether the transcription patterns in tumors reflected those seen in EC cells, such as F9 and PCC4, and whether the different *int-2* promoters responded differently to activation by MMTV. The three promoter domains that have been distinguished, designated P1, P2, and P3 (10, 24) (Fig. 2), have no obvious features in common; the P1 region has an average base composition, P2 is very G+C-rich, while the region now associated with P3 is unusually C+T-rich, analogous to some ribosomal gene promoters (5). Northern (RNA) blotting analyses with DNA probes specific for exons Ia, I, and Ib indicated that in a typical tumor, such as E155A, transcripts are detectable that correspond to initiation at each of the three promoters (Fig. 1a). This was subsequently confirmed by RNase protection with a series of antisense probes, as presented in Fig. 3 and 4. Qualitatively, the RNAs initiated at each promoter are very similar to those seen in EC cells, although there are minor variations in the distribution of cap sites. Quantitatively, each promoter is more active in tumors than in EC cells, and although P2 and P3 are dominant, it is clear that all three promoters function in mammary tumors. This contrasts with the situation in F9 and PCC4 cells, in which only two of the three promoters function in each cell type: P1 and P3 in F9 cells, and P2 and P3 in PCC4 cells (10, 24). Since the sequences involved in directing initiation from each pro-

FIG. 6. RNase mapping of proviruses in exon III of *int-2*. Poly(A)⁺ RNA from the indicated tumors was subjected to RNase mapping with probes that crossed the sites of integration. The location of the provirus (shown by the arrowhead) was deduced from the length of the RNase-resistant fragment (striped line) compared with the input probe (solid line). Each insertion was mapped relative to restriction sites that marked the ends of two independent probes (Fig. 5). The autoradiographs in the lower part of the figure give examples of the data obtained with probes 1.3 (a), 1.1 (b), and 1 (c) on the indicated tumors, with the respective probe and product depicted graphically above each panel. The sizes of the protected fragments (in nucleotides), indicated to the left and right of the panel, were calculated relative to standard markers and a DNA sequence ladder.

moter may overlap (D. Grinberg and J. Thurlow, personal communication), it is possible that MMTV influences gene expression through a common element.

MMTV insertions that perturb *int-2* promoter usage. In most *int-2*-expressing mammary tumors, the MMTV provirus lies outside the transcription unit and the patterns of *int-2* transcription are very similar if not indistinguishable from those described here for tumor E155A (3, 16, 19). However, there are also examples that express anomalous transcripts resulting from proviral insertions in the 5' and 3' untranslated regions of the gene. In three of these, W26, E115, and D157B, the provirus has been accurately located within the *int-2* promoter regions by sequencing across the virus-host junctions. By combining this information with detailed RNase protection analyses, we have accounted for the major *int-2* transcripts in these tumors.

In the first example, W26, *int-2* expression is activated by a single MMTV LTR in exon Ia. We previously concluded

that the major site for *int-2* transcription in this tumor was likely to be the viral promoter in the LTR, based on primer extension analysis (11; data not shown). Here we confirmed this conclusion but showed that transcripts initiated in the LTR may have two structures, depending on whether they undergo splicing between exons Ia and Ib. The calculated size of the major spliced transcript would be 2.75 kb, very similar to that estimated from Northern blot analysis (Fig. 1b), while the less abundant unspliced form would be around 3.2 kb. However, there is also evidence for transcription initiation outside the LTR, mostly from a site in the most 5' region of P2, yielding an approximately 3.0-kb RNA. The reason for initiation at this unusual cap site is unclear but may reflect influences of the MMTV enhancer on the most proximal cap site in the P2 promoter. A bidirectional enhancer might also have been expected to activate transcription from the nearby P1 promoter, but no initiation at P1 was discernible in tumor W26. Nevertheless, it is also clear that

W26 contains minor *int-2*-specific transcripts for which we cannot as yet provide a full explanation, other than the possibility of unspliced or partially spliced forms.

