Volume 11 Number 20 1983 Nucleic Acids Research

Long terminal repeat sequences of intracisternal A particle genes in the Syrian hamster genome: identification of $tRNA^{Phe}$ as a putative primer tRNA

Masao Ono* and Hideyuki Ohishi

Department of Molecular Biology, School of Medicine, Kitasato University, Sagamihara-shi, Kanagawa 228, Japan

Received 2 August 1983; Revised and Accepted 26 September 1983

ABSTRACT

We have determined the nucleotide sequences of long terminal repeat (LTR) regions of Syrian hamster intracisternal A particle (IAP) genes . The size of the LTRs was 350 base-pairs (bp) and 376 bp in two clones, H10 and H18, respectively. Two LTRs at both ends of the IAP gene were linked to directly repeating 6 bp hamster sequences. Many structural features common to the integrated retroviral LTRs such as "CAT" box, "TATAA" box, polyadenylation signal, and terminal inverted repeat (3 bp), were present on each LTR. The estimated length of R region (about 60 bp) was similar to that of the murine leukemia-sarcoma virus. In contrast, the calculated U5 region of 54 bp was the shortest among those of the retroviruses so far studied. Furthermore, from the analysis of primer binding sites, phenylalanine tRNA was for the first time identified as a presumed primer tRNA for reverse transcription. These results clearly distinguish Syrian hamster IAP LTRs from other retroviral ones. Based on the comparison of the sequences between Syrian hamster and laboratory mouse LTRs, the structural features peculiar to the IAP LTRs and the origin of the IAP genes are discussed.

INTRODUCTION

Intracisternal A particles (IAPs) are retrovirus-like structures consistently observed in early mouse embryos and in a variety of mouse tumor cells (1-4). Sometimes IAPs are present in other mammals (5-10). Although isolated IAPs are not infective when injected in mice (11,12), these particles contain certain constituents generally associated with retroviruses such as a reverse transcriptase (13) and polyadenylated RNA molecules (IAP RNAs) of several discrete sizes ranging from 4 to 7 kilo base-pairs (kb) long (14-16).

DNA sequences homologous to IAP RNAs (IAP genes) are present in approximately 1,000 copies per haploid mouse genome (15,17,18) and these genes appear to be interspersed throughout the mouse chromosomes (17). For elucidation of their gene structures, IAP genes were first cloned from laboratory mice (Mus musculus)(15,19). A typical Mus musculus IAP gene was 7.2-7.3 kb and had long terminal repeat (LTR)-like sequences of 0.4 kb at

both ends of the gene (16,20,21).

Although the structural features of the IAP genes well resemble those of the integrated retrovirus genes, IAP genes have no apparent sequence homology with either type B or type C murine retrovirus genes (22). In addition, the number of IAP genes is at least 20 times more than those of any other endogenous retrovirus.

Sequences homologous to Mus musculus IAP genes were widely present in the DNAs prepared from many rodent animals and from mammals such as cat and monkey (23). Recently we isolated IAP genes from Syrian hamster (Mesocricelus auratus)(24) whose progenitor diverged from those of Mus musculus approximately 20 million years ago (25). Highly repetitive and mutually homologous Syrian hamster IAP genes were 7.6 kb in length and contained LTR-like sequences somewhat different from those of Mus musculus IAP genes, and limited sequence homology was observed between these two IAP gene families (24).

The function of IAP genes are still not evident. A possible involvement of IAP genes in the formation of a mouse pseudo α-globin gene hsa been discussed (26). Recently it has been reported that the translocation of an IAP gene in mouse myeloma or hybridoma cells causes either activation of a cellular oncogene, c mos (27,28), or inactivation of immunoglobulin κ-light chain gene (29,30). In these phenomena, LTR sequences must play significant roles in integration and possibly expression of the IAP gene and/or its flanking gene (31,32). Since two LTR sequences of Mus musculus IAP genes have been described in these articles (28,30), we have determined two LTR sequences of Syrian hamster IAP genes in order to elucidate the structure and interrelationship of the LTR sequences.

