Volume 10 Number 10 1982 Nucleic Acids Research

Nucleotide sequences of the murine retrovirus Friend SFFVp long terminal repeats: identification of a structure with extensive dyad symmetry 5' to the TATA box

Stephen P.Clark and Tak W.Mak

The Ontario Cancer Institute, and Department of Medical Biophysics, University of Toronto, 500 Sherbourne St., Toronto, Ontario, Canada M4X 1K9

Received 15 January 1982; Revised 16 March 1982; Accepted 23 April 1982

ABSTRACT

The two long terminal repeats (LTRs) of the integrated provirus of the polycythemia strain of the Friend spleen focus forming virus have been sequenced. Each of the identical LTRs is 514 nucleotides long and together confer on the provirus the features of a transposable element. They are terminated by perfect 11 base-pair inverted repeats, and the entire provirus is flanked by an apparent duplication of host DNA four nucleotides long. The assumed transcription regulatory sequences (Hogness-Goldberg box, CAAT box, polyadenylation signal) can be identified within the LTRs, as well as a region of an imperfect inverted repeat which extends from approximately 140 to 270 nucleotides 5' from the point of transcription initiation. This sequence may form a hairpin-type structure which might have some function in the promotion of transcription.

INTRODUCTION

Retroviruses (RNA tumour viruses) are a family of mammalian and avian viruses capable of inducing malignancies in their hosts (see [1,2] for review). Upon infection, the retrovirus single-stranded RNA genome is converted intracellularly into double-stranded DNA, by the virally encoded enzyme reverse transcriptase, which is subsequently integrated into host nuclear DNA to form a stable provirus (see figure 1). During this process, sequences from each end of the RNA genome become duplicated at both ends of the provirus to form long terminal repeats (LTRs), with the result that the integrated provirus is longer than the RNA genome.

Sequence analysis of retrovirus LTRs has revealed that the proviruses have structures similar to those found in bacterial transposons [3-9]. The LTRs have short inverted repeats of up to 11 bp at their ends, and there are apparent duplications of a few nucleotides of host sequences that flank the provirus. From this and other evidence [10] it has been postulated that the mechanism of integration is similar to transposition, and that perhaps retroviruses evolved from mammalian and avian transposable elements [3].

0305-1048/82/1010-3315$ 2.00/0

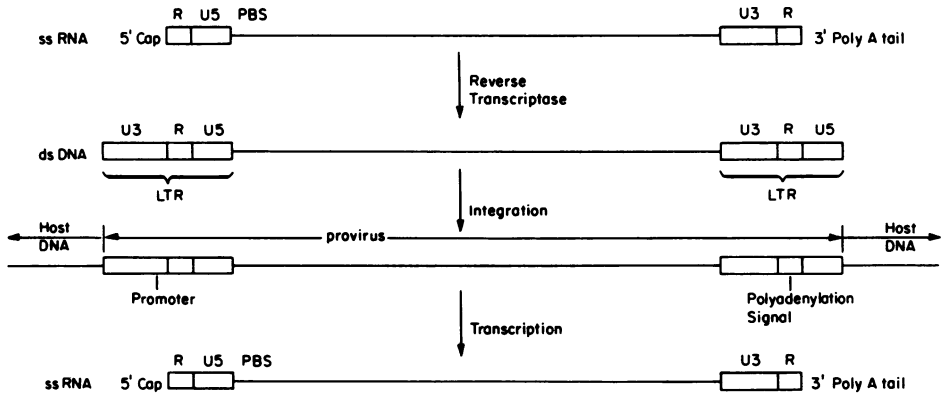


Figure 1: Retrovirus life cycle. R: repeated terminal sequences; U5: unique 5' sequences; U3: unique 3' sequences; LTR: long terminal repeat; PBS: reverse transcriptase primer binding site; ss: single-stranded; ds: double-stranded.

The LTRs are presumed to be responsible for the control of initiation and termination of transcription of the structural genes and regeneration of the RNA genome [11,12]. In vivo, isolated LTRs can promote transcription of genes which have been attached to them [13,14]. Their sequences reveal typical Hogness-Goldberg boxes [15,16] and CAAT boxes [16-18] in the expected orientations for eukaryotic promoter regions. Also present in these LTRs is a sequence identical to the polyadenylation signal [16-19], which is thought to have some function of transcriptional termination in the 3' LTR.

In addition to being responsible for genomic RNA transcription, the promoter properties of the LTRs have been shown to be important in the mechanism of oncogenesis for at least two slowly transforming leukemia viruses, avian leukosis virus (ALV) [20-23] and reticuloendotheliosis virus (REV) [24]. It was found that transcription initiates within the 3' LTR and continues into the flanking cellular sequences. When the provirus is integrated 5' to a cellular gene whose transcription results in activation of the replication machinery, the cell can become transformed. Thus for at least two viruses the LTRs are important in oncogenesis by being responsible for transcription of neighbouring cellular genes, and by possibly being responsible for provirus integration. An understanding of retrovirus LTRs should increase our knowledge of transcriptional promoters.

The polycythemia strain of Friend spleen focus forming virus

($SFFV_P$) is a replication-defective retrovirus that induces a rapid erythroleukemia and polycythemia when injected with its helper virus, Friend murine leukemia virus, into susceptible mice [25]. The integrated form of this provirus has recently been molecularly cloned and found to be biologically active [26]. Sequence analysis of both the 5' and 3' LTRs reveals that, similar to the structures of other retrovirus LTRs, the Friend $SFFV_P$ LTRs confer on the provirus a structure that resembles a transposon, and contain a CAAT box, a Hogness-Goldberg box and a polyadenylation signal. In addition, a region of secondary structure starting 141 bases 5' to the point of transcription initiation, and extending about 130 nucleotides further 5', was identified. Similar, although less stable, structures can also be found in the same area from the sequences of other LTRs, and in other eukaryotic genes. It is possible that a region of non-double-helical DNA may be an important component of some eukaryotic promoters.