With tumor D157B, there was insufficient RNA to conduct a comprehensive analysis, but Northern blots indicated a transcription pattern similar to that in W26 (Fig. 1b). This would be consistent with transcription initiation from the viral promoter, presumably in the 3' LTR of the full-length MMTV provirus present in this tumor. Since the provirus lies in the upstream part of the P2 domain, it has the potential to subvert transcription from both P1 and P2 and may also override P3 by promoter occlusion effects (1, 6).

In tumor E115, on the other hand, the ostensibly intact provirus is in the inverse orientation and cannot program transcription from the normal viral promoter. Nevertheless, the RNase protection experiments we present here suggest that the major transcript in this tumor must initiate within the provirus, presumably at some cryptic promoter on the antisense strand of viral DNA. The position of the provirus would prevent initiation in P2, and there was no evidence for transcription from P1 cap sites. Thus, we conclude that the major 2.8-kb RNA species consistently seen in this tumor is initiated in the provirus, while the less abundant 2.7-kb RNA is a normal P3-derived transcript.

In the only other system in which comparable analyses have been reported, the activation of *c-myc* by avian and murine leukemia viruses, there are close analogies to all three of these scenarios, although the transcription pattern is inherently less complex (8, 12, 20, 21, 23). Thus, proviruses have been found within the first exon or intron that can drive expression from the viral promoter, enhance transcription from the *c-myc* promoters, or induce initiation from cryptic promoters. Although the generalities of the enhancer and promoter insertion models still apply (reviewed in reference 15), it is now clear that the influences of retroviral LTRs on proto-oncogene transcription can be manifold.

MMTV insertions that perturb transcriptional termination. Additional complexity in the transcription of the mouse *int-2* gene arises through usage of two alternative polyadenylation signals, both variants of the consensus (10, 11, 24). Although we cannot rigorously exclude some influences on RNA stability or translational efficiency, it seems likely that this complexity is largely irrelevant. In EC cells, up to 50% of transcripts terminate at the upstream poly(A) site (10, 24), but in MMTV-induced mammary tumors, the majority of *int-2* transcripts terminate at the downstream site, and the 1.8- and 1.6-kb RNA size classes are barely discernible on Northern blots (Fig. 1). One possible explanation would be that the transcription complexes that traverse the gene have different components in different cell types or in MMTV-infected cells. Alternatively, the rate of transcription may simply be higher in tumors than in EC cells, making it less likely to achieve termination at the upstream signal. In some tumors, for example, E146, there is readthrough of both polyadenylation signals, and three sets of transcripts can be resolved (Fig. 1b). Here the provirus is very close to the 3' end of the transcription unit (Fig. 5 and 6), allowing termination at the normal signals within the viral LTR and extending each class of RNA by the length of the U3 region (~1.2 kb). In other tumors, a similar phenomenon occurs, but the provirus lies within the transcription unit (Fig. 5 and 6). Depending on the location of the provirus relative to the two polyadenylation signals, this can result in extended forms of the 2.9- and 2.7-kb RNAs but normal 1.8- and 1.6-kb transcripts, as for tumors V97A, C46, and W106. The latter example is remarkable in that the integration has

occurred very close (within a few bases) to the downstream poly(A) addition site. Alternatively, as in D124 and Z79, the provirus lies upstream of both signals, resulting in only one set of transcripts (Fig. 5 and 6). The loss of *int-2* sequences from the 3' end of these RNAs is compensated for by the addition of viral sequences, making their sizes very similar to the normal 2.9- and 2.7-kb species (Fig. 1b).