MATERIALS AND METHODS

Materials

Restriction endonucleases and terminal deoxynucleotidyl transferase were obtained from Takara Shuzo Co., and digestions were carried out according to the instructions of the supplier.

Recombinant DNA Clones

Two Syrian hamster IAP gene clones, H10 and H18, containing LTR-like sequences at both ends of the gene, have been described (24). The relative positions of the restriction enzyme sites and sequencing strategies of these cloned DNA are shown in Fig. 1.

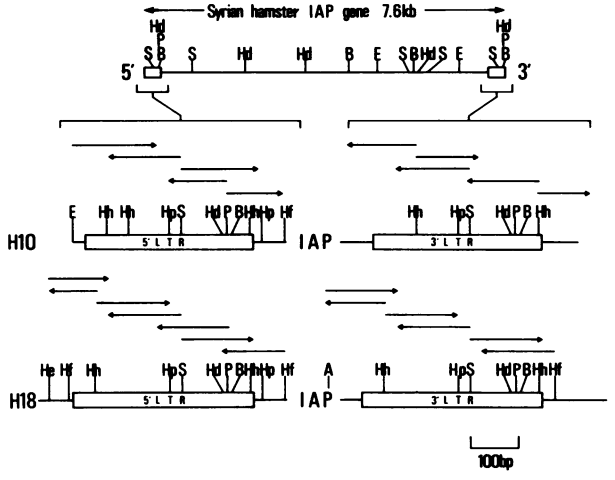

Fig. 1. Restriction maps and sequencing strategy for the terminal regions of the Syrian hamster IAP genes. Top diagram is a restriction map of a Syrian hamster IAP gene with terminal repeats (24). Expanded restriction maps of the terminal regions in the clones, H10 and H18, are shown below. The arrows indicate the extent and direction of the sequence determined. Restriction enzyme abbreviations: A, Alu I; B, Bam HI; E, Eco RI; Hd, Hind III; He, Hae III; Hf, Hinf I; Hh, Hha I; Hp, Hpa II; P, Pst I; S, Sst I.

DNA Sequencing Analysis

DNA fragments for sequencing analysis were labeled at 3'-ends with cordycepin 5'-[α-^{32}P] triphosphate and terminal deoxynucleotidyl transferase (33). The end-labeled fragments were analyzed according to the method of Maxam and Gilbert (34).

RESULTS

Identification of LTR Sequences

As sequences homologous to the Mus musculs IAP gene, Syrian hamster IAP genes were cloned from the partial Eco RI-digested gene library constructed from Syrian hamster liver DNA (24). Since two clones, H10 and H18, possessed directly repeating sequences of several hundred base-pairs at both ends of the 7.6 kb IAP genes, we determined the nucleotide sequences of these repeating regions.

The identical length of the H10 repeats were 350 bp but one base substitution was observed at position 47 (Fig. 2). The H18 repeats, however, each had identical sequences of 376 bp (Fig. 2) Since these sequences possessed many features commonly associated with retroviral LTR structures

