Volume 10 Number 15 1982 Nucleic Acids Research

Use of a cellular polyadenylation signal by viral transcripts in polyoma virus transformed cells

H.Earl Ruley, Luigi Lania*, Furzana Chaudry and Mike Fried

Department of Tumour Virus Genetics, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK, and *Istituto di Biologia Generale e Genetica, Universita di Napoli, Napoli, Italy

Received 26 May 1982; Revised and Accepted 2 July 1982

ABSTRACT

The DNA sequences at and around the junctions between viral and cellular DNA in the polyoma virus transformed mouse cell line, TS-A-3T3, have been determined. No common sequence specificity or structural features at the joins have been observed. The sequence indicates that the 94K truncated large T antigen found in TS-A-3T3 cells is a hybrid protein in which the carboxy-terminal 19 amino acids are encoded by adjacent host sequences. Moreover, the three early region transcripts initiated in viral sequences are also hybrid in nature and appear to utilize a host polyadenylation signal associated with the hexanucleotide, AATAAA, found 100 bp beyond a viral-host join.

INTRODUCTION

Cell transformation by polyoma virus (Py) is accompanied by integration and expression of viral DNA sequences in the genome of the host cell (1). Viral transformed cells invariably retain and express region of the Py genome encoding the viral (55K) middle and (22K) small T antigens (T-ags), but are often unable to make a functional viral (100K) large T-ag (2-7). Frequently the inability to make a functional large T-ag has resulted from recombination with cellular DNA and the loss of sequences encoding the carboxy terminal end of the large T-ag. A further consequence of recombinant joins of this type is the loss of the major polyadenylation signal utilized by the three viral early region mRNAs (8), thus allowing viral transcripts to extend into the adjacent cellular DNA. In some transformed cell lines the viral transcripts may utilize a minor polyadenylation site, located at 99 map units on the viral genome (8). Alternatively, viral transcripts may utilize polyadenylation signals located in the adjacent cellular DNA (8,9). Transformed cells of the latter type may reflect instances in which the viral DNA has inserted into cellular genes.

In this report we describe experiments further characterizing the

0305-1048/82/1015-4515$2.00/0

viral inserts of the Py transformed mouse cell line, TS-A-3T3 (Balb/C 3T3 cells transformed by the Py TS-A mutant (3)). In TS-A-3T3 DNA, viral sequences are present in two inserts of 4Kb and 0.15Kb, separated by approximately 600 bp of cellular DNA (3). The 4Kb insert includes portions of the early region able to code for the middle and small T-ags, however, sequences encoding the carboxy-terminal region of the Py large T-ag as well as the major viral polyadenylation site for the viral early region transcripts have been lost as a result of the join between viral and cellular DNA. Poly A-containing cytoplasmic RNAs of the three early region transcripts in TS-A-3T3 cells extend approximately 100bp into the adjacent cellular DNA (8). In this report the DNA sequences at and around three viral-host joins have been determined, including the region of cellular DNA defining the 3' extent of the viral transcripts. The sequence supports the conclusion that viral transcripts utilize a host polyadenylation signal found approximately 100bp beyond the viral-host join.

MATERIALS AND METHODS

DNA sequence analysis

Cloning of the 7.5Kb fragment containing two blocks of viral DNA into λgtWESλB has been previously described (3). This 7.5Kb fragment and the 1.4 and 6.1Kb fragments generated from it by BamHI cleavage were subcloned into the plasmid pAT153 for sequencing studies. Restriction endonuclease cleavage sites were 5'-end labelled following treatment with calf intestine alkaline phosphate (Boehringer Mannheim) and T4 polynucleotide kinase (P-L Biochemicals) (10). Following secondary enzyme cleavage, labelled fragments were purified from polyacrylamide gels and sequenced by the chemical degradation method of Maxam and Gilbert (11). The sequence was confirmed either by sequencing the opposite strand or in replica experiments. The sequences were analysed for homologies to the Py genome and for repeats, symmetries, dyad symmetries and biased base compositions by the SEQ computer program (12).

Analysis of T-antigen species

The conditions of cell labelling, extraction and immunoprecipitation of TS-A virus infected 3T3 cells have been previously described (13). The transformed TS-A-3T3 cells were grown to semiconfluence in a 5cm dish, and labelled for 3hr at 37°C with ^{35}S-methionine 300μCi/dish (800-1000 Ci/mMole; Radiochemical Centre, Amersham) in Dulbecco's modified Eagle's medium without methionine. After labelling, the cells were detached with

versene, and washed twice with ice-cold Tris-Saline. Subsequent extraction and immunoprecipitation of TS-A-3T3 cells were identical to that of lytically infected cells.