Proviral insertion does not perturb the protein-coding domain. Retroviral activation of cellular proto-oncogenes, whether resulting from transduction into the viral genome or insertional mutagenesis by integrated proviral DNA, can often involve structural alterations to the gene product (reviewed in reference 15). However, there are also examples of insertionally activated oncogenes whose coding potential remains unaltered, implying that it is the normal gene product that contributes to tumorigenesis. Obvious examples of the latter are the *int-1*, *pim-1*, and *myc* genes (12, 13, 20, 23, 25), and here we established that *int-2* belongs in this category. Despite the complexity of transcripts and several examples of proviral insertions that disturb the transcription unit, all the proviruses mapped so far lie outside the open reading frame that specifies the *int-2* protein. At the 5' end of the gene, the most intrusive provirus (E115) is approximately 275 nt upstream of the presumed AUG initiation codon. Although there is recent evidence for initiation at a non-AUG codon, leading to amino-terminally extended forms of the protein (P. Acland, M. Dixon, G. Peters, and C. Dickson, submitted for publication), these would still be unaffected by insertions upstream of P3. At 3' end, the most intrusive provirus that has been mapped to date, in tumor D124, is ~150 nucleotides downstream of the termination codon. Although we cannot rule out point mutations or small alterations in the coding sequences, we would conclude that the contribution of *int-2* to oncogenesis is a function of elevated or inappropriate expression of the normal gene product. It has recently been shown that the product can function as a mitogen for cultured mammary epithelial cells and fibroblasts, in line with its classification as a member of the fibroblast growth factor family (4, 14). Similarly, under appropriate circumstances, the gene can induce morphological transformation of NIH 3T3 cells (R. Deed, W. Walther, C. Dickson, and G. Peters, unpublished data) and lead to epithelial hypertrophy in transgenic mice (W. J. Muller, F. S. Lee, C. Dickson, G. Peters, P. Pattengale, and P. Leder, submitted for publication). However, the normal role of *int-2* is more likely to be as an inductive agent modulating the growth or differentiation of specific embryonic lineages (26, 27). The effect of MMTV integration, in whatever orientation and location relative to the *int-2* transcription unit, is to reactivate the expression of a fetal gene in an adult tissue but not to alter the specificity of the gene product.

ACKNOWLEDGMENTS

We are grateful to Robert Moore, Mark Dixon, Marysia Placzek, and Jan Witkowski for contributing to some aspects of this work and to Roger Watson and Steve Goodbourn for helpful discussions on the manuscript.

LITERATURE CITED

1. Cullen, B. R., P. T. Lomedico, and G. Ju. 1984. Transcriptional interference in avian retroviruses—implications for the promoter insertion model of leukaemogenesis. *Nature (London)* 307:241–245.
2. Dickson, C., and G. Peters. 1987. Potential oncogene product related to growth factors. *Nature (London)* 326:833.
3. Dickson, C., R. Smith, S. Brookes, and G. Peters. 1984. Tumor-