MATERIALS AND METHODS

Subcloning of LTRs

Two subclones, each of which contained only one of the LTRs, were generated (Figure 2) from the cloned integrated SFFV provirus [26]. The 7.4 kilobase (Kb) EcoRI insert of λgtWES λ$SFFV_{P502}$ was separated from the lambda arms by preparative agarose gel electrophoresis and electroelution after EcoRI digestion. This fragment was digested with BamHI and the resulting fragments were separated by preparative agarose gel electrophoresis. The 3.7 Kb EcoRI-BamHI fragment, containing the 3' (left) LTR and 2.2Kb EcoRI-BamHI fragment containing the 5' (right) LTR were electroeluted and individually ligated to EcoRI and BamHI cut pBR322. The ligated sequences were used to transform $CaCl_2$ treated HB101 [27,28]. Ampicillin resistant colonies were selected and the presence of the appropriate insert was verified after small-scale plasmid preparations by the boiling method [29]. Large-scale plasmid purifications were by banding in CsCl/ethidium bromide gradients after lysozyme lysis.

DNA Sequencing

Nucleotide sequencing was performed by the Maxam and Gilbert chemical degradation method [30]. Whenever possible, fragments were labelled with $^{32}PO_4$ at the 3' end by incubating restricted ends possessing 5' extensions (approximately 2 pmoles of DNA) with 20-40 μCi of the appropriate [α^{32}P]dNTP (2000-3000 μCi/mmole) and the Klenow fragment of E. coli DNA polymerase I in Hin buffer (20 mM Tris-HCl pH 7.4, 7 mM $MgCl_2$, 60 mM NaCl).

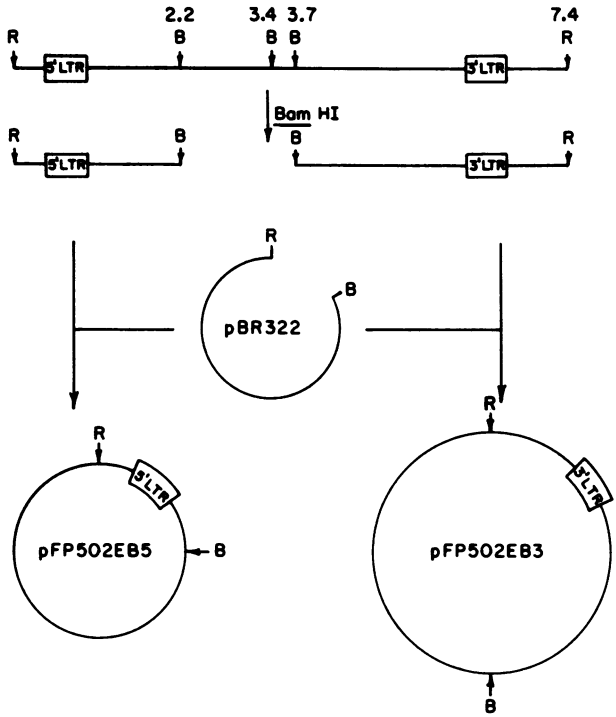


Figure 2: Separation of the two LTRs by subcloning. The EcoR1 (R) and BamHI (B) sites of the 7.4 kilobase fragment are shown on the top with the distances (in kilobases) from the left EcoR1 site shown above them. The resultant plasmids are called pFP502EB5 and pFP502EB3.

Reactions were carried out at 21°C for 15 minutes and chased with 24 molar cold dNTP if necessary for one additional minute (when more than one labelled nucleotide could be added). The 5' ends were labelled after dephosphorylation with T4 polynucleotide kinase and [γ^{32}P]ATP exactly as described [30]. The sequenced fragments were separated on 0.4 mm thick 8%, 12% or 20% polyacrylamide gels (acrylamide:bisacrylamide = 19:1) containing 7 M urea in 100 mM Tris-Borate pH 8.3, 2 mM EDTA. The gels were exposed to Kodak NS-5T films at -20°C.

Identification of Direct and Inverted Repeats

Sequences were examined for direct repeats by making diagonal dot matrix plots, similar to the technique used by Konkel et al. and Efstratiadis et al. [31,32]. Each nucleotide of the sequence is compared to all others and the results are stored in a two-dimensional array. This matrix is manipulated and plotted in a manner such that regions of homology appear as a row of dots that form a diagonal line. The length of the repeat can be determined by the length of the line, and the amount of homology can be partly determined from the number of breaks in the line. In practice, plots are generated which

correspond to 100% or 80% homology to simplify the identification of imperfect repeats.

Inverted repeats are identified by comparing complementary strands of the sequence. Two different sequences can also be compared to determine if there are any homologies or inversions between them.

The energy of formation of the hairpin structures was calculated [33] using the most recent values for base-pairing and loop destabilizing energies [34].

Computer Analysis

The programs used to determine the sequence from individual gel readings were modifications of those supplied by R. Staden [35]. Other programs were written in FORTRAN by S. Clark.

