Cellular DNA Surrounding Integration Sites of an Avian Retrovirus

By RENEE FITTS† AND HOWARD M. TEMIN*
McArdle Laboratory, University of Wisconsin, Madison, Wisconsin 53706, U.S.A.

(Accepted 21 September 1982)

SUMMARY

The size of the direct repeats of cellular DNA next to a spleen necrosis virus (SNV) provirus from infected rat cells and the nature of the cellular DNA surrounding SNV integration sites in chicken DNA were studied. A five-base pair repeat, ATTTT, was observed at the SNV–rat cell junctions. Five of ten SNV proviruses from chicken cells were flanked by unique DNA and five others were flanked by repetitive DNA. No large rearrangements of cellular DNA at SNV integration sites were observed. A clone of uninfected chicken DNA containing the presumptive unoccupied integration site for one SNV provirus was obtained and part of it was sequenced. A small substitution, approximately 60 nucleotides, was observed at the site of virus DNA insertion. Its significance is not known.

INTRODUCTION

During the synthesis of retrovirus DNA, sequences from each end of the RNA genome are duplicated at both ends to generate long terminal repeats (LTRs). Structural and sequence analysis of the LTRs of several different retroviruses (Dhar et al., 1980; Ju & Skalka, 1980; Reddy et al., 1980; Shimotohno et al., 1980; Sutcliffe et al., 1980; Van Beveren et al., 1980; Yamamoto et al., 1980; Donehower et al., 1981; Swanstrom et al., 1981) and the observation that proviruses generate short direct repeats of host DNA upon insertion (Shimotohno et al., 1980; Shimotohno & Temin, 1980; Van Beveren et al., 1982; Hishinuma et al., 1981; Hughes et al., 1981; Majors & Varmus, 1981; Shoemaker et al., 1981) have revealed a remarkable resemblance between retroviruses and transposable elements (Calos & Miller, 1980). It has been proposed that the mechanism of integration of retroviruses may be similar to transposition and that retroviruses evolved from cellular movable genetic elements (Temin, 1980).

Spleen necrosis virus (SNV) is an avian reticuloendotheliosis virus with genus-specific relationships to mammalian type C viruses. When DNA from chicken cells infected with SNV is digested with EcoRI and the resulting fragments are separated by electrophoresis into agarose gels, integrated SNV DNA is found associated with a wide size range of fragments (Keshet & Temin, 1978, 1979). The cell–virus junctions of seven SNV proviruses from infected chicken cells have been cloned and sequenced previously (O'Rear et al., 1980; Shimotohno & Temin, 1980). The length of the cellular repeat sequence for each provirus is five nucleotides, but the sequence repeated is different in each case. These results indicate that SNV can integrate into many different sites in cellular DNA, apparently without regard to the nucleotide sequence at the target site for insertion.

To learn more about the nature of SNV integration, we asked whether the size of the direct repeat adjacent to the provirus is specific for the virus or the species of host cell, whether there is a preference for unique or repetitive DNA at target sites for insertion, and whether there are large or small rearrangements of cell DNA upon viral DNA insertion.

To answer the first question, we cloned an SNV provirus from infected rat cells, sequenced the cell–virus junctions, and found a 5 bp (base pair) repeat, ATTTT. This size is identical to the size of the repeats of SNV proviruses in chickens and, therefore, indicates that the size of the repeat is determined by the virus.

† Present address: Integrated Genetics, 57 New York Ave., Framingham, Massachusetts 01701, U.S.A.
To determine whether large-scale rearrangements occur and to ask about preference for integration in unique or repetitive DNA, we used the cellular sequences adjacent to proviral DNA in ten cloned proviruses from infected chicken cells as hybridization probes for the analysis of uninfected cellular DNA. We found that there do not appear to be any large rearrangements of cellular DNA at SNV integration sites. The DNA adjacent to the provirus is unique in five cases and contains repetitive sequences in five cases.

To study possible fine-scale rearrangements at integration sites, we cloned a region of uninfected chicken cell DNA containing the presumptive integration site for one SNV provirus. Restriction enzyme cleavage site mapping and nucleic acid hybridization demonstrated extensive homology between the clone of uninfected chicken DNA and the cellular sequences adjacent to the provirus. However, DNA sequencing of the clone of uninfected chicken DNA revealed a small substitution, approximately 60 nucleotides, at the site of virus DNA insertion.

METHODS

Cells, viruses and DNA. A library of SNV-infected rat cell DNA in λ Charon 4A was constructed from NRK cells chronically infected with SNV (Keshet & Temin, 1979) as previously described for SNV-infected chicken cells (O’Rear et al., 1980). The single SNV provirus recovered from the library is infectious in DNA transfection assays in chicken cells. Subclones of the 5′ and 3′ cell–virus junctions of this provirus were obtained by ligation of the 5′ and 3′ HindIII fragments into the HindIII site of pBR322 (Fig. 1).

