Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4813–4817, July 1989
Biochemistry

Structure of simian immunodeficiency virus regulatory genes

(*tat*/*rev*/*nef*)

SANDRA COLOMBINI*†, SURESH K. ARYA*, MARVIN S. REITZ*, LINDA JAGODZINSKI‡, BARBARA BEAVER*, AND FLOSSIE WONG-STAAL*

*Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892; and ‡Molecular Biology Department, Biotech Research Laboratories, Rockville, MD 20850

Communicated by Hilary Koprowski, March 8, 1989

ABSTRACT **Three full-length cDNA clones were obtained from cells infected with the simian immunodeficiency virus (SIV) isolated from captive macaques (SIV_{MAC}). Nucleotide sequence analyses suggested that these represented mRNA for the SIV MAC genes *tat*, *rev* (formerly, *art*/*trs*), and *nef* (formerly, 3'*orf*). The putative *tat*-specific clone was active in trans-activation of the SIV MAC long terminal repeat in COS-1 and Jurkat cells. In contrast, the human immunodeficiency virus 1 long terminal repeat was significantly trans-activated only in the COS-1 cells. This suggests that trans-activation by the SIV *tat* gene is modulated by cell-specific factors. The structure of all of the clones suggested an mRNA splicing pattern more complex than that described for human immunodeficiency virus 1.**

The human T-lymphotropic viruses share, among other common properties, the ability to regulate their own expression, mediated by trans-acting viral regulatory proteins. Human immunodeficiency virus 1 (HIV-1) has at least three regulatory genes (*tat*, *rev*, and *nef*) operating at various stages of virus expression (for a review, see ref. 1). In this report, we have adopted the nomenclature proposed by Gallo *et al.* (2). *rev* and *nef* were previously referred to as *trs*/*art* and 3'*orf*, respectively. It has been suggested that the interaction of these genes may determine the latent and replicative phases of virus infection and, therefore, have important implications in the progress of the disease. Furthermore, these regulatory genes are potential targets of anti-viral agents in addition to those applicable to all retroviruses.

Another group of primate retroviruses has been described (3, 4). The human isolates are grouped under the name HIV-2. Similar viruses have been isolated from captive macaques (5, 6) and are called simian immunodeficiency virus (SIV MAC). They show only moderate sequence homology with HIV-1 but have a very similar genomic organization (7–10). All of them possess a trans-activator gene (*tat*) able to greatly increase expression directed by the viral long terminal repeat (LTR) (10–13). We were interested in studying the trans-regulatory functions of this virus subgroup to determine if they have similar mechanisms for viral gene regulation and might, therefore, be subject to types of intervention similar to those designed for HIV-1. Furthermore, comparison of the analogous sequences of corresponding genes of the two distantly related virus subgroups should help define the functionally relevant domains. We, therefore, constructed a cDNA library from SIV MAC-infected cells using a mammalian expression vector (14, 15). In this study, we report the cloning of functional cDNA clones with trans-activation activity and present the complete nucleotide sequence of one such clone.§

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.

MATERIALS AND METHODS

cDNA Library Construction. RNA was obtained from SIV MAC-infected HUT 78 cell lines (generously provided by P. Kanki and M. Essex, Harvard School of Public Health, Boston) by the guanidine thiocyanate procedure (16). The poly(A)-selected RNA [oligo(dT)-cellulose, Collaborative Research] was copied using cloned reverse transcriptase (Life Sciences, Gainesville, FL) and the Okayama–Berg pCDV vector primer (Pharmacia) (14, 15). The cDNA was tailed, *Hin*dIII-digested and circularized utilizing oligo(dG)-tailed PL1 linkers (Pharmacia). Finally, the RNA strand was replaced with ribonuclease H (Pharmacia) and DNA polymerase I (Boehringer Mannheim). Competent HB101 cells (Bethesda Research Laboratories) were transformed with a recovery of 5×10^5 colonies. Duplicate filters of the library were screened with a U3-specific SIV LTR probe and a 17-mer oligonucleotide probe derived from the sequence of the U5 region of the SIV LTR. Four clones hybridized with both probes and were selected for further testing. In addition, several clones containing large inserts that hybridized only with the U3 probe were also studied. DNA sequencing of the selected clones was performed by the chemical degradation method of Maxam and Gilbert (17).

