Nucleotide sequences of murine intracisternal A-particle gene LTRs have extensive variability within the R region

Robert J.Christy, Anne R.Brown¹, Brian B.Gourlie and Ru Chih C.Huang²

Department of Biology, Johns Hopkins University, Baltimore, MD 21218, USA

Received 4 September 1984; Revised and Accepted 27 November 1984

ABSTRACT

Nucleotide sequences of the long terminal repeats (LTRs) of four murine intracisternal A-particle (IAP) genes IAP62, 19, 81 and 14 were determined. Each IAP LTR contains three sequence domains, 5'-U3-R-U5-3', and each is bound by 4 bp imperfect inverted repeats. The transcriptional regulatory sequences, CAAT and TATA, as well as the enhancer core sequence GTGGTAA are conserved and precisely positioned within the U3 region. In the R region, the sequence AATAAA is located twenty base pairs preceding the dinucleotide CA, the polyadenylation site. In IAP19 and IAP81, the 5' and 3' LTRs are flanked by a six nucleotide direct repeat of cellular sequences representing the possible integration sites for these IAP proviruses. Both the size and sequences of different IAP LTRs vary considerably, with the majority of the variation localized within the R regions. The size of R varies from 66 bp in IAP14 to 222 bp in IAP62; in contrast, the U3 and U5 regions are all similar in size. These extra sequences within the R region of large LTRs consist of several unusual directly repeating sequences which account for this variability.

INTRODUCTION

Intracisternal A particles (IAPs) are endogenous retrovirus-like structures that are found budding from the endoplasmic reticulum in normal mouse preimplantation embryos (1-3) and in many mouse tumors (4,5), but rarely in normal mouse cells (6). In Mus musculus there are approximately 1000 integrated copies of IAP genes per haploid genome, which represents 0.2% of the total DNA in mouse (7,8). By restriction endonuclease mapping, heteroduplex formation, and genomic DNA blot hybridization using cloned IAP genes, we (8,9) and others (10-12) have grouped the mouse IAP genes into two classes: type I IAP genes, approximately 7 kb in length, and type II genes, approximately 4 kb in length. Type I genes outnumber type II genes by a ratio of about 10:1 in Mus musculus (10). All IAP genes have conserved 3' coding sequences while the 5' ends vary considerably both in length and in sequence between the IAP genes of the two classes and even among IAP genes within the same class (8,10,11). Three species of IAP transcripts (7.2 kb, 5.3 kb, and 3.8 kb) are found in plasmacytoma cells, MOPC 315 and TEPC 15 (9,19). On the

other hand, IAP transcripts from neuroblastomas are 7.2 and 5.3 kb in length (13), while the major species of IAP transcripts in embryonic teratocarcinomas is 5.3 kb (14). This 5.3 kb species is also the major IAP transcript expressed in preimplantation embryos (15).

Both type I and type II IAP genes are flanked on the 5' and 3' ends by long terminal repeats (LTRs)(11,16). Kuff et al. have determined the nucleotide sequences of the LTRs from a type I IAP gene, MIA14 (17). They have found that the LTRs of MIA14, like those of other retroviruses, contain the presumptive regulatory signals for promotion, initiation, and polyadenylation of transcription. In our laboratory, using S1 nuclease and cDNA extension mapping (18), and more recently with in vitro transcription studies (19), we have shown that IAP RNA initiates within the 5' LTR and terminates in the 3' LTR of genes. Thus, long terminal repeats seem to play an important role in expression and termination of IAP gene transcription. With over a thousand copies of IAP genes in Mus musculus, it is exceedingly difficult to determine which IAP genes are active in transcription. We do not know if all IAP genes are capable of transcription, or whether all the IAP transcripts derive from a small subset of active IAP genes. If so, do active IAP genes have different LTRs from those of inactive genes? To begin to address these questions, we have sequenced the LTRs of four different IAP genes. The long terminal repeats from the four genes which were previously isolated in our laboratory (8), one type I gene (λ IAP81) and three type II genes (λ IAP62, λ IAP19, λ IAP14), were subcloned and sequenced. We found a large variation in sequence and in size among long terminal repeats from different IAP proviruses. Unlike the endogenous avian leukosis virus of chickens where variation in LTR length is due to deletions within the U3 region (20), variability in IAP LTR lengths is likely due to nucleotide deletion or insertion within the R region. Several directly repeating sequences are present in the R regions of long LTRs that are missing in the R regions of short LTRs, and these may play a role in the R region variability.