A λ Charon 4A clone containing sequences homologous to chicken DNA adjacent to the SNV provirus in clone 60 (O’Rear et al., 1980) was obtained by screening the chicken DNA recombinant library constructed by Dodgson et al. (1979). Subcloned DNAs of 5′ and 3′ cellular sequences adjacent to SNV proviruses 60, 70, 3–73, 44, 32, and 4 (O’Rear et al., 1980; Shimotohno et al., 1980) were mixed together, nick-translated, and used as a hybridization probe in screening. All 80 positive plaques were further screened with each subclone DNA individually. All were found to hybridize exclusively with clone 60 DNA. This result was surprising, but probably represents some fortuitous preferential growth of these phages during amplification of the library. One of these isolates, λ60-1, contained six EcoRI restriction fragments. Only one of the six EcoRI restriction fragments in this clone contained sequences homologous to cellular DNA in clone 60 (Fig. 2). This fragment was subcloned by ligation into the EcoRI site of pBR322 to give pRF60-13 and was mapped with several restriction enzymes.

DNA for genomic analysis was obtained from uninfected chicken cells as previously described (Chen et al., 1981). The cells used for this purpose were derived from the same embryo used for infection and subsequent isolation of SNV proviruses described by O’Rear et al. (1980).

Enzymes. All restriction enzymes were obtained from New England Biolabs and were used according to the supplier’s recommendations. T4 DNA ligase was purchased from New England Nuclear. Escherichia coli DNA polymerase I Klenow fragment and T4 polynucleotide kinase were obtained from New England Biolabs.

DNA sequencing. Nucleotide sequencing was performed by the method of Maxam & Gilbert (1980). After restriction enzyme digestion, 3′ ends possessing 5′ extensions were labelled at 37 °C for 30 min with the Klenow fragment of E. coli DNA polymerase I in the same digestion buffer plus 100 to 200 μCi of [α-32P]dATP or [α-32P]dCTP and 50 mM concentrations of the appropriate three remaining nucleoside triphosphates. SaeI-digested DNA was labelled in the same way with [γ-32P]dATP by omitting cold nucleotides from the reaction mix. Reactions were stopped by the addition of 0-1 vol. diethylpyrocarbonate (Sigma) at room temperature for 10 min. The 5′ ends of AvaI- or Hinfl-digested DNA were labelled with [γ-32P]dATP and T4 polynucleotide kinase as described by Maxam and Gilbert (1980).

Southern filter hybridization. Gel electrophoresis, mol. wt. determination, and preparation of hybridization probes were performed as described (Chen & Temin, 1980). DNA was transferred from agarose gels on to nitrocellulose filters as described by Southern (1975).

RESULTS

Cell–provirus junction sequences

Previous studies (Shimotohno et al., 1980; Shimotohno & Temin, 1980) have shown that SNV proviruses are flanked by 5 bp repeats of chicken DNA. To determine whether the size of the repeat is specific to the virus, we cloned an infectious SNV provirus from infected rat cells and sequenced the cell–virus junctions. The DNA sequences at the junctions were determined as indicated in Fig. 1 and are shown in Fig. 3. Those regions of the LTR sequenced for this study are identical to the SNV LTR sequence previously described by Shimotohno et al. (1980). A 5 bp sequence of cellular DNA, ATTTT, is repeated at the junctions in rat cells. This size and the
Characteristics of cellular DNA surrounding integration sites

We wished to determine whether large scale rearrangements of cellular DNA occurred as a result of viral DNA insertion and whether there is a preference for unique or repetitive DNA at integration sites. To accomplish these aims, DNA from uninfected chicken cells was digested with EcoRI, electrophoresed through a 0.5% agarose gel, transferred to nitrocellulose, and hybridized to a pool of 10 cloned SNV proviruses which had been nick-translated. These proviruses and adjoining cellular DNA had previously been isolated as EcoRI fragments from SNV-infected chicken cells (O’Rear et al., 1980). Because EcoRI does not cleave within the SNV genome, whole proviruses and adjacent cellular sequences are obtained from EcoRI-digested DNA. The sizes of EcoRI fragments of uninfected cellular DNA containing the ‘unoccupied’ or

high A + T content correlate with the 5 bp repeats reported for seven SNV proviruses from infected chicken cells (Shimotohno & Temin, 1980). The rat DNA sequenced was A + T rich in general, 74% A + T.
Fig. 3. Nucleotide sequences of host-provirus junctions in clones pRaH10 and pRaH13. Only one strand of DNA is shown, that corresponding to the + strand of virus DNA. All sequences are written from 5’ to 3’. A vertical line indicates the junction of cellular and proviral DNAs. The sequence repeated at the junctions, ATTTT, is underlined with an arrow. Restriction enzymes sites used for sequencing are underlined and labelled.

