Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 6899–6902, November 1982
Biochemistry

Human adult T-cell leukemia virus: Molecular cloning of the provirus DNA and the unique terminal structure

(cDNA clones/DNA sequences/long terminal repeat structure)

MOTOHARU SEIKI, SEISUKE HATTORI, AND MISTUAKI YOSHIDA

Department of Viral Oncology, Cancer Institute, Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan

Communicated by Takashi Sugimura, August 23, 1982

ABSTRACT **Adult T-cell leukemia virus (ATLV) is a human retrovirus closely associated with adult T-cell leukemia. The integrated provirus DNA and cDNA from virion RNA were molecularly cloned and their structures were analyzed. Clone λATM-1 of an integrated provirus DNA in the MT-1 cell line, established from adult T-cell leukemia cells by cocultivation with cord lymphocytes, contained DNA about 13,000 base pairs (bp) long and long terminal repeats (LTR) at both ends of the viral sequence that were about 8,000 bp long. These two LTR sequences were linked to cellular sequences with direct repeats of 7 bp. Each LTR consisted of 754 bp including inverted repeats of 2 bp at the ends and the T-A-T-A-A box, characteristics in common with those of LTRs of other known retroviruses. Adjacent to the 5′ LTR there was a sequence identical to the $tRNA^{Pro}$ binding site in murine leukemia virus, suggesting that $tRNA^{Pro}$ is a primer for reverse transcription of the viral genome. From these structural features, the mechanism of ATLV replication was suggested to be the same as that of other known animal retroviruses. However, the length of the small terminal repeats at the ends of the RNA genome, 228 ± 1 bases, is much longer than the lengths, up to 80 bases, of those in avian, mouse, or primate retroviruses so far analyzed. These findings suggest that ATLV should be classified in a distinct group of retroviruses with bovine leukemia virus which also makes unusually long strong-stop cDNA.**

Adult T-cell leukemia (ATL) is a unique disease of T-cell malignancy in that the birthplaces of the patients show a peculiar distribution in the southwest part of Japan (1). Recently, the retrovirus adult T-cell leukemia virus (ATLV)* was isolated (2, 3) from cell line MT-2 which had been established by cocultivation of the leukemia T cells of an ATL patient with normal cord lymphocytes (4). Furthermore, ATLV was concluded to be closely associated with the leukemia because provirus DNA was found in ATL patients but not in healthy adults and also because all sera from ATL patients tested contained antibody against ATLV protein p24 (3). In 1980, a similar retrovirus was isolated by Poiesz *et al.* (5) from a patient with cutaneous T-cell lymphoma and was reported as a new retrovirus, HTLV (6). Robert-Guroff *et al.* (7) reported immunological crossreactivity of HTLV p24 or p19 with sera from Japanese ATL patients, suggesting that ATLV and HTLV are related viruses. However, no information on the viral genome has been available.

In molecular analyses of animal retroviruses, a number of transforming genes that were acquired from cellular counterparts were identified (8, 9) and long terminal repeats (LTRs) were found in the integrated provirus genome (10) that could be involved in activation of cellular *onc*-related sequences (11). More detailed molecular analysis of ATLV genome should provide some information on the mechanism of malignant transformation of T cells and also on the origin of this human virus and its mode of transmission and replication. Therefore, we isolated a DNA clone of the integrated provirus DNA and also cDNA clones of virion RNA.

In this paper, we report studies on the nucleotide sequences of LTR in ATLV provirus DNA and the virus–cell junction. The data suggest that ATLV should be classified with bovine leukemia virus in a distinct group of retroviruses.

MATERIALS AND METHODS

DNA Cloning of Integrated Provirus of ATLV. High molecular weight DNA from cell line MT-1 (12), which produces ATLV in low titer, was digested with *Eco*RI, and DNA fragments of about 11–15 kilobase pairs (kbp) were separated by agarose gel electrophoresis. The DNA fragments were ligated to the arms of Charon 4A phage DNA (13) and subjected to *in vitro* packaging as described by Blattner *et al.* (14). About 5 × 10^5 plaques of the recombinant phages were screened with viral [^{32}P]cDNA, and one recombinant phage λATM-1 was isolated. The viral sequence had one cleavage site for *Sal* I. The two fragments produced were subcloned in pBR322, and pATM-1 and pATM-3 were obtained. For further analysis, the 5- and 8-kbp insert DNAs in these clones were purified on agarose gel after cleavage with *Eco*RI and *Sal* I.