MATERIALS AND METHODS

Restriction endonucleases were obtained from New England Biolabs or Bethesda Research Laboratories. T4 ligase, Klenow fragment of E. coli DNA polymerase I and M13 15-bp sequencing primer were from New England Biolabs. Endonuclease digestions and DNA modifications were carried out according to the specifications recommended by the manufacturer.

Cloning of IAP LTRs

Isolation and mapping of IAP genomic clones has been described previously (8). IAP DNA inserts from Charon 4A recombinant clones λ IAP62, λ IAP19, λ IAP81 and λ IAP14 were subcloned in plasmid pBR322 (8) and fragments that contained the LTR regions were isolated. Fragments to be sequenced by the dideoxynucleotide chain-termination method were subcloned into the bacteriophage M13 vectors mp8, mp9 or mp10 (21). DNA to be sequenced by the chemical cleavage method was used directly after isolation of the fragments by electroelution (22), and end labelled using T4 polynucleotide kinase (BRL) according to the method of Maxam and Gilbert (23).

Nucleotide Sequencing

The two LTRs and their flanking sequence of IAP19 and the 3' LTR of IAP62 were sequenced using both the chemical cleavage method of Maxam and Gilbert (24) and the M13 dideoxynucleotide chain-termination method of Sanger *et al.* (25). The 3' LTR of IAP14 and the 5' LTR of IAP81 were sequenced using the dideoxynucleotide chain-termination method only.

DNA Sequence Alignment and Analysis

The LTRs were aligned using the NUCALN sequence homology computer program (26) with parameters set at: k-tuple= 2, window size= 20, and gap penalty= 5. Direct and inverted repeats were found using the Los Alamos SEQH program (27) and comparing each LTR to itself.

RESULTS

Primary Nucleotide Sequence of Five IAP LTRs:

In a previous report, we described the isolation and cloning of several mouse IAP genes (8). Some of these cloned sequences were further analyzed by restriction enzyme mapping, DNA filter hybridization and heteroduplex analysis (8,9,18). We found that two IAP DNA clones, IAP19 and IAP81, contain a few hundred nucleotides of long terminal repeat (LTR) sequences at both the 5' and 3' ends of these genes (16). These analyses demonstrated a size and restriction site heterogeneity between these two different IAP LTRs. A comparison of the restriction maps of the IAP LTR clones is shown in Figure 1. Several restriction enzyme sites PstI (pos.155), HinfI (pos.458), and MspI (pos.485) are conserved in all LTRs studied, while an SstI site (pos.308) is unique to long LTRs.

To further characterize IAP LTRs, specifically, to ascertain how these IAP LTRs may differ from each other and from the LTR of MIA14 reported by Kuff *et al.* (17), we have subcloned and sequenced the LTRs from IAP62, 19, 81 and

Figure 1. IAP LTR clones. Restriction maps and sequencing strategies for the long terminal repeats from IAP62, 19, 81 and 14 as described in Table 1. The terminal repeats are aligned in a 5' → 3' (U3-R-U5) orientation by the conserved PstI site. Arrows indicate direction and extent of sequence determined. Dashed line (↔) indicates sequencing by the method of Maxam and Gilbert (24), solid line (↔) sequencing by dideoxy-chain termination by the method of Sanger *et al.* (25). Restriction enzyme sites are denoted: PstI, P; EcoRI, E; SstI, S; BglIII, B; HindIII, Hd; XbaI, X; HinfI, H; and MspI, M. The LTR is an open box (□), IAP coding sequences, shaded line (■) and cellular DNA, solid line (—).

14 (8). Nucleotide sequences of the IAP LTR, MIA14 and rc-mos (the LTR from an IAP gene inserted into the c-mos gene in a mouse myeloma XRPC24 (28-30)), are shown in Figure 2. The nucleotide sequences are aligned to IAP62, the longest LTR, and are numbered in a 5' to 3' orientation. We have compared the IAP LTR sequences with those of other retroviral LTRs to determine whether the sequences thought to be important for integration are also present in IAP LTRs. Most retroviral LTRs and transposable elements are terminated by perfect or imperfect complementary inverted repeats two to sixteen base pairs in length and duplicate cellular sequences at their cellular integration site (31-33). We found that our IAP LTRs have a 4 bp imperfect inverted repeat, 5'TGTT/AAGA3', flanking several hundred nucleotides of terminally repeated sequences. The finding that our IAP LTRs are bound by imperfect inverted