‘preintegration’ sites for these proviruses is predicted by the sum of the lengths of the cellular DNA 5’ and 3’ to the proviruses. Five of the provirus clones, 60, 32, 44, 3-73 and 70, hybridized to single bands of appropriate size, that is, the size predicted by the sum of the 5’ and 3’ cellular sequences in the provirus clones (Fig. 4).

Clone 4 hybridized to two discrete bands, one of predicted size, 6.7 kbp, and one smaller, 6.2 kbp. 5’ and 3’ subclone DNAs of clone 4 hybridized to both bands. Probably the integration event occurred in the 6.7 kbp band and part of the sequence of that band is repeated in the 6.2 kbp band. The remaining clones, 14-44, 36, 63 and 13, gave a long smear when hybridized to genomic DNA, which indicates the presence of repetitive DNA in the probes. An example is shown in lane 14-44 of Fig. 4. Identical results were obtained when pBR322 subclones of 5’ and 3’ cellular sequences of clones 14-44 and 36 were used individually as probes (data not shown).

Sequence of DNA near unoccupied integration site

To determine whether any fine scale rearrangements occur at SNV integration sites, we obtained and characterized a λ Charon 4A clone, λ60-1, of uninfected chicken DNA (from a different chicken) that contained sequences homologous to the presumptive unoccupied integration site for provirus clone 60. An EcoRI fragment of λ60-1 was subcloned into pBR322 to generate pRF60-13. By restriction enzyme site mapping and hybridization to cellular DNA sequences in clone 60, we found that the insert in pRF60-13 has extensive homology with the cellular sequences adjacent to the SNV provirus in clone 60 (Fig. 2). The nucleotide sequence of the region presumed to contain the unoccupied integration site for the provirus was determined. The sequence of cellular DNA 5’ and 3’ to the provirus and the sequence of the corresponding HhaI to HinfI fragment of pRF60-13 are shown side by side in Fig. 5. Reading from the first HhaI site, the two DNA molecules share homology with some base changes for approximately
Retrovirus integration sites

Fig. 4. Southern blot analysis of uninfected chicken DNA. Uninfected chicken DNA was digested to completion with \textit{EcoRI}, electrophoresed through agarose gels, transferred to nitrocellulose filters, and hybridized to nick-translated probes. The SNV proviruses and adjacent cellular DNA used as probes are listed at the top of each lane. The positions of molecular size markers (kbp) are shown at the left. Sizes of cellular DNA sequences 5' and 3' to the provirus in each clone are listed below each lane.

40 nucleotides, at which point the sequences diverge for approximately 60 nucleotides. The remaining 16 nucleotides, with one exception, are the same up to and including the \textit{Hinfl} site. This result indicates a substitution of at least 60 nucleotides surrounding and including the point of provirus insertion. There is no obvious direct or inverted repeat at or near the ends of this substitution.

DISCUSSION

We have investigated several aspects of retrovirus integration: specificity of the size of the direct repeat of cellular DNA; presence of unique or repetitive DNA at integration sites; and possible large or small rearrangements of cellular DNA at integration sites.
Fig. 5. Nucleotide sequence of substituted DNA at the insertion site for SNV provirus. Two sequences are shown for comparison. 60-5' and 60-3' are cellular DNA sequences adjacent to the provirus in clone 60. The 5 bp repeat and presumptive integration site are shown once and boxed. The sequence for the corresponding region from pRF60-13 is shown below that of 60-5' and 60-3'. Only one strand is displayed for each and it is written 5' to 3'. Restriction enzyme sites verified by mapping are denoted by horizontal lines and are labelled. Homologous nucleotides are marked by vertical lines. Sequencing was performed from the HinfI site for both strands of pRF60-13. The sequence of pRF60-13 contained within parentheses is tentative since the sequencing gels contained many compression artefacts. The sequence written is our best approximation from many gels.

To address the first question, we cloned and characterized an SNV provirus from infected NRK cells. This provirus is infectious, but lacks SacI and SalI restriction enzyme cleavage sites present in some other SNV proviruses at 0.76 and 0.86 kbp (O'Rear et al., 1980). When the host-virus junctions of this rat clone were sequenced, a 5 bp repeat of host DNA was found at the ends of the provirus. We assume that this repeat was generated upon insertion of virus DNA, although until the unoccupied integration site from uninfected rat cell DNA is sequenced, we cannot rule out the possibility that integration occurred at a pre-existing tandem repeat in cellular DNA.