Preparation of cDNA Clones of Viral RNA. ATLV RNA containing poly(A) was prepared from the virions (3) purified from the MT-2 cell line (4). cDNA of the viral RNA was synthesized by extending oligo(dT) primers at the 3′ end of poly(A) with reverse transcriptase. The second-strand DNA complementary to cDNA was synthesized by the method described by Taniguchi *et al.* (15). Both ends of the molecules were tailed with poly(dC) by use of terminal transferase and hybridized to the oligo(dG) tails at the *Pst* I site of pBR322. The recombinant plasmids were used for transformation of the χ1776 strain of *Escherichia coli* K-12 and two tetracycline-resistant clones, pATV-1 and pATV-2, containing the cDNA fragment were selected by colony hybridization with viral [^{32}P]cDNA.

Nucleotide Sequence Analysis. DNA fragments from the recombinant plasmids were digested with *Sma* I, *Pst* I *Hin*fI, or *Hpa* II and were labeled either at their 5′ end by using [γ-^{32}P]ATP and polynucleotide kinase as described by Maxam and Gilbert (16) or at their 3′ end by using [α-^{32}P]dCTP and the Klenow fragment of polymerase I. End-labeled DNA fragments were applied to a gel for strand separation or digested with an appropriate restriction endonuclease and the nucleotide sequences of the reisolated fragments were determined by the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: ATL, adult T-cell leukemia; ATLV, adult T-cell leukemia virus; LTR, long terminal repeat; kbp, kilobase pair(s); BLV, bovine leukemia virus.

* The independent isolates HTLV and ATLV are now known to be similar to each other; however, until the International Committee for Retrovirus Nomenclature proposes a systematic nomenclature for these viruses, we will use the name "ATLV."

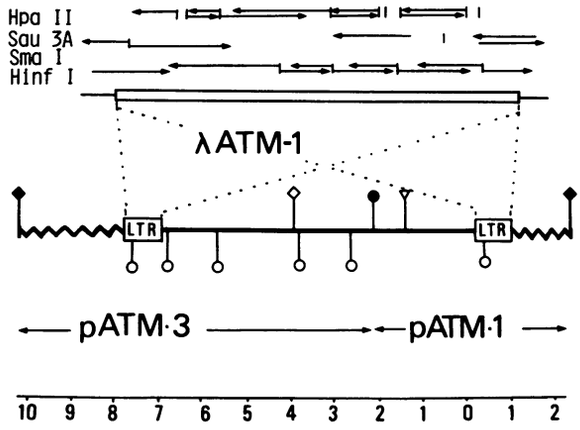

FIG. 1. Restriction map of integrated provirus of ATLV in a clone λATM-1. The regions contained in subclones pATM-1 and pATM-3 are also shown. ●, *Sal*, I; ▽, *Xho* I; ◇, *Xba* I, ○, *Sma* I. Arrows, extent and direction of the sequence obtained by cleaving the DNA with the indicated endonuclease.

procedure of Maxam and Gilbert (16).

Synthesis of Strong-Stop cDNA. ATLV virions purified from MT-2 were incubated in a mixture consisting of 50 mM Tris·HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol, dATP, dGTP, and dTTP each at 0.1 mM, various concentrations of [^{32}P]dCTP, 0.02–0.1% Nonidet P-40, and actinomycin D (50 μg/ml) at 39°C for 30 min. The cDNA was purified by digestion with proteinase K and phenol extraction followed by alkaline treatment.