Figure 2. Comparison of nucleotide sequences from IAP long terminal repeats. The nucleotide sequences of LTRs from IAP19, 81, 14, MIA14 (17) and rc-mos (29) are compared to the largest LTR, IAP62. The putative regulatory signals for transcription are labelled and boxed and the conserved PstI site found in all IAP LTRs is at position 155. Comparisons with IAP62 are denoted as follows: (□, □) deletion, (-) same sequence, (N) nucleotide substitution or insertion. (X) nucleotide not determined. Enhancer core sequence GTGGTAA (37) is present at position 61 to 68. The Z DNA structure is present at position 75 to 83. See Table 1 for descriptions of LTRs.

Table 1 Sizes of Long Terminal Repeats of IAP Genes

IAP Genes ¹	Location of the LTR	Type of IAP Genes ²	Length of LTR
IAP62	3' terminus	II	489
IAP19D	5' terminus	II	468
IAP19B	3' terminus	II	467
IAP81C	5' terminus	I	348
IAP14	3' terminus	II	329
MIA14	5' terminus	I	337
rc-mos ³	3' terminus	II	345

¹LTRs isolated from IAP gene clones designated as in Ref.(8).

²As distinguished by proviral gene size (9,10).

³Size of IAP gene 3' LTR that is found integrated at the 5' end of the c-mos gene in myeloma XRPC24 (28,29).

(34), downstream from the 5' LTR; and a purine-rich region upstream from the 3' LTR in IAP19 (Fig. 3).

Long terminal repeats in other viral systems have been found to contain enhancer sequences, a cis acting element that may be involved in activation of RNA transcription of nearby genes. At position 61 in all the IAP LTRs, we found a sequence similar to that of the SV40 enhancer "core" sequence GTGG^{TTT}_{AAA} (37). The sequences in this region are well conserved, and the enhancer sequence appears to be part of all IAP LTRs. There are also some potential Z-DNA-forming sequences 3' to the enhancer core (position 78-88, Fig. 2) in all IAP LTRs that have been sequenced. Z-DNA forming sequences have also been mapped at position 11-21 in both MIA14 and rc-mos (38). In the LTRs sequenced in our laboratory, this Z-DNA sequence was not found due to a C to G transversion at position 14, thereby decreasing the possibility for left-handed DNA formation. However, the presence of only one Z-DNA forming sequence in LTRs of our IAP clones is unusual; in other viral systems and in IAP LTRs of MIA14 and rc-mos, two pairs of Z-DNA sequences are found approximately 50 to 80 nucleotides apart (38).

Mapping the U3-R-U5 Regions Within the IAP LTRs

Several conserved sequences which are essential for transcriptional regulation have been used to subdivide LTRs into three functional domains,

U3-R-U5 (32). The sequence CCAAT (CAT box) usually occurs in the U3 region 75 bp 5' to R; the sequence $\overline{\text{A}}\text{AT}\overline{\text{A}}\overline{\text{A}}\text{G}$ (Goldberg-Hogness box, TATA) usually occurs 26-32 bp before R. These sequences are thought to be important for transcription promotion (For review, 39). The R region always starts at the capping nucleotide, G, and ends with the poly(A) addition site, CA. Twenty base pairs 5' to the start of the U5 region there is the polyadenylation signal sequence, $\overline{\text{A}}\overline{\text{A}}\overline{\text{A}}\overline{\text{A}}\overline{\text{A}}$. Using our sequence homologies and data from other IAP LTRs, we have searched for these putative regulatory sequences in order to map the boundaries between U3, R and U5 domains within the IAP LTRs; Table 2 summarizes these findings. It appears that, similar to LTRs of other retroviruses, the IAP LTRs contain these putative nucleotide sequences for promotion, initiation, polyadenylation, and termination of viral RNA transcription. These consensus sequences are arranged like those of other retroviral LTRs (33) and genes transcribed by RNA polymerase II (39). At position 190 (Fig. 2 and Table 2) is the so-called "TATA" box (40), and another presumed promotor element, the CAT box (41), is found 30 bp upstream from the TATA-like sequences. The presence and position of the TATA box is consistent with previous data from our laboratory that IAP RNA transcription starts at a position close to the PstI site at position 155 (18). The transcriptional start site of IAP gene MIA14 has recently been localized using S1 nuclease mapping (42). Using these data, the putative viral RNA cap site (the 5' boundary of the R region), normally located 26-34 nucleotides downstream of the TATA box, has been positioned at nucleotide 221, which is 30-32 nucleotides downstream from TATA in our IAP genes. The canonical polyadenylation signal (AATAAA)(43) and acceptor site (CA)(44) are found at the 3' end of the R region with their spacing and sequence well conserved in all the LTRs. Sequences of an IAP cDNA clone have shown that poly-adenylation does occur at this putative acceptor site (CA) (K. Moore, personal comm.). Therefore, we have mapped the 3' boundary of the R region to position 453. Also found, approximately 20 nucleotides downstream from the polyadenylation acceptor site (CA), is a sequence thought to be important in viral RNA termination TGTT (32) (pos. 470).