RESULTS

Structure of viral inserts in TS-A-3T3

The arrangement of Py sequences in TS-A-3T3 cellular DNA as characterized by nucleic acid hybridization, molecular cloning, heteroduplex analysis and by DNA sequence analysis is shown in Figure 1. We have previously reported the cloning of the rightward 7.5 Kb EcoRI fragment in λgtWESλB after size fractionation of EcoRI cleaved TS-A-3T3 DNA (3). This fragment was shown to contain two blocks of viral DNA separated by approximately 600 bp of cellular DNA. This 7.5Kb fragment and the two fragments generated from it by cleavage with BamHI were further cloned into the plasmid pAT153 for DNA sequencing studies. A portion of the map in

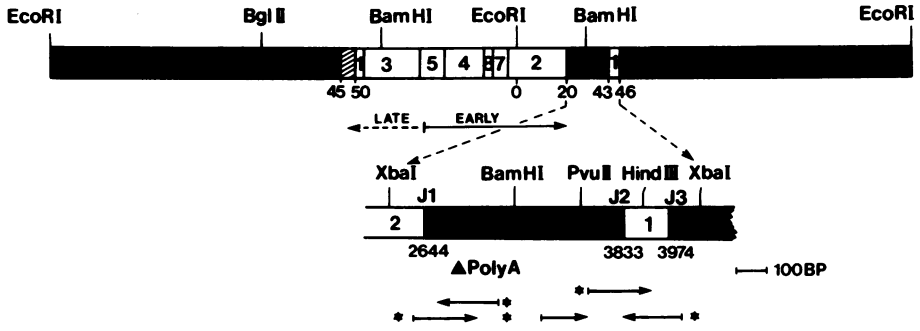

Figure 1 Structure of polyoma virus insert in TS-A-3T3 and strategy for DNA sequencing studies. The arrangement of polyoma virus sequences in TS-A-3T3 cells was characterized by nucleic acid hybridisation, molecular cloning and heteroduplex studies (3) and by DNA sequence analysis (see Figs. 2,4). In the upper portion of the figure, viral sequences (shown in white) are drawn in relation to the eight HpaII cleavage sites in the Py genome, and the map coordinates of the viral sequences present in TS-A-3T3 are indicated (14). Host sequences are shown in black and the region on the left in which the viral sequences joins the host (precise join not yet known) is cross-hatched. A portion of the map is drawn below in expanded scale to indicate the strategy for DNA sequencing. Restriction endonuclease cleavage sites were 5'-end labelled, and the arrows indicate the direction and extent of the sequences determined from each site. The sequence was confirmed either by sequencing the opposite strand or in replica experiments. The three sequenced viral-host joins are indicated (J1, J2 and J3), and the number below each join refers to the last nucleotide colinear with the sequence of the viral insert, based on the numbering system of Soeda et al. (15).

Figure 1 is drawn in expanded scale to indicate the strategy for Maxam-Gilbert DNA sequencing studies.

Analysis of viral-host junctions in TS-A-3T3

The sequence through the three viral and cellular junctions J1, J2 and J3 was determined (Fig. 1). The sequence data confirm the structure of the viral insert in TS-A-3T3 as previously determined and are shown in Figure 2 (3). Viral early region sequences are joined to host cell DNA at nucleotide 2644 (J1), and the small insert from late region is composed of continuous viral sequences extending from nucleotide 3833 (J2) to nucleotide 3974 (J3). The sequences were analysed by the SEQ computer programme (12) for homologies to either strand of the Py genome and for structural features such as Dyad symmetries, inverted or direct repeats and for unusual or biased base compositions. The Py sequences at the sites of joining to cellular DNA were similarly analysed. At each of the three junctions, no common sequence specificity or structural feature in either the viral or cellular sequence is observed.