RESULTS

Strategy for Determining the Sequence of the LTR Regions. A simple restriction map of molecularly cloned ATLV DNA λATM-1 was constructed (Fig. 1). To determine the location of the possible LTR sequences, the fragments formed by *Sma* I digestion were hybridized to strong-stop viral cDNA which represents the 5′ end of the viral RNA. Two fragments, which were separated by a sequence of about 8.0 kbp, were found to hybridize strongly to the probe (data not shown). For sequence determination of these regions separately, two *Sal* I fragments, each containing one of the hybridizable regions, were subcloned in pBR322, and pATM-1 and pATM-3 were obtained. Sequence determination strategy of the hybridizable regions in these clones is indicated in Fig. 1 and the sequence is shown in Fig. 2*A*.

Sequence Organization of LTR Regions and Flanking Sequences. The identical nucleotide sequences that were hybridizable to viral strong-stop cDNA in pATM-1 and pATM-3 were

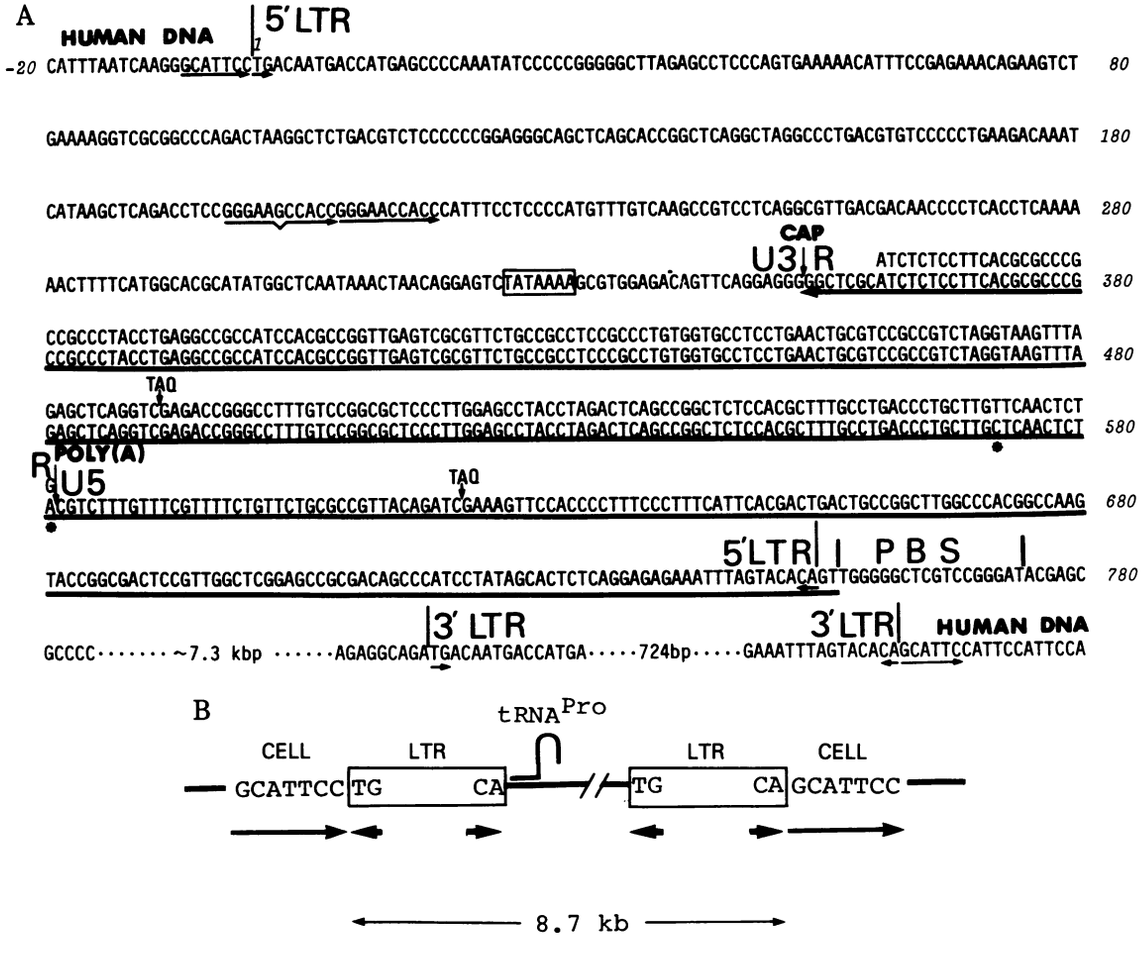

FIG. 2. (*A*) Nucleotide sequences of 5′ LTR in subclone pATM-3 and of cDNA clone pATV-1. The sequence of the (+) strand (same sense as viral RNA) is presented. The shorter sequence around the middle of LTR is that of cDNA clone pATV-1 containing 3′ end of viral RNA. Numbering is from the beginning of the LTR. Regions of interest are marked: U3, sequence from the 3′ end of viral RNA also present at the 5′ end of viral DNA; R, sequence repeated at both ends of viral RNA; U5, sequence from the 5′ end of viral RNA between R and the primer (tRNA^{Pro}) binding site; PBS, 18-bp sequence identical to the tRNA^{Pro} binding site in murine sarcoma virus/murine leukemia virus putative