DNA Transfection. DNA transfection of the cDNA clones and an indicator plasmid pSVIICAT (11) was carried out by the DEAE-dextran procedure using COS-1 cells or Jurkat uninfected lymphocytes. The COS-1 cell line is derived from African green monkey kidney cells and constitutively expresses the simian virus 40 large tumor antigen (18), which supports the replication of plasmids containing the simian virus 40 origin of replication, present in the vector used here. Cytoplasmic extracts were prepared 48 hr after transfection by three cycles of freezing and thawing. Aliquots were assayed at various times for chloramphenicol acetyltransferase (CAT) activity as described (19).

RESULTS

Trans-Activation Activity of cDNA Clones. We assayed the trans-activation capacity of our cDNA clones by cotransfection with a plasmid (pSV11) containing the putative *tat* target sequences (TAR) of the SIV LTR upstream of CAT (11). In addition, cotransfection experiments using pC15CAT (13), containing the TAR of HIV-1 and/or pCV-1, a functional HIV-1 *tat* gene cDNA clone, were carried out in parallel. The results are shown in Fig. 1. Although clones P8 and P6 were inactive, clones P5 and P11 showed marked trans-activating

Abbreviations: SIV, simian immunodeficiency virus; HIV, human immunodeficiency virus; LTR, long terminal repeat; nt, nucleotide(s); ORF, open reading frame; CAT, chloramphenicol acetyltransferase.

†To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M23165).

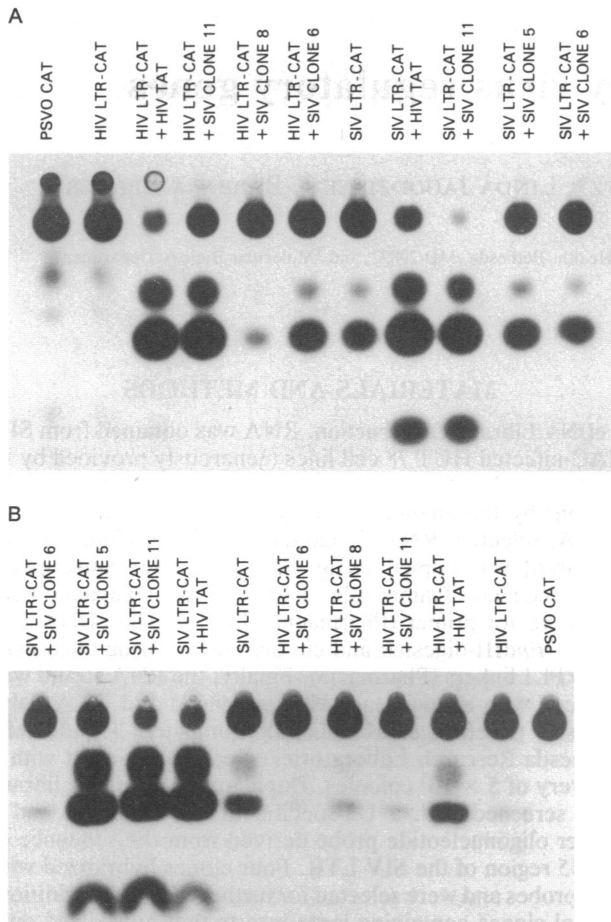

FIG. 1. Activation of SIV and HIV-1 LTR linked to the CAT gene by SIV and HIV-1 cDNA clones. *Tat*, *rev*, and *nef* are, respectively, the P11, P8, and P6 clones. Clone P5 is described in the text, the HIV-1 *tat* is the cDNA clone PCVI (13). Approximately 2×10^6 cells were transfected by the DEAE-dextran protocol and the assay was performed 48 hr later. (*A*) COS-1 cells. (*B*) Jurkat cells. Each value was determined in triplicate.

function. Consistent with the results of others using virus-infected cells (10, 11), HIV-1 *tat* trans-activated the SIV LTR in Jurkat cells, but SIV *tat* (clone P11, see Fig. 1*A*, lane 5) had very little activity with the HIV-1 LTR. In contrast, in COS-1 cells (Fig. 1*B*), we found complete reciprocity between SIV *tat*/LTR and HIV-1 *tat*/LTR pairs. These results suggest that cell-specific factors might be important in determining the specificity of trans-activation. Clone P5 is a 4-kilobase partial reverse transcript of genomic mRNA. It contains open reading frames (ORFs) for the nonstructural genes *tat*, *rev*, and *nef*. This cDNA clone also encodes the gp160 envelope protein, as demonstrated by immunoprecipitation of lysates of transfected cells (data not shown). By contrast, clones P11, P8, and P6 are all full-length transcripts (judged by the presence of U5 and 5′ R sequences) and represent spliced mRNA 2102, 1918, and 1613 nucleotides (nt) long, respectively.