Enzymes

All enzymes were used according to the supplier's recommendations. Calf intestinal alkaline phosphatase, T4 polynucleotide kinase, E. coli DNA polymerase I Klenow fragment, T4 polynucleotide ligase and the restriction enzymes EcoRI, BamHI, BglII and PstI were purchased from Boehringner Mannheim (Montreal, Canada). T4 polynucleotide ligase and the restriction enzymes SstI, Sau3A, KpnI, AvaII and HaeIII were purchased from Bethesda Research Laboratories (Bethesda, Md., USA). The restriction enzyme XmaI was supplied by New England Biolabs (Beverly, Mass., USA).

RESULTS

Sequences of the 3' and 5'Friend $SFFV_p$ LTRs

The restriction maps of the two LTRs are identical, so they were separated by subcloning them individually into plasmids to facilitate sequencing (figure 2). All regions of the two LTRs were sequenced at least once on each strand and all restriction sites were sequenced through. The sequences are illustrated in Figure 3. The top line is the sequence of the 5' LTR and the bottom line represents those of the 3' LTR, and both correspond to the (+) strand of the RNA genome. The vertical lines denote homology between the two strands. The Friend $SFFV_p$ LTRs represent a perfect duplication of 514 nucleotides at the ends of the provirus. On each end of the LTRs is an eleven base-pair long inverted repeat of sequence TGAAAGACCCC at position 1-11 and GGGGTCTTTCA at positions 504-514. Flanking the provirus directly 5' of the left LTR is the rat sequence ATCC. The same sequence can also be found immediately 3' of the right LTR in the rat DNA at position 515-518.

Within the LTRs are also sequences which resemble transcriptional

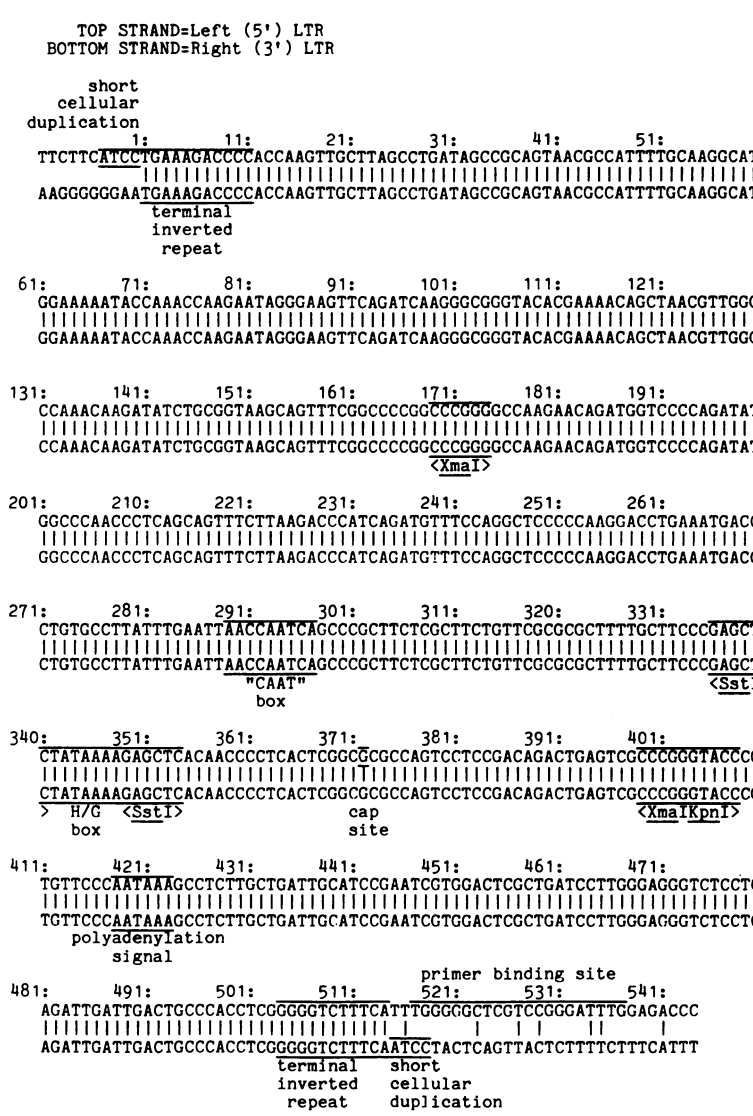


Figure 3: Comparison of the sequences from the 5' (top stand) and the 3' (bottom strand) LTRs. The vertical lines represent homology between the bases. Regions of interest are marked on the sequence. H/G box: Hogness-Goldberg box; cap site: presumed point of transcription initiation.

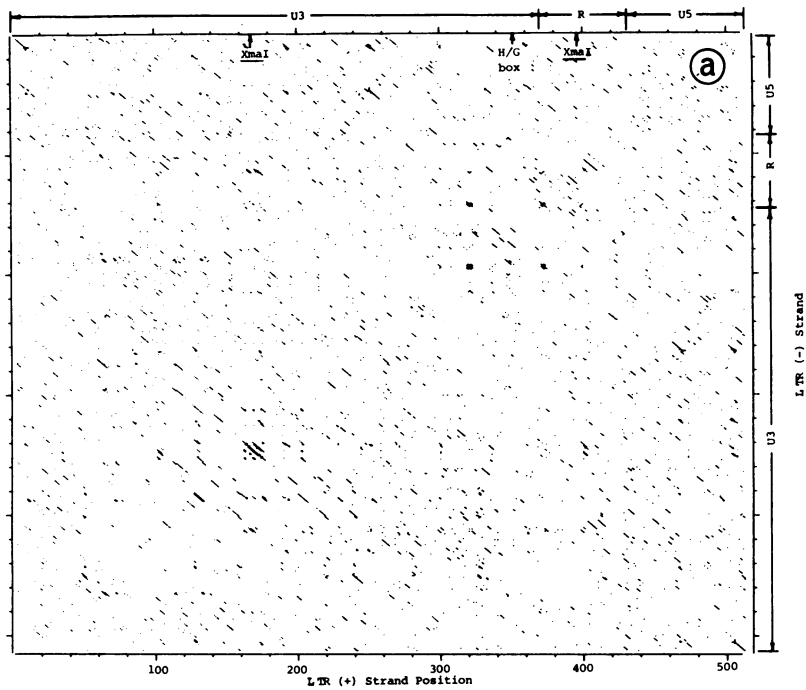
controlling elements found in eukaryotic promoters. The point of transcription initiation as identified by extrapolation from the Moloney murine leukemia virus (MoMuLV) LTR (see Discussion), is located at position