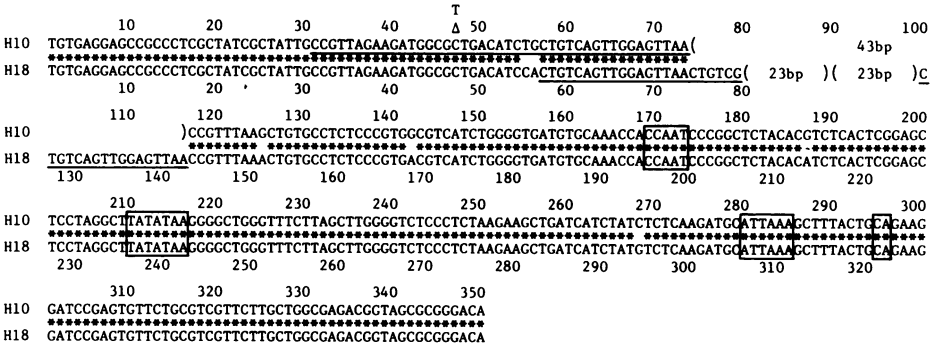

Fig. 2. Nucleotide sequences of LTRs positioned at the 5'-ends of the Syrian hamster IAP genes, H10 and H18. LTR sequences of the strand having the same polarity as IAP RNA (24) are presented. The sequences of H18 LTRs were identical at both ends, whereas H10 LTR situated at 3'-end had a substituted T at position 47 (Δ) instead of C. Asterisks indicate nucleotide identity. "CAT" box, "TATAA" box, polyadenylation signal, and polyadenylation site are boxed. Direct repeats (43 bp in H10 and 23 bp in H18) are underlined first and enclosed by parentheses next. The 17 bp sequence positioned at the end of three repeats of 23 bp in H18, is identical but lacks the last 6 bp of the 23 bp sequence.

(31,32), as mentioned below, we defined these as LTR sequences of the Syrian hamster IAP genes.

On the H10 LTR, two identical sequences of 43 bp were observed side by side at position 31-116. Starting at position 57, H18 LTR had three consecutive direct repeats of 23 bp followed by a 17 bp sequence which was identical to but lacked the last 6 bp of the 23 bp sequence. The presence of the repeating sequences is a common feature of the retroviral LTRs even though their function is unknown (31,32). Although the length and number of these repeats were different, H10 and H18 shared almost identical sequences except for the absence of a 6 bp sequence starting at position 74 of the H18 LTR. By comparison of the corresponding sequences (307 bp) between H10 and H18 LTRs, base substitution at 6 positions dispersed on the LTRs was observed, so that the sequence homology between these LTRs could be calculated as 98%.

Structural Features of LTR Sequences and Their Flanking Regions

Although the lengths and the sequences are fairly different among retroviral LTRs, they share many structural features (31,32). Each LTR has inverted repeats of 2-16 bp at the ends and is bounded by 5'TG---CA3'. Transcription of the retrovirus genes always starts with G which is positioned at about 30 bp downstream from the promotor sequence of RNA

polymerase II called the "TATAA" box (35). At 40-50 bp upstream from the "TATAA" box, a canonical promotor sequence called the "CAT" box is located (35). The polyadenylation signal is presumed to be $A^{A}_{T}TAAA$ (35) and a poly(A) tail is added next to CA which is placed 10-20 bp downstream from the polyadenylation signal (31).

Two TMP residues are usually found immediately after the 5'LTR. The next 18 nucleotides form a primer-binding site complementary to the last 18 bases of a certain cellular tRNA called the primer tRNA. Right before the 3'LTR, a purine rich sequence of 10-20 bp is observed followed usually by two AMP residues. On integration of the retrovirus gene, cellular DNA of 4-6 bp adjacent to the 5'LTR is directly repeated next to the 3'LTR. Depending on the retrovirus species, the number of nucleotides of these repeats is constant although the sequences are always different. We therefore examined whether the LTRs of the Syrian hamster IAP genes had these common structural features.

The hamster IAP LTR had inverted repeats of 3 bp the sequence of which was 5'TGT---ACA3'. Starting at position 210 in H10 and 236 in H18, both LTRs had the sequence, TATATAA, which was identical to the consensus sequence of the "TATAA" box ($TATA^{A}_{T}A^{A}_{T}$)(35). The sequences, CCAAT ("CAT" box), were observed in both LTRs placed at intervals 36 bp upstream from the "TATAA" box. The polyadenylation signals, ATTAAA, were found at intervals of 63 bp downstream from the "TATAA" box. The polyadenylation sites, CA, were located at intervals of 9 bp downstream from the polyadenylation site.

The retroviral LTR structure can be subdivided into 3 regions, 5'-end U3-R-U5 3'-end (31,32). The R region always starts with G positioned at about 30 bp downstream from the "TATAA" box, and ends with the polyadenylation site, CA. If we suppose the G nucleotide to be located between the 26th and 34th nucleotide downstream from the initial T of the "TATAA" box as the 5'-end of the R region (35), since the nucleotide sequence of both 5'- and 3'-ends of the IAP RNA has not been determined yet, then the length of the R region can be calculated as 57-60 bp. This value is close to those of murine leukemia-sarcoma viruses but fairly larger than those of avian leukosis-sarcoma viruses (31,32). In contrast, the length of the U5 region was estimated to be 54 bp which is the shortest value ever reported.