Our data also show that the U3 and U5 regions are approximately the same size in all IAP LTRs. The size of the U5 region, 55 ± 2 bp, is smaller than those reported in other retroviral LTRs, but is consistent in size with IAP LTRs isolated from Syrian hamster (34) and *Mus caroli* (35). Previous reports indicate that differences in LTRs among virus species and within the same species are due to variations in U3. This is not the case in IAP LTRs where

Table 2 Sequence of Putative Transcriptional Signals and Size of LTR Functional Domains

LTR CLONE ¹	U3		R	U5	
	→ "CAT" Box TGTT CCAAT	← "TATA" Box Length (bp) ² AATATAA	← Poly(A) Site G Length (bp) ³ AATAAA CA	← Length (bp) ⁴ AACA	←
IAP62	TGTT	XCAAT (212)	AGGATGA	G (222) AATAAA CA	(55) AAGA
IAP19B	TGTT	XCAAT (213)	AATATAA	G (201) AATAAA CA	(54) AAGA
IAP19D	TGTT	XCAAT (212)	AATATAA	G (202) AATAAA CA	(53) AAGA
IAP81C	TGTT	XCAAT (209)	AGGATAA	G (84) AATAAA CA	(55) AAGA
IAP14	TGTT	XCAAT (210)	AGGATAA	G (66) AATAAA CA	(53) AAGA
MIA14	TGTT	CCAAT (214)	AATATAA	G (66) AATAAA CA	(57) AACA
rc-mos	TGTT	CCAAT (207)	AGGAGAA	G (81) AATAAA CA	(57) AACA

¹See Table 1.²Length between 5' inverted repeat and capping nucleotide, G.³Length of the R region.⁴Length between polyadenylation site, (CA), and 3' inverted repeat.

the U3 region is the same size in all LTRs, 220 ± 4 , and similar in size to the U3 region of avian leukosis-sarcoma virus (20,32). In IAP LTRs, the difference in length is due to variation in the R region, 66 bp in IAP14 and MIA14 to 222 bp in IAP62. Mapping analysis by cDNA extension and S1 nuclease have indicated that IAP transcripts initiate within the 5' LTR and terminate within 3' LTR (18). The size of the 5' cDNA extension of 780 nucleotides (Ref. 18, Fig. 6 and Fig. 7) could only result from an IAP gene with a long R region, such as IAP19 or IAP62, thus indicating that IAPs with long R regions are being transcribed in MOPC315 myeloma cells.

Localization of Short Oligonucleotide Repeats Within and at the Boundary of the R Region.

We have shown (Table 1) that LTRs of IAP62 and IAP19 are approximately one hundred forty nucleotides longer than those from IAP81, IAP14, and MIA14 (17), and rc-mos (29). These size variations are largely due to changes in the R regions of the LTRs (Fig. 2 and Table 2); the long LTRs contain several directly repeated sequences not present in the short LTRs. A diagram showing the location of these short direct repeats is presented in Fig. 4. All LTRs studied contain at least one copy of a 28 nucleotide repeat (CTGGCCCTGAAGC-TGTAAGCAATAAAG near the 3' end of the R region (positions 417-444) that includes the polyadenylation signal AATAAA (Fig. 2) and makes the 3' end adenine rich. This 28 nucleotide stretch is completely repeated at positions 284 to 311 and the first 19 nucleotides of the repeat are found at positions 355 to 374 in the R regions of both IAP62 and IAP19; these two repeats are