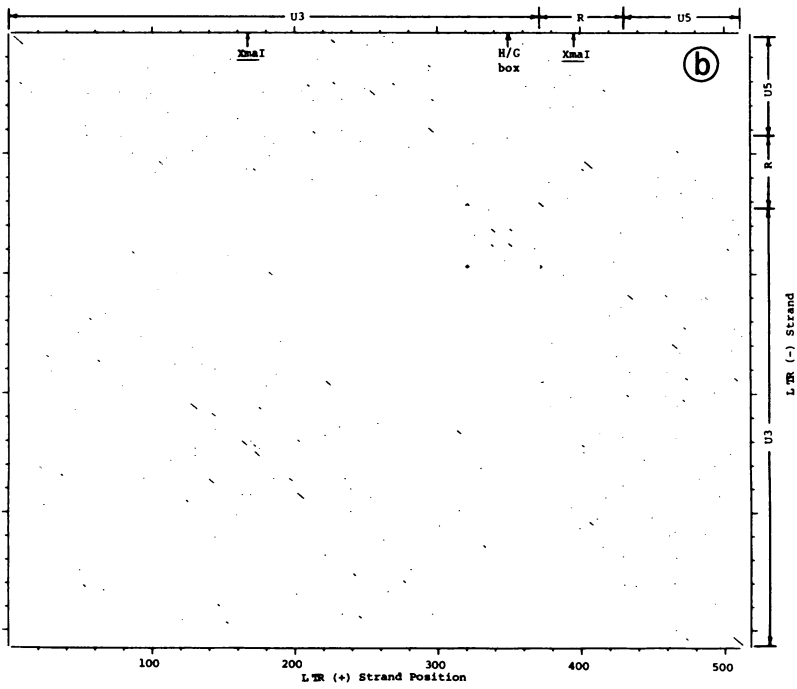
372 of the $SFFV_p$ LTR. Thirty nucleotides upstream from this point at base #341 is the consensus Hogness-Goldberg box with a sequence of TATAAAA [16]. Two SstI sites are situated immediately 5' and 3' to this structure. A further 53 nucleotides upstream is the CAAT box (AACCAATCA compared to the canonical sequence of GGPyCAATCT [16]). These two structures, because of their homology to sequences found in other eukaryotic promoters [16], are presumed to be involved in the initiation of transcription of the Friend $SFFV_p$ viral genome. Fourteen nucleotides upstream from position 432 of the $SFFV_p$ LTR, which corresponds to the 3' end of the MoMuLV RNA genome [11], is the sequence AATAAA, called the polyadenylation signal, which is found near the end of eukaryotic transcripts and is thought to be involved in transcription termination [19]. Another feature which can be found in all retrovirus genomes, the primer binding site, is located two nucleotides 3' to the left LTR at position 517-537. This sequence is complementary to the 3' end of the $tRNA^{pro}$ which is the primer for the initiation of reverse transcription [36].

Analysis of the $SFFV_p$ LTR for direct or inverted repeats

Diagonal dot matrix analysis was used to examine the LTR for direct and inverted repeats. Unlike the Moloney murine sarcoma virus (MoMuSV) LTRs [5,6], no extensive direct repeats within the Friend $SFFV_p$ LTR were found. However, comparison of the (+) strand by diagonal dot matrix analysis to the (-) strand (figure 4) revealed a region from approximately positions 100-230 that forms an extensive imperfect inverted repeat which could form the hairpin structure shown in Figure 5a. The free energy of this structure, assuming that G-T pairs are not stable, is -69 Kcals/mole, relative to the single-stranded form. All other hairpins discussed in this paper could not be easily detected on the plot which required perfect homology (see figure 4 legend).

To determine whether chance alone was enough to result in this structure, five random sequences 500 nucleotides long having the same base composition (±1% per base) as the Friend $SFFV_p$ LTR, and five with the same composition as the hairpin, were generated and analyzed for inverted repeats. In the former case, seven structures were found altogether ranging in stability from -14 Kcals/mole to -51 Kcals/mole (average = -30 Kcals/mole). The two most stable structures had an approximately 10% higher GC content than the $SFFV_p$ hairpin. Of the random sequences with the same composition as the $SFFV_p$ hairpin, six hairpin structures were found with a range of -7 Kcal/mole to -28 kcal/mole and an average of -17 Kcals/mole. Again the most


U3
R
U5
XmaI
H/C box
XmaI
a
U5
R
U3
LTR (-) Strand
100
200
300
400
500
LTR (+) Strand Position


U3
R
U5
XmaI
H/C box
XmaI
b
U5
R
U3
LTR (-) Strand
100
200
300
400
500
LTR (+) Strand Position

stable structures had a 5-10% higher GC content than the LTR hairpin.