Nucleotide Sequence Analysis of the cDNA Clones. We have determined the structures of the three spliced cDNA clones by obtaining the complete nucleotide sequence of clone P11 and partial sequences of clones P8 and P6. The structures of these clones are presented in Fig. 2. Clone P11 is a complete cDNA clone, starting at the cap site and ending in a $(dT\text{-}dA)_{23}$ tail. A comparison with the available genomic sequences of SIV/HIV-2 (7–10) suggested the presence of at least four splice junctions in this clone.

The first exon starts at the junction between U3 and R and utilizes a donor splice site at position +64. By comparison of the sequence of this first exon with the sequence of the HIV-1 LTR, this splice site should correspond to the end of the TAR-responsive sequence in the HIV-1 genome (19). The second exon, which contains the $tRNA^{Lys}$ binding site and the leader sequence, utilizes a 3′ (acceptor) splice site at position 209 and 5′ (donor) at nt 473 [numbering system is that of Franchini *et al.* (8)]. The donor and acceptor splice site sequences are GT and AG, respectively. This first intron of 145 nt contains the AATAAA polyadenylylation signal. The fact that the same splicing pattern is evident in cDNA clones P8 and P6 as well strongly suggests that this is a true splicing event rather than an artifact of cDNA cloning. A consensus donor splice site has not been identified in the same region in HIV-1 LTR but is present in all the HIV-2 provirus sequenced so far. The third exon of P11 utilizes an acceptor at position 5163 and a donor at position 5240 and is derived from the *vif* region. Sequences of this exon are not present in clones P8 and P6. The fourth exon of clone P11 starts in a region of the genome that was deleted in a cloned provirus previously sequenced in our laboratory (8). However, other cloned SIV and HIV-2 proviruses contain these sequences (7, 9, 10, 20). Clone P8 contains a shorter exon in this position, utilizing a 3′ splice site 214 nt further downstream, but uses the same 5′ splice site at position 5722. Clone P6 does not contain any sequences derived from this region. The fifth exon is common to all three cDNA clones and utilizes an acceptor donor site at nt 7937.

Inspection of the complete nucleotide sequence of clone P11 (Fig. 3) revealed three ORFs, each starting with a methionine codon. The positions of these three ORFs are similar to those of the *tat*, *rev*, and *nef* coding regions in a functional tat cDNA clone (pCV-1) of HIV-1 (13). The first ORF (nt 424–810) has a coding capacity of 130 amino acids and has significant amino acid identity (31.7%) with the HIV-1 tat amino acid sequence. The fact that clone P11, which contains this ORF, has strong trans-activating activity, while clones P8 and P6, which lack the ORF, do not, suggests that this ORF indeed encodes the tat of SIV.

Comparison of the SIV and HIV-1 Regulatory Genes. A comparison of the predicted amino acid sequences of the SIV

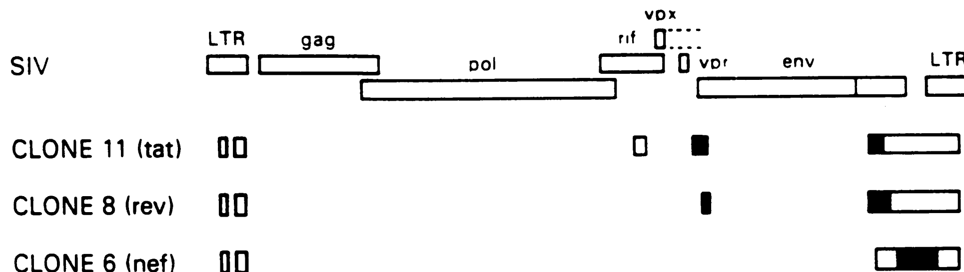

FIG. 2. Diagrammatic representation of the tat_{SIV}, rev_{SIV}, nef_{SIV} gene cDNA clones. The coding region present on the three ORFs is marked by a solid box.