The size of this repeat is the same as that found for seven SNV proviruses from chicken cells (Shimotohno & Temin, 1980). This result indicates that the size of the direct repeat is a function of the virus rather than the host cell. The corollary to this experiment of examining one particular virus in more than one cell type is to determine repeat sizes for several different viruses in one cell type. These data are available. In rat cells, Rous sarcoma virus (Hughes et al., 1981) and MMTV (Majors & Varmus, 1981) generate 6 bp repeats, Mo-MuLV causes a 4 bp repeat (Shoemaker et al., 1981), and SNV causes a 5 bp repeat. Furthermore, in chickens ev-1 has a 6 bp direct repeat (Hishinuma et al., 1981). Thus from virus to virus in one cell type the repeat size is different, but for one virus, SNV, in two cell types the repeat size is the same.

To ask about large-scale rearrangements and repetitive DNA, we have analysed uninfected chicken genomic DNA for the size and constitution of EcoRI restriction fragments containing sequences homologous to cellular DNA adjoining SNV provirus DNA. Five of the 10 provirus clones used as probes hybridized to single restriction fragments of length approximately equal to the sum of the 5' and 3' cellular DNA sequences adjacent to the proviruses. Clone 4 hybridized to two EcoRI fragments, one of the expected size and one smaller. The remaining four clones hybridized in broad, smeared patterns, which result from the presence of repetitive DNA in the probes.

We had anticipated the possibility that retroviruses appear to integrate into a large number of sites because they preferentially insert into or near repetitive sequences dispersed throughout the host genome. This does not appear to be true. Half of the provirus clones used in this study hybridized to unique sequence DNA. Although it is conceivable that integration occurs at or near a very small repetitive sequence that might be missed due to the strong single band signal, most repetitive sequences in the chicken are greater than or equal to 2 kb in length (Eden & Hendrick, 1978).
By analogy with transposable elements, there is the possibility that rearrangements of DNA occur as a consequence of viral DNA insertion. In six cases studied here, no major changes (over 0.3 kbp) were evident on the basis of EcoRI restriction fragment size. In the other four instances, the presence of repetitive DNA made that evaluation difficult. Rearrangements may occur in a small region at the site of insertion or at a frequency not detectable in this sample.

Tests were also made to determine whether there was preferential integration at under-methylated sites in chicken DNA or in non-transcribed DNA. DNAs of clones 32, 70, 4, 60, 36, 44 and 14-44 were used as hybridization probes to analyse HpaII- and MspI-digested uninfected cell DNA and to assay total cytoplasmic RNA from uninfected chicken cells in dot blots. No evidence was found either for integration in under-methylated DNA or for integration in DNA without stable transcripts (data not shown).

In any experiments examining specificity at integration sites or frequency of rearrangements at integration sites, there is the potential problem of using a non-representative population of proviruses. For example, in our study eight out of ten proviruses were cloned from chronically infected cells. Since these cells persist in culture for some time after infection, they may result from a selection for non-lethal integration events (Battula & Temin, 1977). Selection of this type or any other type restricts interpretation of these experiments.

To study possible small-scale DNA sequence changes that result from the integration event, we attempted to secure a clone of uninfected chicken DNA containing the presumptive unoccupied integration site in proviral clone 60. Unfortunately, we used a DNA clone library made from a chicken different from that used for preparing clone 60. Since clone 60 represented unique DNA, we foresaw no problems in this difference.

Furthermore, the clone of uninfected cell DNA we secured, λ60-1, contained extensive sequence homology with proviral clone 60 by restriction enzyme mapping and nucleic acid hybridization. However, DNA sequence analysis indicated that a substitution of approximately 60 nucleotides was present in clone λ60-1 at the location where we expected the unoccupied integration site. Therefore, it appears that there may be DNA sequence polymorphism at this site in uninfected chickens. As a result, we can draw no conclusions about possible small-scale DNA sequence changes resulting from the integration event. The presence of such a small change of DNA sequence so near the site of viral DNA insertion is intriguing, but its significance remains unclear.

We can also only speculate about how such a substitution could have occurred. It might have been the result of an insertion of cellular DNA in that region followed by deletion of part of the inserted DNA plus adjoining sequences. These events would generate an apparent substitution.

We thank Tak Mak for construction and screening of the rat library and H. J. Kung, E. Fritsch, and J. D. Engel for providing libraries of chicken DNA. We also thank C. Miller, K. Shimotohno, and K. Wilhelmsen for helpful comments on the manuscript. This investigation was supported by Public Health Service research grants CA-07175 and CA-22443 from the National Cancer Institute. H.M.T. is an American Cancer Society Research Professor.

REFERENCES


(Received 4 May 1982)