DISCUSSION

The sequences of both the 5' and 3' long terminal repeats of the Friend $SFFV_p$ provirus were determined, then analyzed for transcriptional regulatory signals, potential secondary structure, and properties similar to those of bacterial insertion elements.

Transcription Regulatory Signals

The cap site (the point of transcriptional initiation) has not been identified experimentally, so it was determined by comparison to the MoMuLV LTR, for which the points of transcription initiation and termination have been identified [11,12]. Although the Friend SFFV and MoMuLV induce entirely different leukemias in mice and are under the control of different host loci [37,38], the LTRs [4] are 86% homologous between these two viruses, with mostly single base substitutions and only three single base deletions (data not shown). The homology from the CAAT box to the point of transcription termination is 91%. This high degree of homology between the LTRs of the two viruses and the fact that both the Moloney and the Friend LTRs are able to promote the transcription of non-viral genes linked to them [13,14] make it likely that the Friend cap site and termination site are located at positions 372 and 432, respectively. Determination of the cap site allows the Hogness-Goldberg box and the CAAT box to be identified. The latter, at position 291, does not resemble the canonical sequence [16] perfectly (the central 6 out of the 9 total match), but the Hogness-Goldberg box does. It is interesting that this sequence is flanked by two SstI sites. Experiments are in progress to determine the effect of removing or replacing the Hogness-Goldberg box on the promoter properties of the LTR.

The presumed polyadenylation signal is found at position 418. It is not known how transcription can initiate in the 5' LTR and continue past the polyadenylation signal, which is only 47 nucleotides downstream from the cap site. A mechanism to accomplish this has been proposed for the MoMuSV [11].

Figure 4: Diagonal dot matrix analysis of the LTR for inverted repeats. The horizontal axis represents the position on the LTR (+) strand while the vertical axis represents the complimentary strand. Some regions are marked as landmarks. H/G box: Hogness-Goldberg box. A dot on the graph represents homology between the sequences represented on the axes. In (a) up to 20% mismatches are allowed, but none are allowed in (b). The diagonal lines that represent the inverted repeat can most easily seen in (b) towards the lower left, while (a) is more useful for delimiting the boundaries of the repeat.

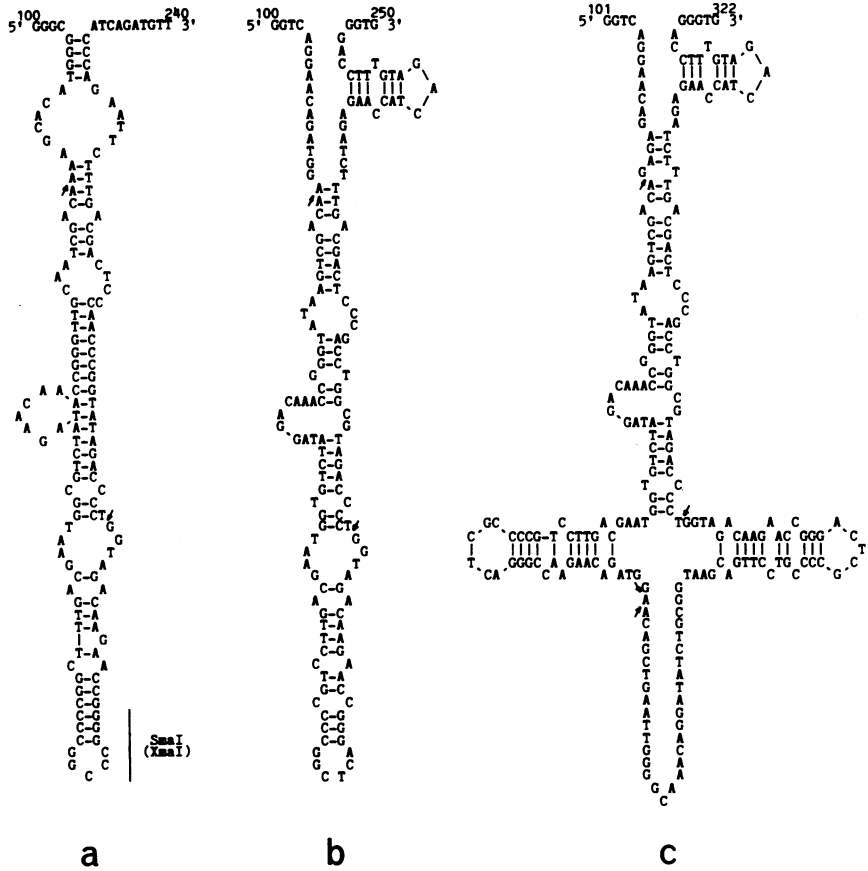


Figure 5: Putative hairpin structures located at identical positions within the Friend and Moloney LTRs. The numbers above the sequences are the distances from the 5' end of the LTR. The start (arrow pointing up) and end (arrow pointing down) of the region that is duplicated in the MoMuSV LTR is marked. (a) Friend SFFVp LTR. (b) MoMuLV LTR [4] which does not contain a large duplication. (c) MoMuSV LTR [5] which contains a 72 base-pair duplication. XmaI and SmaI are isoschizomers which cut at different positions within the recognition sequence.