Three kinds of tRNA, Trp(36), Pro(37) and Lys(38), have been identified as primer tRNAs for reverse transcription. The nucleotide sequences of the primer-binding site of the two hamster IAP genes did not coincide with those

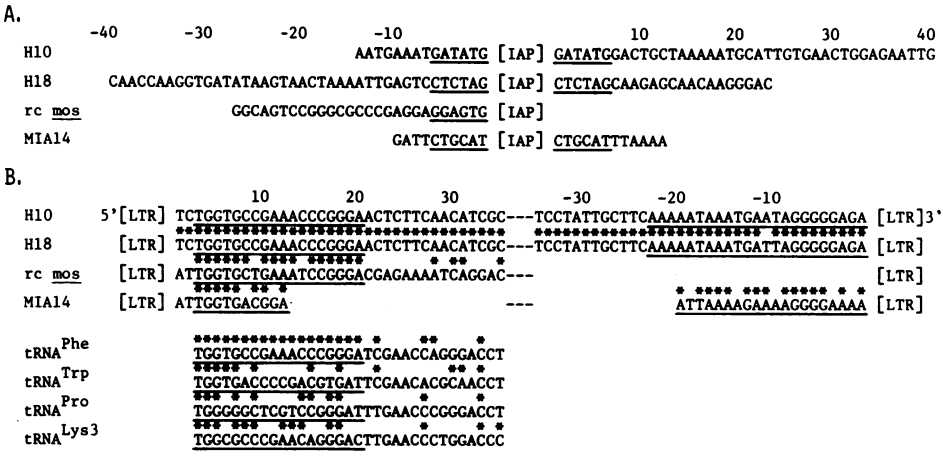

Fig. 3. Nucleotide sequences of the flanking regions of the IAP LTRs. A. Sequences of cellular DNA adjacent to the IAP LTRs. The direct repeats of 6 bp are underlined. B. Sequences of the primer-binding site and the purine rich region. The primer-binding site and the purine rich region are underlined. Sequences complementary to the last 33 nucleotides of the known primer tRNAs (36-38) and the mammalian $tRNA^{Phe}$ (39,40) are shown below. Asterisks indicate nucleotide identity with the H10 sequence.

reported previously, but were perfectly complementary to the last 18 nucleotides of mammalian phenylalanine tRNA (39,40)(Fig. 3B). Adjacent upstream from the 3'LTR were observed the purine rich sequences 23 bp in length (Fig. 3B). Cellular direct repeats of 6 bp were always present immediately outside both LTRs, but these sequences were not mutually homologous at all (Fig. 3A).

From these results, we may conclude that the LTR sequences of the Syrian hamster IAP genes possess all the structural features common to the retroviral LTRs but should be grouped into a distinct class.

Comparison of the Sequences between Syrian hamster and laboratory mouse LTRs

Recently, two LTR sequences of the Mus musculus IAP genes have been reported (27,28,30). To elucidate the interrelation between the Syrian hamster and Mus musculus IAP genes, we compared the LTR sequences and their flanking regions cloned from two phylogenetically distant rodent animals (25). The lengths of the Mus musculus LTRs (338 bp in MIA14 and 351 bp in rc mos) were somewhat different from those of the hamster LTRs but possessed all the characteristics generally associated with retroviral LTRs (Fig. 3,4 and TABLE I).

TABLE I. Size of IAP LTR Regions

Clone	LTR	5'End	+++	CAT Promotor	+++	TATAA Promotor	+++	Poly(A) Signal	+++	CA	+++	3'End
General Consensus				CCAAT		TATA$^{A}_{T}A^{A}_{T}$		A$^{A}_{T}$TAAA		CA		
H10	350	168		CCAAT	36	TATATAA	63	ATTAAA	9	CA		54
H18	376	194		CCAAT	36	TATATAA	63	ATTAAA	9	CA		54
rc mos	351	152		CCAAT	37	TTAAGAG	76	AATAAA	9	CA		57
MIA14	338	155		CCAAT	37	TAAAAAA	60	AATAAA	9	CA		56

The regulatory signals and the number of nucleotide between the signals are indicated.