primer. The possible promoter sequence is shown in a box. Horizontal arrows indicate the sequences of direct or inverted repeats. The long thick underlining represents strong-stop cDNA. *Taq* I sites are shown by small vertical arrows. (*B*) Schematic illustration of the arrangement of provirus DNA in λATM-1.

defined as the LTR regions of the integrated provirus genome of ATLV (Fig. 2A). Each LTR consisted of 754 bp, and one terminus of each LTR was linked to a directly repeated cellular 7-bp sequence, G-C-A-T-T-C-C (Fig. 2B). LTR had a terminal inverted repeat of 2 bp. These structural features of the virus–cell junction are consistent with those found generally for other animal retroviruses (17) and are also similar to those of transposable elements in lower organisms (18).

In general, the LTR sequence consisted of one copy each of the 3′ end (U3) and 5′ end (U5) regions and terminal repeats (R) of the viral RNA. To determine the region of the 3′ end of viral RNA in LTR, we determined the nucleotide sequences of two cDNA clones, pATV-1 and pATV-2. These two clones contained identical sequences adjacent to a long poly(dA) stretch, thus indicating that these cDNA clones represented a sequence at the 3′ end of viral RNA, but not a sequence primed at a small oligo(dA) cluster within the viral genome. The nucleotide sequence in the cDNA clones was also found to be identical to that upstream from position 581 in LTR (Fig. 2A), except for alterations of two bases (positions 572 and 581). Thus, the poly(A) site for viral RNA was identified at position 581, defining 173 bases from this point to the end of LTR as U5. Differences of 2 bases between LTR and cDNA clones could be explained by