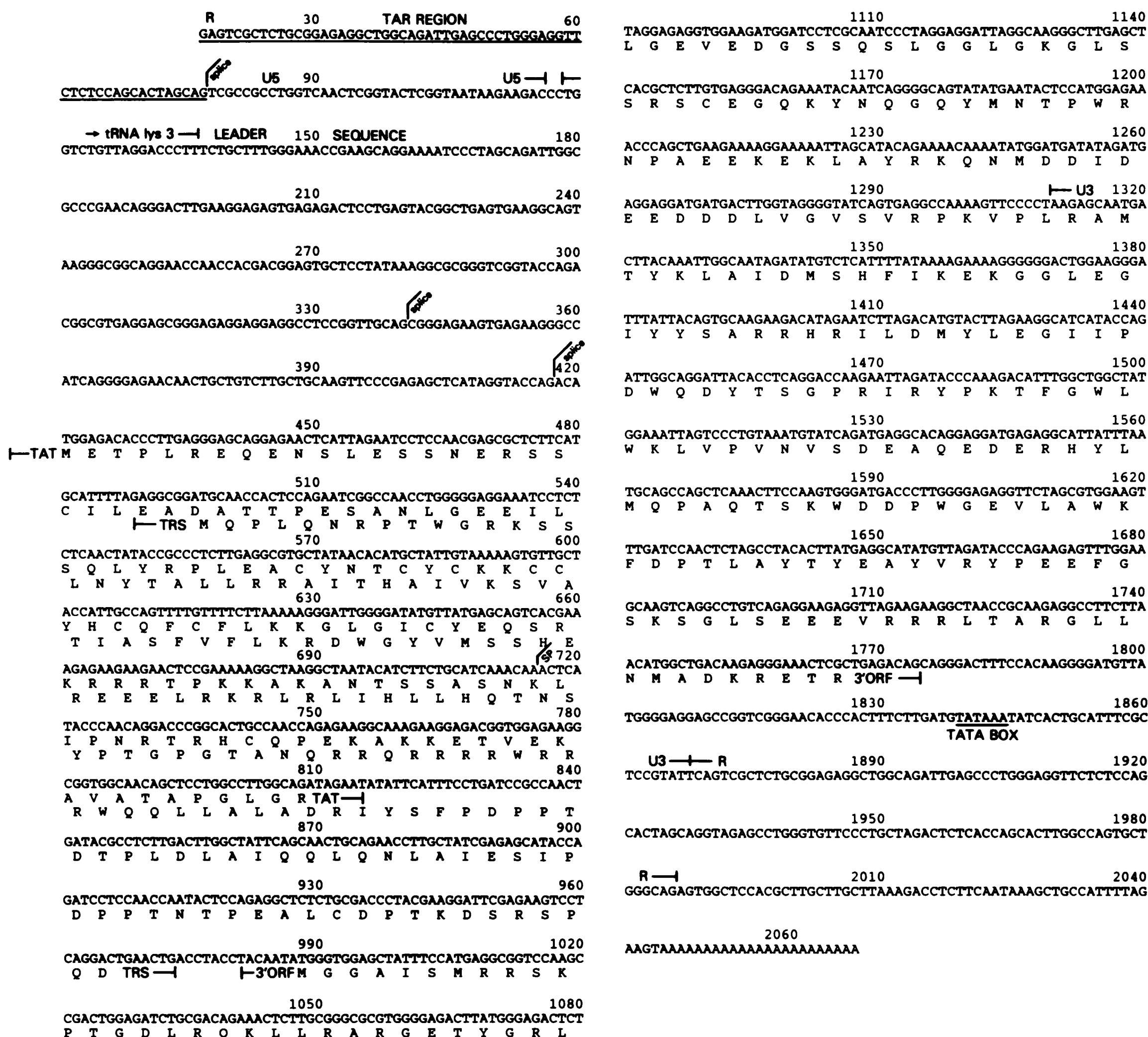

FIG. 3. Nucleotide sequence of clone P11. The first exon of the clone corresponding to HIV-1 TAR-responsive sequence by nucleotide alignment is marked as are the others—regulatory elements and the splice sites. The derived amino acid sequence is presented using the single-letter amino acid code. TRS, rev; 3'ORF, nef.