Although the overall sequence homology of the corresponding regions between H10 and rc mos LTR was estimated to be 60%, highly homologous regions were dispersed on the LTRs (Fig. 4). Two highly homologous regions, A and C, were located between the 5'-end and "CAT" box, and their homologies between H10 and rc mos were 83% and 79%, respectively. Another region, F,

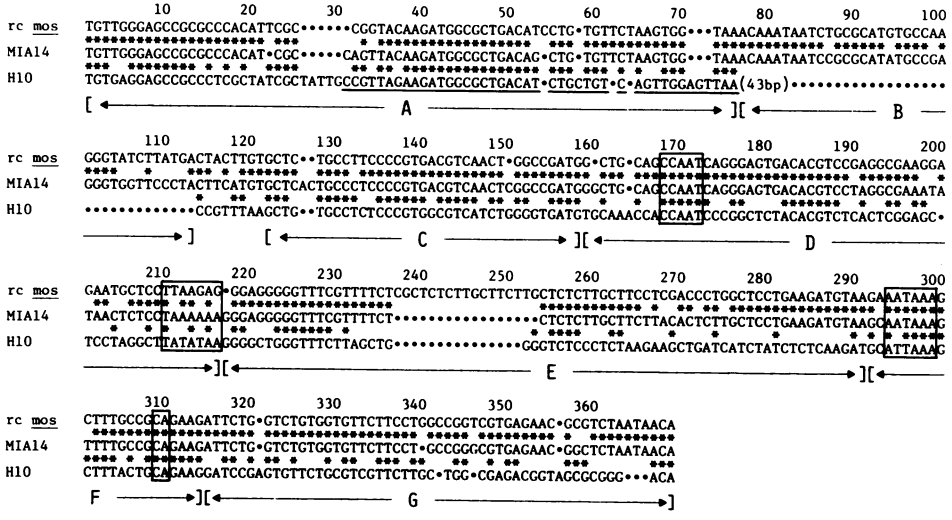

Fig. 4. Comparison of the sequences between Syrian hamster and laboratory mouse LTRs. The nucleotide sequence of the Syrian hamster LTR, H10, is compared with the mouse LTRs, rc mos and MIA14 (27,28,30). LTR sequences positioned at the 5'-end of each IAP gene are presented. Asterisks indicate nucleotide identity with rc mos. The dot indicates the absence of a nucleotide. The regulatory signals for transcription are boxed. The directly repeating sequences in H10 LTR are indicated as in Fig. 2.

having a sequence homology of 83%, was positioned almost between the polyadenylation signal and polyadenylation site. Since the primer-binding site of the rc mos IAP gene was 89% homologous to that of H10, in spite of the limited sequence homology downstream from these regions (Fig. 3B), these highly conserved regions must be indispensable for the function of the LTR structure of the IAP genes.

Both the distance between the "CAT" box and "TATAA" box, and the length of the putative U5 region, were fairly constant in two kinds of LTRs (TABLE I); however, the extent of the sequence homology in these regions was limited (49% in region D and 52% in region G). In these regions, the constant distance between the functional signals must be significant for the functional LTR. It was found, for instance, that the "TATAA" box is positioned at roughly 3 turns of the DNA helix from the mRNA start site (35). On the IAP LTRs, the distance between the "CAT" box and "TATAA" box was about 4 turns of the DNA helix, while the polyadenylation site was located at about one turn downstream from the polyadenylation signal.

More heterogeneity was observed in the region E (36%), and the sequence of region B was absent in the Syrian hamster LTR.