Resemblance of the SFFV provirus to transposable elements

Bacterial transposable elements are structures several kilobases long which encode a few proteins and are terminated by inverted or direct repeats of several hundred nucleotides (reviewed recently in [39,40]). These terminal repeats are often insertion elements, which are themselves able to be mobilized, and are terminated by perfect or imperfect inverted repeats of several tens of bases long. When either an insertion element or a transposon

integrates into a DNA target sequence, a duplication of a few base pairs is generated at the target, which then flank the transposon. Because of transcriptional promoters and terminators within these elements, they often induce polar mutations. Transposons have also been identified in Drosophila (copia, 412, 297) and yeast (Tyl).

In many respects the Friend provirus resembles a transposable element. The LTRs resemble an insertion element with their 11 bp inverted terminal repeats and transcriptional start and stop sequences. The region between the LTRs codes for at least two proteins [41], and integration leads to an apparent duplication of 4 bp of host DNA. While the data presented here do not prove that this duplication arose as a result of integration (the provirus may have integrated into the middle of the rat sequence ATCCATCC), all integrated proviruses that have been examined are flanked by such duplications [3,5,6,8,9,42] which has been proven to be present only once in the absence of the provirus [8,9,42]. Mutations which interfere with normal gene expression are detected when REV [23] or ALV [22] proviruses integrate next to c-myc (the cellular homolog of the transforming gene of myelocytomatosis virus-29), which results in cellular transformation, or when MoMuLV integrates within an integrated Rous sarcoma provirus, which results in loss of the transformed phenotype [43] or when an endogenous ecotropic murine leukemia virus is present at a specific locus, which results in the dilute coat colour [44]. There is no reason to suspect that SFFV cannot induce similar mutations, but its rapid pathology in vivo makes them difficult to detect.

Analysis of the Friend LTR for Homology to Other Genetic Elements

It was noted [7] that there are short regions of direct or inverted homology between Rous associated virus-2 (RAV-2) and the bacterial insertion element IS2-6 [45] and the eukaryotic elements copia [46], Tyl [47], and SV-40 [48], so these sequences and several chicken retrovirus LTRs [3,7,8], were compared to the Friend LTR by diagonal dot matrix analysis. Many short areas of potential homology were found, but because of their large number, none of them can be considered to be significant without further evidence. However, the Friend LTRs, like many transposable elements [8], end with the dinucleotides TG.....CA.

As mentioned above, the Friend LTR is 86% homologous to the Moloney LTR, while the 11 bp inverted terminal repeats are identical between these two viruses. Analysis of a circular unintegrated MoMuLV has shown [10] that the Moloney inverted repeat is actually 13 bp long, the terminal two nucleotides

presumably being lost during integration. Other retroviruses which have been analyzed in sufficient detail show the same phenomenon. While no data are available on the structure of the unintegrated form of the Friend SFFV, it is likely that its inverted repeat is also 13 bp long prior to integration. The identity of these inverted repeats suggests that a similar or identical protein is involved in the provirus integration for both viruses. This is supported by the observation that MoMuLV can function as a trans-acting helper for Friend SFFV [49], if it is hypothesized that the integration protein is encoded by the virus genome.

Inverted Sequence Upstream from the CAAT Box

Figure 5a shows a hairpin structure which can be drawn between positions 104 and 230. Whether or not such a structure (or a related one) could exist in the chromatin containing the provirus is not known. It is unlikely that this structure can be explained simply as a result of chance, as shown by the lack of such extensive structures in random sequences with the same base composition. The most stable structures that were found in the random sequences had a much higher GC content than the postulated hairpin in the LTR (GC pairs are twice as stabilizing as AT pairs) and were even then less stable than the LTR hairpin. In view of this, it is possible that this proposed structure provides some function for the virus. Further evidence that such hairpins may serve some purpose is the finding that the region corresponding to bases 117-189 is duplicated in MoMuSV [5,6] (see arrows in figure 5). This duplication results in a cloverleaf type structure rather than a hairpin, as shown in figure 5c. The hairpin structure of MoMuLV LTR, which is the same length as the Friend LTR and lacks this 72 base-pair duplication, is shown in figure 5b for comparison.

On purely thermodynamic grounds, it would seem unlikely at first that such a structure exists. The free energy of -138 Kcals/mole (-69 Kcals/mole per strand) seems small compared to the approximately -350 Kcals/mole for the normal heteroduplex, each relative to two single strands. However, these calculations are based on values derived for oligonucleotides which have no torsional constraints and are devoid of all proteins. Neither of these conditions apply to chromatin, where DNA is negatively supercoiled and where proteins can bind, to either stabilize or destabilize such structures. It has been found that negatively supercoiled plasmid DNA which contains regions potentially able to form 9-13 base pair long stems with 3-6 base loops are cut by S1 nuclease precisely within the loop, while the same DNA in the relaxed state was not cut [50]. This indicates that stem-loop

structures can exist in negatively supercoiled DNA probably because they relieve the supercoiling somewhat.

The role of protein-DNA interactions on the hairpin of figure 5a may be difficult to assess. It is known that the chromatin upstream from some genes is hypersensitive to DNAse I, indicating some structural change of this DNA compared to other chromatin. In the case of the transcriptionally active chicken endogenous virus-3, the DNAse I hypersensitive site maps to a region apparently within 100 base-pairs upstream from the Hogness-Goldberg box in both LTRs (A. Larsen and M. Groudine, personal communication). The chromatin is also S1 nuclease sensitive in this region, a strong indication that this DNA does not exist in the normal double-helical conformation. These results can be explained by a hairpin stabilized by protein.