Figure 4. Repeat sequences in the R regions of IAP LTRs. The R region (IAP62, 19, 81, 14, MIA14 and rc-mos) is represented by an open box (□) and sequences deleted in LTR (▤). The numbering from 220–460 bp corresponds to nucleotide sequence of LTRs as in Fig. 2. Sequences indicated by (•••••) and (▨▨▨▨▨) are present once in all LTRs studied and are directly repeated three (•••••) and two (▨▨▨▨▨) times in the long LTRs; IAP62 and IAP19. Direct repeat sequences unique to long LTRs (▬▬▬▬▬) are also found within the R regions of IAP62 and IAP19. The repeat sequence (|—|) is present in IAP62, 19, and 81 with the middle repeat lacking 3 nucleotides from its 3' end. Specific direct repeats unique to individual LTRs are also present: IAP62, (Δ—Δ); IAP19, (▲—▲, ■—■), IAP81, (x—x); IAP14 and MIA14, (*—*); rc-mos, (o—o). Inverted complementary repeats (→, ←) are found in IAP19. They are spanning the missing sequences (—▤→, ←—) in the R regions of IAP14 and MIA14.

absent in the R regions of IAP81, IAP14, MIA14 and rc-mos LTRs. In addition, we found another 23 nucleotide stretch (CTCTCTGCGCTGCGCTCTGCGC ▬▬▬▬▬) devoid of A that is repeated three times in the R region of IAP62, twice in IAP19, but not present in IAP81, IAP14, MIA14, and rc-mos. There are also smaller direct and inverted repeats which are unique for each IAP LTR, and these often contain portions of the larger 23 bp adenine-poor direct repeat units found in IAP62 and IAP19 (Δ—Δ, ▲—▲, ■—■, Fig. 4). These are repeated two or three times, and the middle repeat is often adjacent to the last repeat and overlaps nucleotides within the last repeat unit.

There are also direct repeats specific for each short LTR which are not found in the long LTRs (x—x, *—*, o—o). These direct repeats are

paired and are often found adjacent to each other in short LTRs. It is interesting to note, that the computer aligns the largest direct repeats in IAP14, and MIA14 (*—*) to span the deleted region with one repeat on each side, but in rc-mos the repeats are both found 5' to this region (o—o), and in IAP81 the largest direct repeats are at the 3' side of the deleted region (|—|). The alignment can alternatively be made so that both direct repeats of IAP14 and MIA14 are 5' to the deletion while still maintaining greater than 80% homology to IAP62. There are also small direct repeat sequences (4-6 bp) found in these small LTRs. They are found at the U3-R junction and demonstrate the variability of sequences in this region among all the LTRs. The R regions of IAP19 and MIA14 also contain inverted repeat sequences. The possible significance of these sequences remains to be explored.

We have arranged these repeat sequences in order to find the best homology possible. It should be noted that within these larger repeats, there are smaller 3-6 bp repeat sequences. Since the R region is extremely deficient in A, it is possible to arrange these repeats with many smaller simple sequences than we have illustrated.

DISCUSSION

Since mouse intracisternal A particle genes belong to a repetitive gene family, it has long been assumed that IAP genes are basically similar in structure and are bound by conserved LTR sequences (16,17,28,35). Although some differences in restriction enzyme sites within LTR regions were noticed in several cloned IAP genes, they were attributed to random mutations after gene integration (17,29,34,35). In the present study we demonstrate that the LTR regions of IAP genes are, in fact, very heterogeneous. This heterogeneity is unusual in that it is confined to one section of the LTR, the R region. Two classes of R regions are observed whose size range from short (66-84 bp) as in IAP14 and IAP81 to very long (201-222 bp) as in IAP62 and IAP19. The long R region domains in the latter consist of several repeated oligonucleotide stretches not found in the LTRs of IAP14, MIA14, or other retroviral LTRs that we compared.

Comparing the overall sequences in the R regions of five LTRs, we find that IAP62, 19 and 81 may be grouped separately from IAP14 and MIA14. The R regions of IAP19 and IAP81 are similar to IAP62, but contain small (IAP19) and large (IAP81) deletions in their R regions. The deletions of IAP19 and IAP81 LTRs start at the same place in the 5' end of the R region (pos.232) and they contain repeats similar to IAP62 3' to the deletion (Fig. 3). On the other

hand, the R regions of IAP14 and MIA14 are almost identical in size and sequence, while the LTR of rc-mos is similar, but slightly larger. The short LTR sequences are similar to each other, but are very different from the R region sequences of IAP62, 19 and 81.