DISCUSSION

LTR sequences are the direct repeats of 300-1400 bp situated at both ends of the provirus genome of the retrovirus and play significant roles in the reverse transcription of viral RNA, integration of the provirus DNA into cellular DNA, and expression of viral genes (31,32). Nucleotide sequencing analysis demonstrated that the LTR sequences and their flanking regions of the Syrian hamster IAP genes not only possessed all the structural features commonly observed on the retroviral LTRs and their flanking sequences, but had some features peculiar to the IAP genes. The first feature had the shortest length of the putative U5 region. The second feature identified the phenylalanine tRNA as the presumed primer tRNA. In addition, two nucleotides bound immediately either after the 5'LTR (TC in the hamster and AT in the mouse), or before the 3'LTR (GA in the hamster), were different from the general consensus nucleotides, 5'[LTR]TT---AA[LTR]3' (31,32), although MIA14 had the common nucleotides, AA, right before the 3'LTR (Fig. 3B).

A rough survey by visual inspection suggested that the IAP LTR has no marked sequence homology with those of either murine leukemia-sarcoma virus, mouse mammary tumor virus, or avian leukosis-sarcoma virus. A more detailed

comparison facilitated by the use of a computor will be necessary for elucidating the function and origin of the IAP LTR, since the sequence homology between avian leukosis-sarcoma virus and copia-like transposable element of Drosophila can be found by detailed analysis (41).

As the IAP genes, which are dispersed throughout the cellular genes, behave as cellular genes, they have taken on mutations in proportion to the alternation of generations. In the case of rodents, about 8×10^5 years are required for the accumulation of the mutations to cause a 1% difference in the nucleotide sequences of the chromosomal unique genes (25). The same mutation rate was reported for the nonfunctional mouse α-globin pseudo gene (42). If it is supposed that the mutation rate of the IAP LTR is equal to that of the unique sequence genes of the rodent, then we can estimate when the integration and divergence of the particular IAP gene took place.

Since the LTR sequences positioned at both ends of the provirus gene have been reported to be mutually identical immediately after its integration (31,32), the difference in the nucleotide sequence between these two LTRs must have been caused by random mutation following integration. These differences in H10 and H18 were 0.3% and 0%, respectively, so that we can calculate that less than 1.5×10^5 years passed after the integration of these genes in the hamster genome. A two percent difference observed in the MIA14 LTRs (30) indicates the integration time as 8×10^5 years ago. The sequence homology between rc mos and MIA14 LTRs was estimated as 88% (28) and thus the ancestors of these two genes must have diverged from each other about 4.8×10^6 years ago.

The estimated age of divergence between the ancestors of H10 and rc mos LTRs (about 1.6×10^7 years ago) was similar to that between the Syrian hamster and the laboratory mouse (2.0×10^7 years)(25). Since the sequences homologous to the Mus musculus IAP genes are found in chromosomal DNAs in almost all of rodent (23), the ancestors of the IAP genes must have been present on the chromosomes of the common progenitors of the rodent prior to their divergence, and accumulated mutations following species differentiation.

It is still not clear as to when the integration of the huge number of IAP genes of the Syrian hamster or laboratory mouse took place. Since both cloned and genomic hamster IAP genes were considerably homogeneous (24), the integration of these genes must have occurred at about the same time as that of H10 and H18 sequences. In the same manner, variations and deletions

observed in the mouse IAP genes (15,16) suggest an integration time similar to that of the MIA14 sequence.

ACKNOWLEDGEMENT

We thank Dr. M. Kawakami for encouragement during this work. This work was supported in part by a research grant from the Japanese Ministry of Education, Science and Culture.