Finally, a comparison of the flanking cellular and displaced viral DNA sequences (i.e. viral DNA sequences adjacent to the viral inserts which have been replaced by cellular DNA) reveals that the transition from viral to cellular sequence is abrupt, while beyond the join there are small patches of homology between the flanking cellular and displaced viral DNAs (Fig. 2). The significance of these homologies was tested both by

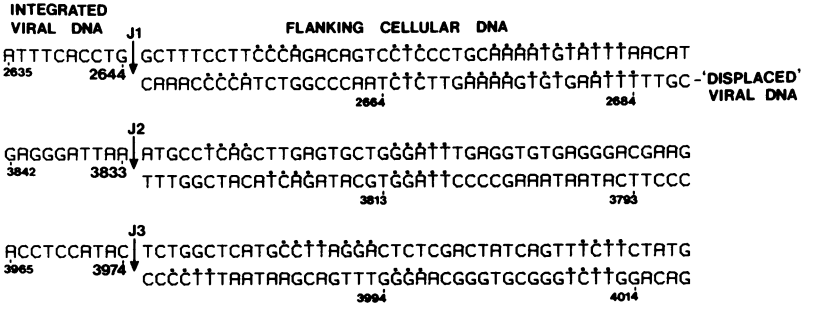

Figure 2 DNA sequence through three viral-host joins in TS-A-3T3 cells. The sequences of the three junctions (J1, J2, J3) between viral and cellular DNA indicated in Figure 1 are shown. The sequence flanking the viral insert (flanking cellular DNA) is drawn above the displaced viral DNA (i.e. viral sequences which are no longer present as a result of recombination between viral and cellular DNA). The over-scored nucleotides indicate homologous patches of three or more nucleotides found between the flanking cellular and displaced viral DNAs.

comparing randomly chosen sequences and by the probability algorithm employed by the SEQ program (12,16). None of the homologies shown in Figure 2 is statistically significant. Thus even the more extensive homology found between the flanking cellular and displaced viral DNAs beyond J1 would be expected to occur by chance at least once if ten pairs of random sequence forty nucleotides long were compared (E>.1).

A hybrid truncated large T-Antigen is synthesised as a result of the join between viral and cellular DNAs

Immunoprecipitation of labelled proteins from TS-A-3T3 with anti-tumour sera reveals a 94K-reacting species in addition to the viral middle and small T-ags (Fig. 3). As TS-A-3T3 cells do not produce the Py

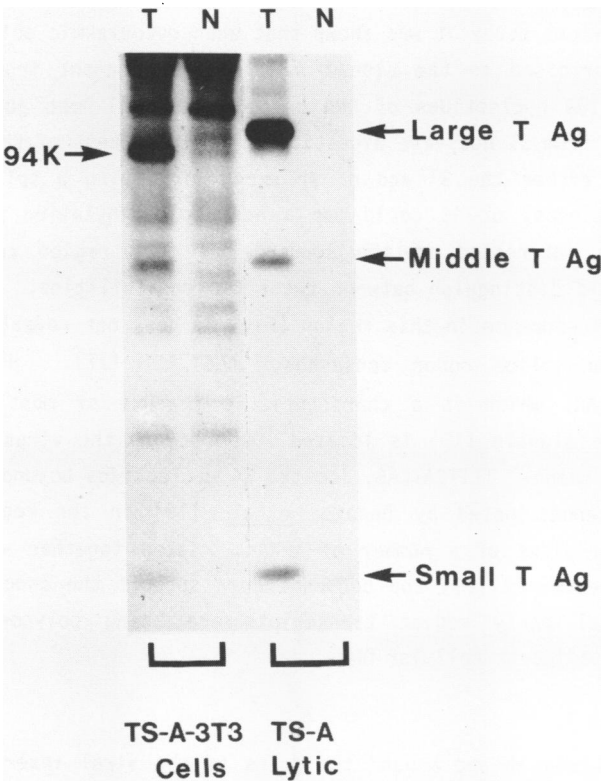

Figure 3 T-Antigens in TS-A-3T3 cells. Cell extracts labelled with ^{35}S methionine from TS-A virus infected mouse 3T6 cells or TS-A-3T3 cells incubated either with normal serum (N) or antitumour serum (T) were precipitated with S.aureus and fractionated on 10% SDS-polyacrylamide gels. The truncated TS-A-3T3 94K large T-Ag and the Py 100K large, 55K middle and 22K small T-Ags are indicated.