If the structure shown in figure 5a exists and is part of the promoter, one would expect to find similar structures (but probably with different sequences) in the promoter regions of other genes. Table 1 lists stabilities of some of the hairpins which can be found, using the method described in materials and methods, and their distances from the cap site, for some retrovirus LTRs and eukaryotic genes. It is possible that many of these genes can also form hairpin structures, with many of them starting 100-150

Table 1: Positions and stabilities of potential hairpin structures for several eukaryotic promoter regions.

Sequence	Distance from cap site to base of stem (nucleotides)	Stability (Kcal/mole)	Source of sequence data
SFFV LTR	140	-69	this paper
Spleen necrosis virus LTR	140-150	-45	[3]
MoMuLV LTR	150	-39	[4]
RAV-2 LTR	100	-29	[7]
Chicken endogenous virus-1 LTR	80	-11	[8]
	80	-54	[51]
Chicken α2 type I collagen	50	-54	[51]
	70	-46	[51]
	35	-51	[51]
Rabbit β-globin	100	-42	[52]
Mouse α-globin	30	-39	[53]
Mouse β-globin (major)	70	-31	[31]
Sea urchin Histone H2A	60	-30	[54]
Goat β-globin	140	-30	[55]
Human ε-globin	125	-28	[56]
Mouse β-globin (minor)	55	-25	[31]
Drosophila 70K heat shock protein (G13)	150	-22	[57]
Chicken conalbumin	115	-22	[58]

nucleotides from the cap site (approximately 140 base-pairs of DNA are wound on one nucleosome [59]). The Friend $SFFV_p$ LTR, however, appears to be the most stable structure when compared to these other putative hairpin structures. No attempt was made to determine the significance of the structures in other promoters by analyzing random sequences with comparable base compositions.

Not all 5' flanking regions are found to have easily identifiable hairpins. Perhaps hairpins are not the only structures associated with promoters (cf the Moloney LTR which contains a 72 base repeat - it can form a semi-cloverleaf structure (figure 5c)), or perhaps some genes do not require any such structure. Many of the genes examined were members of multigene families, where it was uncertain that the genes identified were actually expressed ([57] and data not shown).

The inverted repeat found in the Friend $SFFV_p$ LTR is suggested to form a hairpin in the DNA, although it is possible that it has some function in the RNA genome, such as a signal for transcription termination [11], or perhaps a structural role. It may even serve two purposes: one in the RNA and one in the DNA.

Function of Secondary DNA Structure in Promoters

One proposed mechanism of promoter function is to "phase" the nucleosomes [60] so that they are aligned properly to expose regulatory signals. A hairpin would be an ideal structure for this purpose. It may not be a coincidence that the base of the hairpin for the Friend LTR is exactly one nucleosome distance from the cap site. Other investigators have pointed out the possibility that hairpin structures can be drawn in the 5' flanking regions of certain genes, but these were associated with the cap site [51,53,61] Hogness-Goldberg box [61] or CAAT box [51]. It would be interesting to determine if there is a correlation between hairpin stability or position and promoter strength. Note that the mouse β-globin major gene has a more stable structure than the mouse β-globin minor gene (Table 1). A hairpin structure in the DNA with occasional bulges would provide many places for regulatory molecules to bind, and since each hairpin could have a different structure, it would be easy for a regulatory molecule to find the correct promoter.

ACKNOWLEDGEMENTS

We thank C. Astell and M. Smith for initial help with the Maxam and Gilbert sequencing, and Y. Yamamoto and A. Bernstein for the λgtWESλ

$SFFV_{P502}$ preparation. S. Clark is a recipient of a studentship from the Medical Research Council of Canada. This work was supported by grants from the National Cancer Institute of Canada and the Medical Research Council of Canada.

REFERENCES

1. Bishop, J.M. (1978) in Ann. Rev. Biochem., Snell, E.S., Boyer, P.D., Meister, A. and Richardson, C.C. Eds., Vol. 47,, pp 35-88, Annual Reviews Inc., Palo Alto, Calif.
2. Coffin, J.M. (1979) J. Gen. Virol. 42, 1-26.
3. Shimotohno, K., Mizutani, S. and Temin, H.M. (1980) Nature 285, 550-554.
4. Van Beveren, C., Goddard, J.C. Berns, A. and Verma, I.M. (1980) Proc. Natl. Acad. Sci. 77, 3307-3311.
5. Dahr, R., McClements, W.L., Enquist, L.W. and Vande Woude, G.F. (1980) Proc. Natl. Acad. Sci. 77, 3937-3941.
6. Reddy, E.P., Smith, M.J., Canaani, E., Robbins, K.C., Tronick, S.R., Zain, S. and Aaronson, S.A. (1980) Proc. Natl. Acad. Sci. 77, 5234-5238.
7. Ju, G. and Skalka, A.M. (1980) Cell 22, 379-386.
8. Hishinuma, F., DeBona, P.J., Astrin, S. and Skalka, A.M. (1981) Cell 23, 155-164.
9. Majors, J.E. and Varmus, H.E. (1981) Nature 289, 253-258.
10. Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S.W. and Baltimore, D. (1980) Proc. Natl. Acad. Sci. 77, 3932-3936.
11. Benz, Jr., E.W., Wydro, R.M., Nadal-Ginard, B. and Dina, D. (1980) Nature 288, 665-669.
12. Fuhrman, S.A., Van Beveren, C. and Verma, I.M. (1981) Proc. Natl. Acad. Sci. 78, 5411-5415.
13. McClements, W.L., Dhar, R., Blair, D.G., Enquist, L., Oskarsson, M. and Vande Woude, G.F. (1980) Cold Spring Harbor Symposia on Quant. Biol. 45, 699-705.
14. Joyner, A., Yamamoto, Y. and Bernstein, A. (1981) Proc. Natl. Acad. Sci., in press.
15. Goldberg, M.L. (1979) Ph.D. Thesis, Stanford University.
16. Chambon, P. and Breathnach, R. (1981) in Ann. Rev. Biochem., Snell, E.S., Boyer, P.D., Meister, A. and Richardson, C.C. Eds., Vol. 50, pp 349-383.
17. Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142.
18. Liebhaber, S.A., Goossens, M.J. and Wai Kan, Y. (1980) Proc. Natl. Acad. Sci. 77, 7054-7058.
19. Proudfoot, N.J. and Brownlee, G.G. (1974) Nature 252, 359-362.
20. Payne, G.S., Courtneidge, S.A., Crittenden, L.B., Fadly, A.M., Bishop, J.M. and Varmus, H.E. (1981) Cell 23, 311-322.
21. Neel, B.G., Hayward, W.S., Robinson, H.L., Fang, J. and Astrin, S.M. (1981) Cell 23, 323-334.
22. Hayward, W.S., Neel, B.G. and Astrin, S.M. (1981) Nature 290, 475-480.
23. Fung, Y.-K.T., Fadly, A.M., Crittenden, L.B. and Kung, H.-J. (1981) Proc. Natl. Acad. Sci. 78, 3418-3422.
24. Noori-Daloii, M.R., Swift, R.A., and Kung, H.-J. (1981) Nature 294, 574-576.
25. Mak, T.W., Gamble, C.L., MacDonald, M.E. and Bernstein, A. (1980) Cold Spring Harbor Symposia on Quant. Biol. 44, 893-899.
26. Yamamoto, Y., Gamble, C.L., Clark, S.P., Joyner, A., Shibuya, T.,