and HIV-1 *tat* genes revealed some interesting features that may be useful in localizing of the functional domains of these genes (Fig. 4*A*). The SIV *tat* gene is larger by 44 amino acids. There is little homology between the two genes except for a central core sequence that includes two potential nucleic acid binding domains. The first is a cysteine-rich region that has been shown to be involved in metal binding and dimerization of the tat protein (22). A second highly conserved potential nucleic acid binding domain is located downstream and contains a hydrophilic, highly basic stretch of lysine and arginine residues. It has been shown (13) that a *tat* gene lacking sequences on the 3' side of this hydrophilic stretch still retained significant trans-activating potential. Consistent with the idea that these 3' sequences are not essential is the low homology between HIV *tat* and SIV *tat* in this region.

The second ORF of clone P11 (nt 498–968) encodes 157 amino acids as shown (Fig. 2). Clone P8 contains a truncated version (106 amino acids) of this ORF and uses an alternate downstream 3' splice junction. Although both ORFs are potentially equivalent to the *rev* gene of HIV-1, the shorter ORF of clone P8 more closely approximates the size of the HIV-1 *rev*. Alignment of the shorter orf with rev from the first methionine residue showed a high homology with the first half of the gene, which includes a stretch rich in arginine residues (Fig. 4*B*).

The third ORF, which is present in all cDNA clones, is derived from the 3' end of the provirus genome and corresponds to the *nef* gene of HIV-1. Clone P6 contains only this ORF and presumably is derived from the mRNA of this gene. Alignment with the HIV-1 *nef* gave only 30% amino acid identity, but the overall size and hydropathy of the predicted proteins are similar. Moreover, a possible core sequence is 70% identical to the *nef* of HIV, as observed (8).

DISCUSSION

We have obtained cDNA clones corresponding to the regulatory genes (*tat*, *rev*, and *nef*) of SIV. The tat cDNA clone was functional in trans-activation of a reporter gene (CAT) linked to the SIV LTR. A complete reciprocity in trans-activation was observed between SIV and HIV-1 when the assay was carried out in COS-1 cells. However, in human T cells (Jurkat), the HIV-1 *tat* trans-activates the SIV LTR much more efficiently than is true with the reciprocal heterologous system.

Analysis of the structures of these cDNA clones suggested some distinctive features of mRNA processing for SIV. First, an intron is found within the LTR. It has not been reported that HIV-1 splices out this intron. The TAR-responsive region for HIV-2 has been found to be longer than the HIV-1