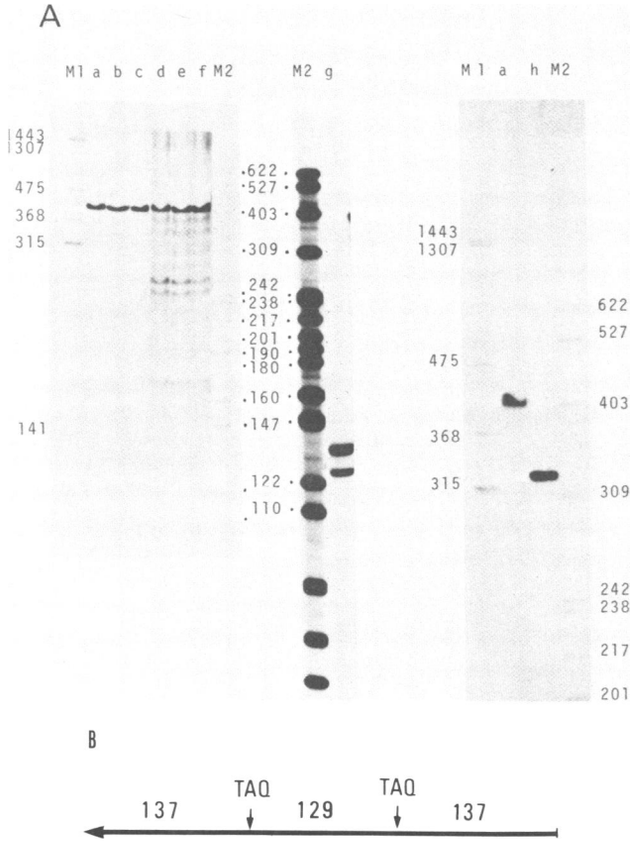

FIG. 3. Analysis of strong-stop viral cDNA of ATLV and bovine leukemia virus (BLV) (*A*) and the physical map of *Taq* I sites (*B*). (*A*) cDNAs were synthesized with 0.1 mM [^{32}P]dCTP (lanes a–c) or with 5 μM [^{32}P]dCTP (lanes d–f) in the presence of various concentrations of Nonidet P-40 (0.02% for a and d, 0.05% for b and e, 0.1% for c and f). The reaction was carried out at 39°C for 30 min. The cDNA extracted from lane a was digested with *Taq* I and run in a separate gel (lane g). ATLV cDNA in lane a was compared with that of BLV strong-stop cDNA (lane h). The size markers of pBR322 fragments formed by *Taq* I (M1) or *Hpa* II digestion (M2) are also included and the sizes were shown by nucleotide numbers. (*B*) The arrow in the map represents strong-stop cDNA and its direction of synthesis.

Table 1. Sizes of various regions in the LTR sequence

Virus	LTR	IR	R	R + U5	Primer tRNA
ATLV	754	2	228 ± 1	403 ± 1	Pro (?)
BLV	750	—	—	320 ± 1	
ALV/SV*	326–344	12	21	99	Trp
SNV*	543–615	3	79	174	Pro
MLV/SV*	515–588	11	60–68	136–151	Pro
MMTV	1,330	6	16	136	Lys

ATLV, adult T-cell leukemia virus; BLV, bovine leukemia virus; ALV/SV, avian leukosis-sarcoma virus; SNV, spleen necrosis virus; MLV/SV, Molony murine leukemia and sarcoma viruses; MMTV, mouse mammary tumor virus.

* Values for these viruses are from the review by Temin (17). IR, inverted repeat at ends of LTR in integrated provirus; R, repeated sequence at both ends of viral RNA; R + U5, strong-stop viral cDNA.

point mutations among the different isolates of ATLV: LTR clones were isolated from the MT-1 cell line established from one patient with ATLV, whereas virion RNA for cDNA clones was isolated from the MT-2 cell line established from another, unrelated, patient with ATL (4, 12).

To determine the site of the 5′ terminus of viral RNA, we analyzed the strong-stop viral cDNA. Purified ATLV virions were incubated under the conditions for general cDNA synthesis except that a higher concentration of dNTP and slightly higher temperature (39°C) were used. In this way, cDNA of a discrete size was synthesized after a short reaction period (Fig. 3A). The maximal size of the main product was 403 ± 1 nucleotides and was not affected by the concentration of detergent or substrates. Therefore, cDNA of 403 ± 1 nucleotides was concluded to be strong-stop viral cDNA which had been primed by tRNA on the genome RNA and stopped at the 5′ end of the genome. Three fragments of cDNA were obtained by digestion with *Taq* I, and these were mapped as shown in Fig. 3B. For this mapping, immature small products, which were synthesized with a low concentration of dCTP as shown in lanes d–f of Fig. 3A, were used. The sequences in LTR contained a map of *Taq* I sites identical to that in cDNA (Fig. 2A). Thus, the strong-stop cDNA corresponds to the region shown with thick underlining in Fig. 2A. Therefore, the cap site for the 5′ terminus of viral RNA was identified as position 354 ± 1, defining the upstream 353 ± 1 nucleotides as U3 and the downstream 228 ± 1 nucleotides as the terminal repeat sequence at the ends of viral RNA. This length of 228 ± 1 for the terminal repeat sequence is much longer than such values in other known retroviruses (Table 1).