The origin and possible function of these repeated nucleotide stretches in the R domains are thus far unknown. It seems unlikely that they are generated during cloning or plasmid propagation since the sizes of LTRs in the plasmid clones are identical to those found in the original IAP gene lambda clones from the mouse genomic library (8). Furthermore, both the 5' LTR and 3' LTR sequences of IAP19 share 98% homology, and have long R domains in spite of having been propagated separately in different plasmid clones. It seems more likely that the LTR differences originated from variations in the terminal repeats (r) of the IAP RNAs which served as templates in proviral DNA synthesis by reverse transcription. The different IAP RNAs may contain either a long repeat (r_L) or a short repeat (r_S) at its termini. Reverse transcription of r_S RNA will generate a provirus with short LTRs, as found in IAP14 and MIA14, while replication from r_L RNA will yield a provirus with long LTRs as in IAP62. This interpretation, however, leaves the mechanism for the generation of IAP19 and IAP81 LTRs, which are intermediates between long and short LTRs, unanswered. The R domain in IAP19 is similar to that in IAP62, but one oligonucleotide stretch is missing, position 231-259 (Fig. 4). The R domain of IAP81, on the other hand, has a large portion of sequences deleted; in fact, the size of R in IAP81 is close to that of IAP14, but they have quite different R region sequences (Fig. 4). Since the DNA sequences between position 375 and 450 (Fig. 4) are largely conserved in IAP62, 19 and 81, they are likely transcribed from IAP RNAs with the same long r (r_L) sequence. It is feasible that imperfect reverse transcription of IAP19 and IAP81 viral RNAs, or the faulty jumping of strong stop DNA (either intramolecularly or intermolecularly) may account for the observed size differences in proviral DNA sequence of IAP19 and IAP81. A plausible mechanism for this hypothesis can be described as follows: during viral DNA synthesis, reverse transcriptase copies the 5' region of IAP RNA to produce strong stop DNA; if the reverse transcription is efficient, strong stop DNA will contain complimentary sequences of entire U5 and r_L . Subsequently, proviral IAP genes would have large LTRs as found in IAP62; if the reverse transcription is incomplete, the r_L would only be partially copied and strong stop DNA, having only the 3' end of r_L would be made. During viral DNA synthesis, this short, strong stop DNA jumps and hybridizes to the 3' end of the IAP RNA and, instead of hybridizing correctly at the 3' repetitive sequence at position 413-444, it anneals to the

repeat at position 283-311. As a result, IAP genes with short LTRs will be formed with the center of the r_L sequences deleted from the proviral DNA. IAP19 is a provirus whose LTR contains a short deletion within R, while the LTR of IAP81 has a long deletion. We are currently analyzing the repeated sequences in different IAP RNA species, 7.2 kb, 5.3 kb, and 3.5 kb, in order to ascertain whether both long and short terminal repeat sequences do exist in these IAP transcripts.

ACKNOWLEDGEMENT

This work was supported by National Institutes of Health grants 5R01AG04350 and 5T32AG00069 to RCCH.

¹Present address: Division of Biophysics, School of Hygiene and Public Health, Johns Hopkins University, Baltimore, MD 21205, USA