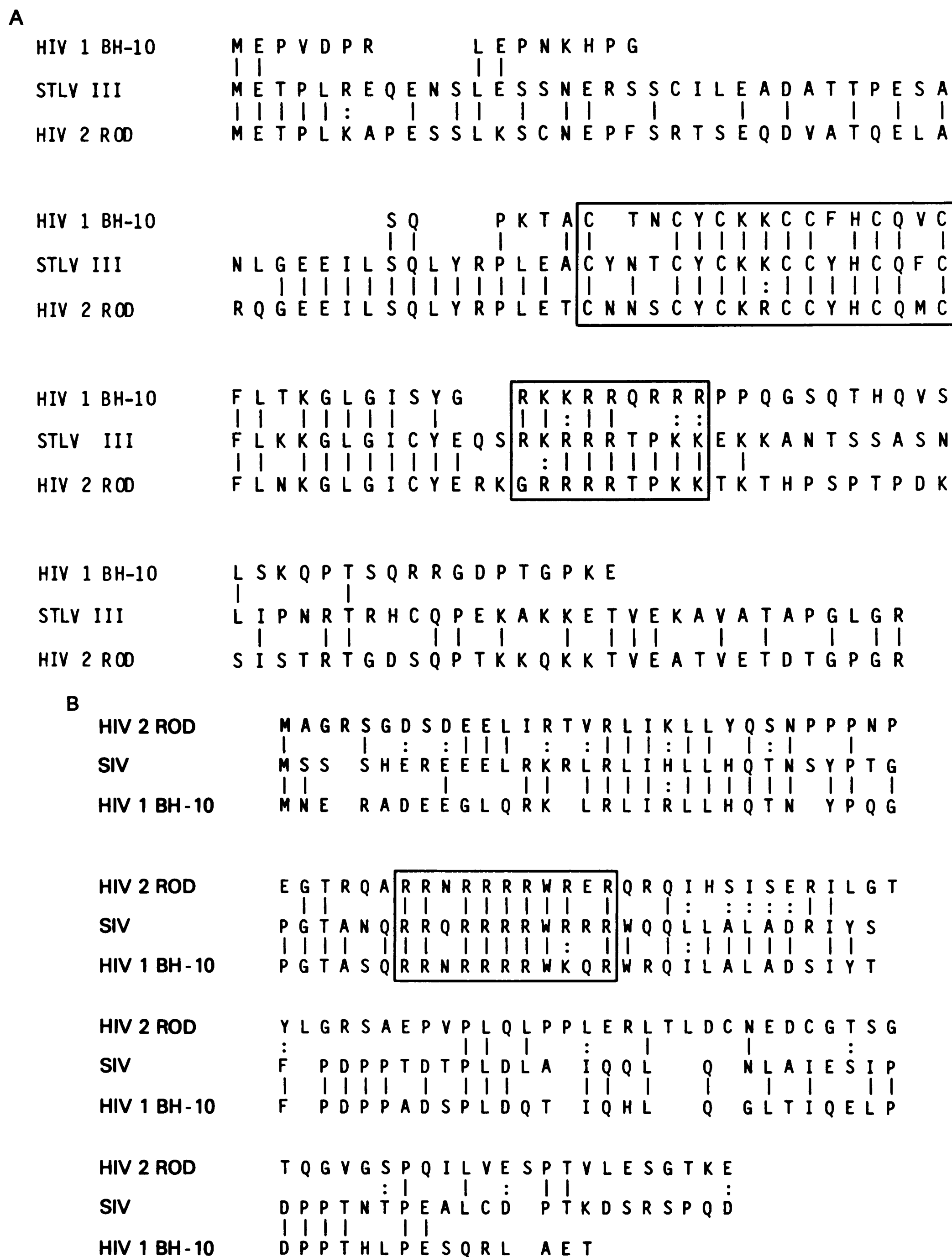

FIG. 4. Amino acid alignment for HIV-1 (strain $HTLV_{IIIB}$), SIV, and $HIV\text{-}2_{ROD}$ tat (*A*) and rev (*B*) proteins. Gaps were introduced to optimize the alignment; vertical lines indicate amino acid identity; dots indicate conservative changed. The regions for nucleotide binding capacity are enclosed in boxes. HIV-1 sequences are from Ratner *et al.* (21) and $HIV\text{-}2_{ROD}$ sequences are from Guyader *et al.* (10). The single-letter amino acid code is used.

TAR and includes a stretch of 40 nt from the U5 region that are missing in HIV-1. These 40 nt are required for full trans-activation by SIV *tat* (23). These same nucleotides would be spliced out in the cDNAs that we analyzed. It has been suggested that *tat* may interact with either viral DNA or RNA to enhance viral gene expression transcriptionally or post-transcriptionally. The fact that part of the SIV TAR region is not present in the viral mRNA may indicate that at least one mode of *tat* action is mediated either at the DNA or at the RNA level before splicing occurs. Furthermore, the splicing of this intron destroys the stem–loop structure in TAR of the mRNA, which has been suggested to inhibit efficient translation (25). Consistent with this result is our observation that SIV mRNAs are efficiently translated in cell-free systems (S.C., unpublished data). The second feature of SIV that differs from the previously accepted view of HIV-1 is that at least one form of the *rev* gene of SIV appears to be transcribed as a separate mRNA from that of *tat*, utilizing a methionine codon 214 nt downstream from the first methionine codon of its ORF. This laboratory has demonstrated (24) that *rev* of HIV-1 is indeed expressed from a distinct mRNA species.

As shown (7, 8, 11), the structural comparison of the regulatory genes of the two viral subgroups, HIV-1 and SIV/HIV-2, shows clusters of high homology. The conserved regions likely represent important functional domains of these

gene products. This suggests that SIV and HIV-2 diverged from a common ancestor more recently than from HIV-1.