- igenesis by mouse mammary tumor virus: proviral activation of a cellular gene in the common integration region *int-2*. *Cell* 37:529–536.
4. Dixon, M., R. Deed, P. Acland, R. Moore, A. Whyte, G. Peters, and C. Dickson. 1989. Detection and characterization of the fibroblast growth factor-related oncoprotein INT-2. *Mol. Cell. Biol.* 9:4896–4902.
 5. Dudof, K. P., and R. P. Perry. 1986. Properties of a mouse ribosomal protein promoter. *Proc. Natl. Acad. Sci. USA* 83: 8545–8549.
 6. Emermann, M., and H. M. Temin. 1984. Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism. *Cell* 39:459–467.
 7. Jakobovits, A., G. M. Shackelford, H. E. Varmus, and G. R. Martin. 1986. Two proto-oncogenes implicated in mammary carcinogenesis, *int-1* and *int-2*, are independently regulated during mouse development. *Proc. Natl. Acad. Sci. USA* 83: 7806–7810.
 8. Linial, M., and M. Groudine. 1985. Transcription of three *c-myc* exons is enhanced in chicken bursal lymphoma cell lines. *Proc. Natl. Acad. Sci. USA* 82:53–57.
 9. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual, p. 202. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 10. Mansour, S. L., and G. R. Martin. 1988. Four classes of mRNA are expressed from the mouse *int-2* gene, a member of the FGF gene family. *EMBO J.* 7:2035–2041.
 11. Moore, R., G. Casey, S. Brookes, M. Dixon, G. Peters, and C. Dickson. 1986. Sequence, topography and protein coding potential of mouse *int-2*: a putative oncogene activated by mouse mammary tumour virus. *EMBO J.* 5:919–924.
 12. Nottenburg, C., and H. E. Varmus. 1986. Features of the chicken *c-myc* gene that influence the structure of *c-myc* RNA in normal cells and bursal lymphomas. *Mol. Cell. Biol.* 6: 2800–2806.
 13. Nusse, R., A. van Ooyen, D. Cox, Y. K. T. Fung, and H. Varmus. 1984. Mode of proviral activation of a putative mammary oncogene (*int-1*) on mouse chromosome 15. *Nature* 307:131–136.
 14. Paterno, G. D., L. L. Gillespie, M. S. Dixon, J. M. W. Slack, and J. K. Heath. 1989. Mesoderm-inducing properties of INT-2 and kFGF: two oncogene-encoded growth factors related to FGF. *Development* 106:79–83.
 15. Peters, G. 1989. Oncogenes at viral integration sites, p. 23–66. In D. M. Glover and B. D. Hames (ed.), *Frontiers in molecular biology: oncogenes*. IRL Press, Oxford.
 16. Peters, G., S. Brookes, M. Placzek, M. Schuermann, R. Michalides, and C. Dickson. 1989. A putative *int* domain for mouse mammary tumor virus on mouse chromosome 7 is a 5' extension of *int-2*. *J. Virol.* 63:1448–1450.
 17. Peters, G., S. Brookes, R. Smith, and C. Dickson. 1983. Tumorigenesis by mouse mammary tumor virus: evidence for a common region for provirus integration in mammary tumors. *Cell* 33:369–377.
 18. Peters, G., and C. Dickson. 1987. On the mechanism of carcinogenesis by mouse mammary tumor virus, p. 307–319. In D. Medina, W. Kidwell, G. Heppner, and E. Anderson (ed.), *Cellular and molecular biology of mammary cancer*. Plenum Publishing Corp., New York.
 19. Peters, G., A. E. Lee, and C. Dickson. 1986. Concerted activation of two potential proto-oncogenes in carcinomas induced by mouse mammary tumour virus. *Nature (London)* 320:628–631.
 20. Robinson, H. L., and G. C. Gagnon. 1986. Patterns of proviral insertion and deletion in avian leukosis virus-induced lymphomas. *J. Virol.* 57:28–36.
 21. Shih, C.-K., M. Linial, M. M. Goodenow, and W. S. Hayward. 1984. Nucleotide sequence 5' of the chicken *c-myc* coding region: localization of a noncoding exon that is absent from *myc* transcripts in most avian leukosis virus-induced lymphomas. *Proc. Natl. Acad. Sci. USA* 81:4697–4701.
 22. Selten, G., H. T. Cuypers, and A. Berns. 1985. Proviral activation of the putative oncogene *Pim-1* in MuLV induced T-cell lymphomas. *EMBO J.* 4:1793–1798.
 23. Selten, G., H. T. Cuypers, M. Zijlstra, C. Melief, and A. Berns. 1984. Involvement of *c-myc* in MuLV-induced T cell lymphomas in mice: frequency and mechanisms of activation. *EMBO J.* 3:3215–3222.
 24. Smith, R., G. Peters, and C. Dickson. 1988. Multiple RNAs expressed from the *int-2* gene in mouse embryonal carcinoma cell lines encode a protein with homology to fibroblast growth factors. *EMBO J.* 7:1013–1022.
 25. Van Ooyen, A., and R. Nusse. 1984. Structure and nucleotide sequence of the putative mammary oncogene *int-1*; proviral insertions leave the protein-encoding domain intact. *Cell* 39: 233–240.
 26. Wilkinson, D. G., S. Bhatt, and A. P. McMahon. 1989. Expression pattern of the FGF-related proto-oncogene *int-2* suggests multiple roles in fetal development. *Development* 105:131–136.
 27. Wilkinson, D. G., G. Peters, C. Dickson, and A. P. McMahon. 1988. Expression of the FGF-related proto-oncogene *int-2* during gastrulation and neurulation in the mouse. *EMBO J.* 7: 691–695.