*To whom correspondence should be addressed

REFERENCES

1. Weivel,N.A. and Smith,G.H. (1971) Int. J. Cancer 7, 167-175.
2. Calarco,P.G. and Szollosi,D. (1973) Nature New Biol. 243, 91-93.
3. Biczysko,W., Pienkowski,M., Solter,D. and Koprowski,H. (1973) J. Natl. Cancer Inst. 51, 1041-1051.
4. Chase,D.G. and Piko,L. (1973) J. Natl. Cancer Inst. 51, 1971-1973.
5. Shipman,Jr.C., Weide,G.C.V. and IlMa,B. (1969) Virology 38, 707-710.
6. Cesarini,J.P. and deMicco,C. (1972) Int. J. Cancer 10, 174-185.
7. Weber,H.W., Geddes,A. and Stellwagen,R.H. (1978) J. Natl. Cancer Inst. 60, 919-923.
8. Opler,S.R. (1967) J. Natl. Cancer Inst. 38, 797-800.
9. Feldman,D.G. and Gross,L. (1970) Cancer Res. 30, 2702-2711.
10. Tumilowicz,J.J. and Cholon,J.J. (1971) Proc. Soc. Exp. Biol. Med. 136, 1107-1110.
11. Hall,W.T., Hartley,J.W. and Sanford,K.K. (1968) J. Virol. 2, 238-247.
12. Kuff,E.L., Wivel,N.A. and Lueders,K.K. (1968) Cancer Res. 28, 2137-2148.
13. Wilson,S.H. and Kuff,E.L. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1531-1536.
14. Paterson,B.M., Segal,S., Lueders,K.K. and Kuff,E.L. (1978) J. Virol. 27, 118-126.
15. Ono,M., Cole,M.D., White,A.T. and Huang,R.C.C. (1980) Cell 21, 465-473.
16. Kuff,E.L., Smith,L.A. and Lueders,K.K. (1981) Mol. Cell. Biol. 1, 216-227.
17. Lueders,K.K. and Kuff,E.L. (1977) Cell 12, 963-972.
18. Shen-Ong,G.L.C. and Cole,M.D. (1982) J. Virol. 42, 411-421.
19. Lueders,K.K. and Kuff,E.L. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3571-3575.
20. Cole,M.D., Ono,M. and Huang,R.C.C. (1981) J. Virol. 38, 680-687.
21. Cole,M.D., Ono,M. and Huang,R.C.C. (1982) J. Virol. 42, 123-130.
22. Lueders,K.K. and Kuff,E.L. (1979) J. Virol. 30, 225-231.
23. Lueders,K.K. and Kuff,E.L. (1981) Nucleic Acid Res. 9, 5917-5930.
24. Suzuki,A., Kitasato,H., Kawakami,M. and Ono,M. (1982) Nucleic Acid Res. 10, 5733-5746.
25. Benveniste,R.E., Callahan,R., Sherr,C.J., Chapman,V. and Todaro,G.J. (1977) J. Virol. 21, 849-862.
26. Lueders,K.K., Leder,A., Leder,P. and Kuff,E.L. (1982) Nature 295, 426-428.
27. Rechavi,G., Givol,D. and Canaani,E. (1982) Nature 300, 607-611.
28. Kuff,E.L. Feenstra,A., Lueders,K., Rechavi,G., Givol,D. and Canaani,E. (1983) Nature 302, 547-548.

29. Hawley,R.G., Shulman,M.J., Murialdo,H., Gibson,D.M. and Hozumi,N. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 7425-7429.
30. Kuff,E.L., Feenstra,A., Lueders,K., Smith,L., Hawley,R., Hozumi,N., and Shulman,M. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 1992-1996.
31. Temin,H.M. (1981) Cell 27, 1-3.
32. Varmus,H.E. (1982) Science 216, 812-820.
33. Tu,C-P.D. and Cohen,S.N. (1980) Gene 10, 177-183.
34. Maxam,A.M. and Gilbert,W. (1980) Methods Enzymol. 65, 499-560.
35. Breathnach,R. and Chambon,P. (1981) Ann. Rev. Biochem. 50, 349-383.
36. Harada,F., Sawyer,R.C. and Dahlberg,J.E. (1975) J. Biol. Chem. 250, 3487-3497.
37. Harada,F., Peters,G.G. and Dahlberg,J.E. (1979) J. Biol. Chem. 254, 10979-10985.
38. Peters,G. and Glover,C. (1980) J. Virol. 35, 31-40.
39. Roe,B.A., Anandaraj,M.P.J.S., Chia,L.S.Y., Randerath,E., Gupta,R.C. and Randerath,K. (1975) Biochem. Biophys. Res. Comm. 66, 1097-1105.
40. Keith,G. and Dirheimer,G. (1978) Biochim. Biophys. Acta. 517, 133-149.
41. Kugimiya,W., Ikenaga,H. and Saigo,K. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3193-3197.
42. Miyata,T. and Yasunaga,T. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 450-453.