Mager, D., MacDonald, M.E., Bernstein, A. and Mak, T.W. (1981) Proc. Natl. Acad. Sci., 78, 6893-6897.
27. Boyer, H.W. and Roulland-Dussoix, D. (1960) J. Mol. Biol. 41, 459-472.
28. Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) Proc. Natl. Acad. Sci. 69, 2110-2114.
29. Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193-197.
30. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds., Vol. 65, pp 499-560, Academic Press, New York.
31. Konkel, D.A., Maizel, J.V. and Leder, P. (1979) Cell 18, 865-873.
32. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Sprite, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Barelle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653-668.
33. Tinoco, Jr., I., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.L., Crothers, D.M. and Gralla, J. (1973) Nat. New Biol. 246, 40-41.
34. Salser, W. (1977) Cold Spring Harbor Symposia on Quant. Biol. 62, 985-1002.
35. Staden, R. (1980) Nucl. Acids Res. 8, 3673-3694.
36. Peters, G. and Dahlberg, J.E. (1979) J. Virol. 31, 398-407.
37. Steeves, R. and Lilly, F (1977) Ann. Rev. Genet. 11, 277-296.
38. Pincus, T. (1980) in Molecular Biology of RNA Tumor Viruses, Stephenson, J.R., Ed., pp. 77-130.
39. Starlinger, P. (1980) Plasmid 3, 241-259.
40. Miller, J.H. and Calos, M.P. (1980) Cell 20, 579-595.
41. MacDonald, M.E., Reynolds, Jr., F.H., Van de Ven, W.J.M., Stephenson, J.R., Mak, T.W. and Bernstein, A. (1980) J. Exp. Med. 151, 1477-1492.
42. Hughes, S.H., Mutschler, A., Bishop, J.M. and Varmus, H.E. (1981) Proc. Natl. Acad. Sci. 78, 4299-4303.
43. Varmus, H.E., Quintrell, N. and Ortiz, S. (1981) Cell 25, 23-26.
44. Jenkins, N.A., Copeland, N.G., Taylor, B.A. and Lee, B.K. (1981) Nature 293, 370-374.
45. Ghosal, D. and Saedler, H. (1978) Nature 275, 611-617.
46. Levis, R., Dunsmuir, P. and Rubin, G.M. (1980) Cell 21, 581-588.
47. Farabaugh, P. and Fink, G.R. (1980) Nature 286, 352-356.
48. Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L. and Weissman, S.M. (1978) Science 200, 494-502.
49. MacDonald, M.E., Johnson, G.R. and Bernstein, A. (1981) Virol. 110, 231-236.
50. Lilley, D.M.J. (1981) Nucl. Acids Res. 9, 1271-1289.
51. Vogeli, G., Ohkubo, H., Sobel, M.E., Yamada, Y., Pastan, I. and De Crombrugghe, B. (1981) Proc. Natl. Acad. Sci. 78, 5334-5338.
52. Dierks, P., Van Ooyen, A., Mantei, N. and Weissmann, C. (1981) Proc. Natl. Acad. Sci. 78, 1411-1415.
53. Nishioka, Y. and Leder, P. (1979) Cell 18, 875-882.
54. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. 77, 7102-7106.
55. Haynes, J.R., Rosteck, P. and Lingrel, J.B. (1980) Proc. Natl. Acad. Sci. 77, 7127-7131.
56. Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 621-626.
57. Ingolia, T.D., Craig, E.A. and McCarthy, B.J. (1980) Cell 21, 669-679.
58. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. and Chambon, P. (1979) Nature 282, 567-574.
59. Felsenfeld, L. (1978) Nature 271, 115-122.
60. Igo-Kemenes, T. and Zachau, H.G. (1981) Cell 24, 597-598.
61. Ziff, E.B. and Evans, R.M. (1978) Cell 15, 1463-1475.