100K large T-ag, it seemed likely that the 94K protein represents a truncated large T-ag species resulting from the join with cellular DNA. Analysis of the sequence beyond the join with the Py early region (J1) indicates that the coding sequence for the carboxy terminal 85 amino acids of large T has been deleted and the reading frame for the large T protein remains open for an additional 56 nucleotides before an inframe termination codon (TGA) is encountered (Fig. 4). Thus the DNA sequence predicts that the 94 T-ag species found in TS-A-3T3 is both a truncated large T-ag and a hybrid protein, containing 695 and 19 amino acids derived from viral and host sequences, respectively.

The 3' extent of viral transcripts in TS-A-3T3 is located near the cellular sequence, AATAAA

In a previous study it was shown that when cytoplasmic polyadenylated mRNAs were hybridised to the cloned 7.5Kb EcoRI fragment from TS-A-3T3, approximately 100 nucleotides of DNA beyond the viral-host junction (J1) were protected from S1 nuclease digestion (8). The extent of the protected DNA could map either the 3' end of an exon, signifying a splice further into host sequences, or it could map a host polyadenylation signal. We were therefore interested in the sequence of this region to determine whether we could distinguish between these two possibilities. Examination of the host DNA sequence in this region (Fig. 4) does not reveal a sequence related to the splice donor consensus, $\overset{A}{C}AGGT\overset{A}{G}AGT$ (17). However, the sequence, AATAAA, which is a characteristic feature of most eucaryotic polyadenylation signals (18) is located 100bp beyond the virus-host join. Further, the sequence, TTTTCATCAG, located 24 nucleotides beyond the AATAAA resembles sequences noted by Benoist et al. (19) in the region of the polyadenylation sites of a number of mRNAs. Taken together with the S1 protection experiments (8), the sequence data support the conclusion that the three viral early region transcripts are being polyadenylated by signals in the adjacent cellular DNA.

DISCUSSION

The sequences at and around the joins of the viral inserts in TS-A-3T3 cellular DNA have been determined in order to correlate structure and expression of viral sequences in Py transformed cells. This analysis has revealed that the 94K T-ag species found in TS-A-3T3 cells is a hybrid truncated large T-ag, containing 19 amino acids which are encoded by cellular sequences. Moreover, the three early region mRNAs are also hybrid

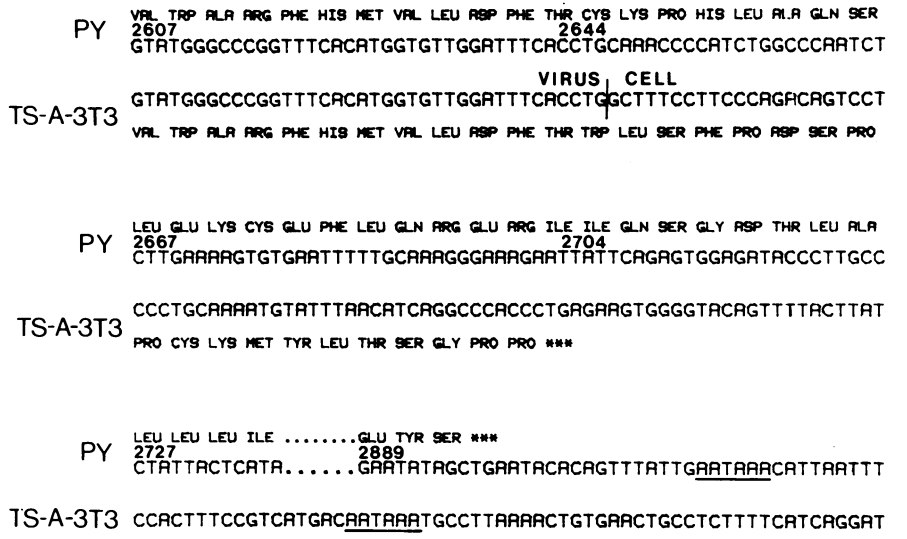

Figure 4 Sequence of adjacent cell DNA transcribed in TS-A-3T3 cells. Viral early region transcripts extend approximately 100 nucleotides past the viral-host join, J1 (see Fig. 1), and the sequence of this region is shown (TS-A-3T3). Py indicates the sequence of the Py TS-A mutant (M. Jones, personal communication) which was the transforming virus used to generate the TS-A-3T3 cell line (3). The translation products in the large T-antigen reading frame are shown below the TS-A-3T3 and above the Py DNA sequences. A comparison of these two sequences shows that the viral-host join is located at nucleotide 2644 in the Py genome, interrupting the coding region for the viral large T-antigen. The cellular sequences remain open in the large T-Ag reading frame for an additional 19 amino acids as shown. The underlined TS-A-3T3 sequence, AATAAA, is located 100 nucleotides beyond the virus host junction.