*To whom correspondence should be addressed

REFERENCES

1. Kelly, F., and Condamine, H. (1982) *Biochim. Biophys. Acta* **651**: 105-141.
2. Chase, D. G. and Piko, L. (1973) *J. Natl. Cancer Inst.* **51**: 1971-1973.
3. Calarco, P. G. and Szollosi, D. (1973) *Nature New Biol.* **243**: 91-93.
4. Dalton, A. J., Potter, M., and Merwin, R. M. (1961) *J. Natl. Cancer Inst.* **26**: 1221-1267.
5. Perk, K. and Dahlberg, J. E. (1974) *J. Virol.* **14**: 1304-1306.
6. Wivel, N. A., and Smith, G. H. (1971) *Int. J. Cancer* **7**: 167-175.
7. Lueders, K. K., and Kuff, E. L. (1977) *Cell* **12**: 963-972.
8. Ono, M., Cole, M. D., White, A. T., and Huang, R. C. C. (1980) *Cell* **21**: 465-473.
9. Morgan, R. A., and Huang, R. C. C. (1984) *Cancer Res.*, In Press.
10. Shen-Ong, G. L., and Cole, M. D. (1982) *J. Virol.* **42**: 411-421.
11. Kuff, E. L., Smith, L. A., and Lueders, K. K. (1981) *Mol. Cell. Biol.* **1**: 216-227.
12. Lueders, K. K., and Kuff, E. L. (1980) *Proc. Natl. Acad. Sci. USA* **77**: 3571-3575.
13. Paterson, B. M., Segal, S., Lueders, K. K., and Kuff, E. L. (1978) *J. Virol.* **27**: 118-126.
14. Hojman-Montes de Oca, F., Dianoux, L., Pewries, J., and Emanoil-Rovicovitch, R. (1983) *J. Virol.* **46**: 307-310.
15. Piko, L., Hammons, M. D., and Taylor, K. D. (1984) *Proc. Natl. Acad. Sci. USA* **81**: 488-492.
16. Cole, M. D., Ono, M., and Huang, R. C. C. (1981) *J. Virol.* **38**: 680-687.
17. Kuff, E. L., Feenstra, A., Lueders, K., Smith, L., Hawley, R., Hozumi, N., and Shulman, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**: 1992-1996.
18. Cole, M. D., Ono, M. and Huang, R. C. C. (1982) *J. Virol.* **42**: 123-130.
19. Wujcik, K. M., Morgan, R. A., and Huang, R. C. C. (1984) *J. Virol.*, In Press.
20. Ju, G., and Skalka, A. M. (1980) *Cell* **22**: 379-386.
21. Viera, J. and Messing, J. (1982) *Gene* **19**: 259-268.
22. McDonnell, M. W., Simmon, M. N., and Studier, F. W. (1977) *J. Mol. Biol.* **110**: 119-135.

23. Maxam, A. M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74: 560-564.
24. Maxam, A. and Gilbert, W. (1980) *Methods Enzymology* 65: 499-560.
25. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74: 5463-5467.
26. Wilber, W. G. and Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* 80: 726-730.
27. Kanehisa, M. I. (1982) *Nucleic Acids. Res.* 10: 183-196.
28. Kuff, E. L., Feenstra, A., Lueders, K., Rechavi, G., Givol, D., and Canaani, E. (1983) *Nature* 302: 547-548.
29. Canaani, E., Dreazen, O., Klar, A., Rechavi, G., Ram, D., Cohen, J. B. and Givol, D. (1983) *Proc. Natl. Acad. Sci. USA* 80: 7118-7122.
30. Cohen, J. B., Unger, T., Rachavi, G., Canaani, E. and Givol, D. (1983) *Nature* 306: 797-799.
31. Hishinuma, F., DeBona, P. J., Astrin, S., and Skalka, A. M. (1981) *Cell* 23: 155-164.
32. Temin, H. (1981) *Cell* 27: 1-3.
33. Varmus, H. (1982) *Science* 216: 812-820.
34. Ono, M., and Ohishi, H. (1983) *Nucleic Acids Res.* 11: 7169-7179.
35. Ono, M., Kitasato, H., Ohishi, H., and Motogayashi-Nakajima, Y. (1984) *J. Virol.* 50: 352-358.
36. Roe, B. A., Anandaraj, M. P. J. S., Chia, L. S. Y., Randerath, E., Gupta, R. C., and Randerkath, K. (1975) *Biochem. Biophys. Res. Comm.* 66: 1097-1105.
37. Weihler, H., Konig, M., and Gruss, P. (1983) *Science* 219: 626-631.
38. Nordheim, A., Rich, A. (1983) *Nature* 303: 674-679.
39. Breathnach, R., and Chambon, P. (1981) *Ann. Review Biochem.* 50: 349-383.
40. Corden, J. ., Wasyluk, B., Buchwalder, A., Sassone-Corsi, P., Keding, C. and Chambon, P. (1980) *Science* 209: 1406-1414.
41. Efstradiatis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slighton, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and Proudfoot, N. J. (1980) *Cell* 21: 653-668.
42. Leuders, K.K., Fewell, J.W., Kuff, E.L., and Koch, T. (1984) *Mol. Cell. Biol.* 4: 2128-2135.
43. Proudfoot, N. J., and Brownlee, G. C. (1974) *Nature* 252: 359-362.
44. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) *Nucleic Acids Res.* 8: 127-142.