DISCUSSION

General Features of Provirus and LTR Structures. The main functions of LTR in retroviruses are (*a*) reverse transcription of viral RNA, (*b*) integration of the provirus DNA into cellular DNA, and (*c*) regulation of the synthesis of viral RNA. The LTR of the ATLV provirus was shown to have the same structural organization as that of other animal retroviruses—that is, U3-R-U5. This structural organization suggests that this LTR has functions similar to those of other LTRs in viral replication.

LTR is bound by 2-base inverted repeats, T-G and C-A. Known inverted repeats in LTR are larger than 3 bases. Thus, ATLV provides an example of minimal inverted repeats at the ends, which are thought to be important for integration of the provirus DNA. The viral sequence of about 8.7 kbp in λATM-1 had LTR sequences at both ends and was linked to cellular sequences by a 7-base direct repeat sequence, G-C-A-T-T-C-C. Clone λATM-1 seemed to contain the whole viral genome

of ATLV, judging from the presence of two LTRs in the same direction.

Adjacent to 5′-LTR in pATM-3, after 2 bases there is a sequence of 18 bases identical to the $tRNA^{Pro}$ binding site in murine sarcoma virus (19) (Fig. 2A). This sequence suggests that $tRNA^{Pro}$ is a primer for reverse transcription of the viral genome as murine leukemia virus/sarcoma virus, although the sequence of $tRNA^{Pro}$ of human origin is not available.

All these structural features suggest that the mechanism of replication of ATLV is the same as that of other animal retroviruses, although the infectivity of ATLV is hard to demonstrate. The size of the viral sequence, about 8.0 kbp, is consistent with the previous findings (3) that the largest viral RNA in infected cell lines was 35 S. This size is similar to that of animal lymphatic leukosis viruses which do not contain a transforming gene.

Sequences Possibly Related to Transcriptional Control. As in other LTRs, a consensus sequence T-A-T-A-A-A, which is thought to be the signal for contributing to promote the transcriptional initiation of the proviral genome, was found at position 325–331. This sequence is located 29 ± 1 bases upstream from the cap site at position 354 ± 1, thus being consistent with the general features for sequence arrangements of T-A-T-A-A and initiation site (20). Short direct repeats of 10 bases were present at about 130 bases before the T-A-T-A-A box. This could be an enhancer signal for initiation of transcription, although the direct repeat is shorter than those in the known viral LTRs (18, 19).

The end of the terminal repeat—that is, the poly(A) site—is T-A at position 581. This T-A was changed to T-G in both cDNA clones. In any case, these ends are unique because all terminal repeat sequences of other known retroviruses end with C-A (17). Furthermore, A-A-T-A-A-A, the so-called poly(A) signal, was not present 10–30 bases upstream from the poly(A) site. Without these signal sequences, ATLV viral RNA contained poly(A) at the 3′ end, although the sequence A-A-T-A-A-A is thought to be required for polyadenylylation (21). Almost all known eukaryotic mRNAs containing poly(A) have this signal sequence, although a few exceptions have been reported: surface glycoprotein mRNA in *Trypanosoma* (22) and transcripts from the *cyc1* locus in yeast (23). Our findings together with these previous exceptions suggest that the consensus sequence A-A-T-A-A-A is essential for efficient polyadenylylation but is not absolutely required.

Unusual Structure of the ATLV LTR. The terminal repeat sequence R, which is repeated at both ends of viral RNA, was 228 ± 1 bases. This was much longer than known R sequences (Table 1), among which 79 bases in spleen nephrosis virus of chicken was the longest (17). This great difference in the size of the R sequence suggests that ATLV is a member of a different group from the known avian, mouse, or primate retroviruses. This unusually long R sequence and the absence of a poly(A) signal may be related to the inefficient replicative nature of this virus. We also analyzed strong-stop viral cDNA of BLV, which also replicates inefficiently. As shown in lane h of Fig. 3A, the strong-stop cDNA (U5 + R) was 320 ± 1 bases and, although different from that of ATLV, it too was unusually long. These results suggest that ATLV and BLV should be classified in a class of retroviruses distinct from other known retroviruses because the LTR sequences were well conserved among the viruses related (24).