in nature and appear to utilise a host polyadenylation signal associated with the hexanucleotide, AATAAA, located 100bp beyond a viral-host join. In TS-A-3T3 DNA, viral sequences are present in two inserts of 4Kb and 0.15Kb separated by approximately 600bp of cellular DNA. The present analysis of the viral inserts does not make it clear how this arrangement was generated. For example, independent integrations of two pieces of viral DNA may have occurred at the same site in the host genome. Alternatively, a piece of host DNA, either before or after integration, may have recombined with a continuous stretch of viral sequence. The sequence through the viral-cell junctions reported in this study give no indication that a sequence-specific mechanism lies behind their formation. The sequences lack direct repeats characteristics of transposition elements (20) and lack structural features such as symmetries, specific sequences or

base compositions which might facilitate recombination.

The sequences determined at the three viral host junctures show an abrupt transition from viral to host sequences. As noted in Fig. 2, there are small patches of homology between displaced viral sequences and the flanking cellular DNA. These patches of homology raise questions as to whether the sequences are virus-derived as a result of patchwork recombination between viral and cellular DNAs, or whether the sequences played a role in recombination by providing stretches for homologous pairing. An answer to the first question requires cloning and sequencing the site in untransformed 3T3 cellular DNA into which the viral DNA has inserted. While it is not possible to answer the second question directly, it is noteworthy that none of the homologies shown in Fig. 2 is statistically significant (i.e. similar homologies are found if randomly chosen sequences are compared). Consequently, if recombination depended on such small homologies, this would not greatly limit recombination between sequences, and there would be a large number of sites in the host DNA into which viral DNA could become integrated. The sequences through the viral-host junctions are indicative of a non-specific or illegitimate mechanism of recombination. Similar results have been observed for the joins between Py and host sequences in other Py transformed cell lines (H.E. Ruley, unpublished results; Hayday et al., manuscript in press) and also for viral-host junctions in SV40 transformed cell lines (21,22).

Infection of mouse cells with Py virus generally results in cell death since mouse cells are permissive for Py replication (1). As a result, isolation of Py transformed mouse cell lines imposes a strong selection against viral replication. Frequently, this is accomplished by the loss of sequences specific for the viral replication protein, large T-ag, as a result of recombination with host DNA. Such is the case for TS-A-3T3 where sequences specific for the viral large T-ag are interrupted by recombination with cellular DNA. Analysis of the T-ags of TS-A-3T3 indicates that a 94K truncated T-ag species is made in addition to the viral 55K-middle and 22K-small T-ags. Truncated T-antigen species are frequently seen in Py transformed cells (2,3,7,23,24), and often arise when sequences coding for viral T-ags are interrupted by host DNA. The sequence beyond the join (J1) indicates that this is the case of TS-A-3T3 (Fig. 4). The join with host DNA at nucleotide 2644 results in the deletion of 85 amino acids from the carboxy-end of the large T-ag to which are added 19 additional amino acids coded for by the adjacent cell DNA.

A further consequence of the join (J1) with cellular DNA is that sequences containing the major polyadenylation site for the three early region mRNAs has been lost. Transcription studies of Py transformed cell lines have shown there are two potential classes of early region transcripts in cells where the major polyadenylation signal at 26 map units (mu) has been deleted. In the first class, nuclear transcripts extending thousands of nucleotides into adjacent host sequences have been detected (7,8), but cellular sequences are not represented in cytoplasmic mRNA, presumably due to the lack of 3' processing signals. Instead, cytoplasmic viral transcripts terminate at an alternative polyadenylation signal at 99 mu on the viral genome. This signal is less efficient than the signal located at 26 mu since only 5-10% of the early region transcripts are polyadenylated at this position during lytic infection, but it may become the predominate signal in the absence of a stronger distal signal. The second class of cytoplasmic transcripts are hybrid in nature arising when the viral transcripts utilise polyadenylation signals located in the adjacent cell DNA. The majority of the transcripts in TS-A-3T3 belong to this second category, terminating approximately 100 nucleotides beyond the viral-host junction. The DNA sequence through this region has features characteristic of other sequenced polyadenylation signals (18,19), strongly suggesting that the early region transcripts are using a host polyadenylation signal.