1. Wong-Staal, F. (1987) in *Viral Carcinogenesis: Functional Aspects*, eds. Kjeldgaard, N. O. & Forchhammer, J. (Munksgaard, Copenhagen), pp. 55–60.
2. Gallo, R. C., Haseltine, W., Montagnier, L., Wong-Staal, F. & Yoshida, M. (1988) *Nature (London)* **333**, 504.
3. Kanki, P. J., Barin, F., M'Boup, S., Allan, J. S., Romet-Lemonne, J. L., Marlink, R., McLane, M. F., Lee, T. H., Arbeille, B., Denis, F. & Essex, M. (1986) *Science* **232**, 238–243.
4. Clavel, F., Guetard, D., Brun-Vezinet, F., Chamaret, S., Rey, M. A., Santos-Ferreira, M. O., Laurent, A. G., Dauguet, C., Katlama, C., Rouzioux, C., Klatzmann, D., Champalimaud, J. L. & Montagnier, L. (1986) *Science* **233**, 343–346.
5. Daniel, M. D., Letvin, N. L., King, N. W., Kannagi, M., Sehgal, P. K., Hunt, R. D., Kanki, P. J., Essex, M. & Desrosiers, R. C. (1985) *Science* **228**, 1201–1204.
6. Kanki, P. J., McLane, M. F., King, N. W., Letvin, N. L., Hunt, R. D., Sehgal, P., Daniel, M. D., Desrosiers, R. C. & Essex, M. (1985) *Science* **228**, 1199–1201.
7. Hirsch, V., Riedel, N. & Mullins, J. I. (1987) *Cell* **49**, 307–309.
8. Franchini, G., Gurgo, C., Guo, H. G., Gallo, R. C., Collalti, E., Fargnoli, K. A., Hall, L. F., Wong-Staal, F. & Reitz, M. S. (1987) *Nature (London)* **328**, 539–543.
9. Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M. D., Desrosiers, R. C., Tiollais, P. & Sonigo, P. (1987) *Nature (London)* **328**, 543–547.
10. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. & Alizon, M. (1987) *Nature (London)* **326**, 662–669.
11. Arya, S. K., Beaver, B., Jagodzinski, L., Ensoli, B., Kanki, P. J., Albert, J., Fenyo, E., Biberfeld, G., Zagury, J. F., Laure, F., Essex, M., Norrby, E., Wong-Staal, F. & Gallo, R. C. (1987) *Nature (London)* **328**, 548–550.
12. Clavel, F., Guyader, M., Guetard, D., Salle, M., Montagnier, L. & Alizon, M. (1986) *Nature (London)* **324**, 691–695.
13. Arya, S. K., Josephs, S. F., Guo, C. & Wong-Staal, F. (1985) *Science* **229**, 69–73.
14. Okayama, H. & Berg, P. (1983) *Mol. Cell. Biol.* **3**, 280–282.
15. Okayama, H. & Berg, P. (1982) *Mol. Cell. Biol.* **2**, 151–155.
16. Chirgwin, J. M., Przybyla, A. E., McDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
18. Gluzman, Y. (1981) *Cell* **23**, 175–182.
19. Seigel, L., Ratner, L., Josephs, S. F., Derse, D., Feinberg, M. B., Reyes, G. R., O'Brien, S. J. & Wong-Staal, F. (1986) *Virology* **148**, 226–231.
20. Hahn, B., Kong, L. I., Lee, S. W., Kuhar, P., Taylor, H. E., Arya, S. K. & Shaw, G. M. (1987) *Nature (London)* **330**, 184–186.
21. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* **313**, 277–284.
22. Frankel, A. D., Bredt, D. S. & Pabo, C. O. (1988) *Science* **240**, 70–73.
23. Emerman, M., Guyader, M., Montagnier, L., Baltimore, D. & Huesing, M. A. (1987) *EMBO J.* **6**, 3755–3760.
24. Sadaie, M. R., Rappaport, J., Benter, T., Josephs, S. F., Willis, R. & Wong-Staal, F. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9224–9228.
25. Parkin, N. T., Cohen, E. A., Darveau, A., Rosen, C., Haseltine, W. & Sonenberg, N. (1988) *EMBO J.* **7**, 2831–2837.