This conclusion is supported by the preliminary finding that strong-stop cDNA of BLV hybridizes significantly with the ATLV LTR sequence under nonstringent conditions. The relatedness of BLV with HTLV has been reported by Oroszlan *et al.* (25) who showed that virion p24 of BLV and HTLV has a similar amino acid sequence. Because similarity between ATLV and HTLV were suggested from the immunological crossreactivity of p24 or p19 (7), our conclusion from ATLV genome analysis is consistent with that on HTLV by Oroszlan *et al.* (25). From these results, ATLV and BLV are suggested to be derived from a common ancestor of retroviruses.

The authors thank Dr. H. Sugano for valuable discussion and encouragement during this work and also Ms. Y. Hirayama for excellent technical assistance. This work was partly supported by a Grant-in-Aid for Cancer Research from the Ministry of Education, Science and Culture of Japan.

1. Uchiyama, T., Yodoi, J., Segawa, K., Takatsuki, K. & Uchino, H. (1977) *Blood* **50**, 481–492.
2. Hinuma, Y., Nagata, K., Hanaoka, M., Nakai, M., Matsumoto, T., Kinoshita, K., Shirakawa, S. & Miyoshi, I. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6476–6480.
3. Yoshida, M., Miyoshi, I. & Hinuma, Y. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2031–2035.
4. Miyoshi, I., Kubonishi, I., Yoshimoto, S., Akagi, T., Ohtsuki, Y., Shiraishi, Y., Nagata, Y. & Hinuma, Y. (1981) *Nature (London)* **294**, 770–771.
5. Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7415–7419.
6. Reitz, M. S., Poiesz, B. J., Ruscetti, F. W. & Gallo, R. C. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 1887–1891.
7. Robert-Guroff, M., Nakao, Y., Notake, K., Ito, Y., Sliski, A. & Gallo, R. C. (1982) *Science* **215**, 975–978.
8. Bishop, J. M. (1982) *Sci. Am.* **246**, 69–78.
9. Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. (1976) *Nature (London)* **260**, 170–173.
10. Hsu, T. W., Sabaran, J. L., Mark, G. E., Guntaka, R. V. & Taylor, J. M. (1978) *J. Virol.* **28**, 810–818.
11. Hayward, W. S., Neel, B. G. & Astrin, S. M. (1981) *Nature (London)* **290**, 475–480.
12. Miyoshi, I., Kubonishi, I., Sumida, M., Hiraki, S., Tsubota, T., Kimura, I., Miyamoto, K. & Sato, J. (1980) *Gann* **71**, 155–156.
13. Williams, B. G. & Blattner, F. R. (1979) *J. Virol.* **29**, 555–575.
14. Blattner, F. R., Blechl, A., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. L., Tucker, P. W. & Smithies, O. (1978) *Science* **202**, 1279–1284.
15. Taniguchi, T., Palmieri, M. & Weissman, C. (1978) *Nature (London)* **274**, 223–228.
16. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
17. Temin, H. M. (1981) *Cell* **27**, 1–3.
18. Shimotohno, K., Mizutani, S. & Temin, H. M. (1980) *Nature (London)* **285**, 550–554.
19. Beveren, C. V., Straaten, F. V., Galleshaw, J. A. & Verma, I. (1981) *Cell* **27**, 97–108.
20. Corden, J., Wasylyk, B., Buckwalder, A., Sassone-Corsi, P., Kedinger, C. & Chamborn, P. (1980) *Science* **209**, 1406–1414.
21. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
22. Majumder, H. K., Boothroyd, J. C. & Weber, H. (1981) *Nucleic Acids Res.* **9**, 4745–4753.
23. Zaret, K. S. & Sherman, F. (1982) *Cell* **28**, 563–573.
24. Lovinger, G. G. & Schochetman, G. (1980) *Cell* **20**, 441–449.
25. Oroszlan, S., Sarngadharan, M. G., Copeland, T. D., Kalyanaraman, V. S., Gilden, R. V. & Gallo, R. C. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1291–1294.