Transcripts of the second class are probably found in most cell lines producing truncated T-ag species which lack the 26 mu polyadenylation signal. This has been the case in all cell lines in which both transcription and translation products have been studied (8,9, Ruley et al., unpublished results). In no cell line studied has a truncated large T-antigen been detected corresponding to the translation product of a message terminating at 99 mu. Since many Py transformed cell lines which lack viral sequences containing the 26 mu polyadenylation signal produce truncated T-ag species, it appears that viral sequences frequently integrate in proximity to sequences capable of serving as polyadenylating signals. In the case of TS-A-3T3, it is not known whether the polyadenylation signal is actually part of a cellular gene. It is possible that sequences capable of serving as polyadenylation signals exist in the cellular genome in regions which are not transcribed.

Acknowledgements

We would like to thank Sharon Boast and Moira Griffiths for excellent technical assistance and Dr. M. Jones for providing sequence data of the TS-A virus.

REFERENCES

1. Tooze, J. ed. (1980) DNA Tumour Viruses. The Molecular Biology of Tumour Viruses, Part II: DNA Tumour Viruses. Cold Spring Harbor Laboratory, New York.
2. Lania, L., Gandini-Attardi, D., Griffiths, M., Cooke, B., De Cicco, D. and Fried, M. (1980). Virology 101, 217-232.
3. Lania, L., Hayday, A., Bjursell, G., Gandini-Attardi, D. and Fried, M. (1980). Cold Spring Harb. Symp. Quant. Biol. 44, 597-603.
4. Novak, U., Dilworth, S.M. and Griffin, B.E. (1980). Proc. Natl. Acad. Sci. USA 77, 3278-3282.
5. Hassell, J.A., Topp, W.C., Rifkin, D.B. and Moreau, P.E. (1980). Proc. Natl. Acad. Sci. USA 77, 3978-3982.
6. Israel, M.A., Vanderryn, D.F., Meltzer, M.L. and Martin, M. (1980). J. Biol. Chem. 255, 3798-3805.
7. Lania, L., Hayday, A. and Fried, M. (1981). J. Virol. 39, 422-431.
8. Kamen, R., Favaloro, J., Parker, J., Treisman, R., Lania, L., Fried, M. and Mellor, A. (1980). Cold Spring Harb. Symp. Quant. Biol. 44, 63-75.
9. Fenton, R.G. and Basilico, C. (1981). J. Virol. 40, 150-163.
10. Weaver, R.F. and Weissmann, C. (1979). Nucl. Acids Res. 7, 1175-1193.
11. Maxam, A.M. and Gilbert, W. (1980). Methods Enzymol. 65, 499-560.
12. Brutlag, D.L., Clayton, J., Friedland, P. and Kedes, L.H. (1982). Nucl. Acids Res. 10, 279-294.
13. Ito, Y. (1979). Virology 98, 261-266.
14. Griffin, B.E., Fried, M. and Cowie, A. (1974). Proc. Natl. Acad. Sci. USA 71, 2077-2081.
15. Soeda, E., Arrand, J.R., Smolar, N., Walsh, J.E. and Griffin, B.E. (1980). Nature 283, 445-453.
16. Korn, L.J., Queen, C.L. and Wegman, M.N. (1977). Proc. Natl. Acad. Sci. USA 74, 4401-4405.
17. Mount, S.M. (1982). Nucl. Acids Res. 10, 459-472.
18. Proudfoot, N.J. and Brownlee, G.G. (1974). Nature 252, 359-362.
19. Benoist, C., O'Harre, K., Breathnach, R. and Chambon, P. (1980). Nucl. Acids Res. 8, 127-142.
20. Calos, M.P. and Miller, J.H. (1980) Cell 20, 579-595.
21. Botchan, M., Stringer, J., Mitchison, T. and Sambrook, J. (1980). Cell 20, 143-152.
22. Stringer, J.R. (1981). J. Virol. 38, 671-679.
23. Dailey, L., Colantuoni, V., Fenton, R.G., La Bella, F., Zouzias, D., Gattoni, S. and Basilico, C. (1982). Virology 116, 207-220.
24. Ito, Y. and Spurr, N. (1980). Cold Spring Harb. Symp. Quant. Biol. 44, 149-157.