JOURNAL OF VIROLOGY, July 1982, p. 191–200
0022-538X/82/070191-10$02.00/0

Vol. 43, No. 1

Sequence of the Long Terminal Repeat and Adjacent Segments of the Endogenous Avian Virus Rous-Associated Virus 0

STEPHEN H. HUGHES

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724

Received 9 December 1981/Accepted 9 March 1982

Rous-associated virus 0 (RAV-0), an endogenous chicken virus, does not cause disease when inoculated into susceptible domestic chickens. An infectious unintegrated circular RAV-0 DNA was molecularly cloned, and the sequence of the long terminal repeat (LTR) and adjacent segments was determined. The sequence of the LTR was found to be very similar to that of replication-defective endogenous virus EV-1. Like the EV-1 LTR, the RAV-0 LTR is smaller (278 base pairs instead of 330) than the LTRs of the oncogenic members of the avian sarcoma virus-avian leukosis virus group. There is, however, significant homology. The most striking differences are in the U_3 region of the LTR, and in this region there are a series of small segments present in the oncogenic viruses which are absent in RAV-0. These differences in the U_3 region of the LTR could account for the differences in the oncogenic potential of RAV-0 and the avian leukosis viruses. I also compared the regions adjacent to the RAV-0 LTR with the available avian sarcoma virus sequences. A segment of approximately 200 bases to the right of the LTR (toward *gag*) is almost identical in RAV-0 and the Prague C strain of Rous sarcoma virus. The segment of RAV-0 which lies between the end of the *env* gene and U_3 is approximately 190 bases in length. Essentially this entire segment is present between *env* and *src* in the Schmidt-Ruppin A strain of Rous sarcoma virus. Most of this segment is also present between *env* and *src* in Prague C; however, in Prague C there is an apparent deletion of 40 bases in the region adjacent to *env*. In Schmidt-Ruppin A, but not in Prague C, about half of this segment is also present between *src* and the LTR. This arrangement has implications for the mechanism by which *src* was acquired. The region which encoded the gp37 portion of *env* appears to be very similar in RAV-0 and the Rous sarcoma viruses. However, differences at the very end of *env* imply that the carboxy termini of RAV-0, Schmidt-Ruppin A, and Prague C gp37s are significantly different. The implications of these observations are considered.

Rous-associated virus 0 (RAV-0), an endogenous chicken virus spontaneously produced by certain lines of inbred white leghorn chickens, is closely related to the other members of the avian sarcoma virus-avian leukosis virus (ASV-ALV) group. It has, however, a distinct host range, defined by the virus-encoded envelope gene, and, in contrast to the other ASV-ALV viruses, is nononcogenic even in long-term (over 1 year) in vivo infections of domestic chickens (7, 20).

The ASV-ALV viruses transform cells by at least two distinct mechanisms: certain oncogenic viruses that can directly and inevitably cause transformation carry homologs of cellular genes (oncogenes) (1); the ALV viruses which lack oncogenes can, at least in some cases, activate resident cellular oncogenes after integrating nearby (14, 21, 22a, 23). As a consequence, the ALV viruses are less efficient in causing neoplastic diseases than are retroviruses carrying oncogenes. RAV-0, which does not carry a cellular oncogene, is apparently also incapable of activating cellular oncogenes in domestic chickens by insertion since it does not cause tumors in these birds. There are two simple hypotheses: that the RAV-0 proviruses do not integrate in the same places that ALV proviruses integrate, i.e., near the cellular oncogene *c-myc*, or that RAV-0 proviruses can integrate in the same sites, but fail to activate *c-myc*. Whatever model is correct, there must be some underlying difference between the oncogenic ALVs and RAV-0.

Since the long terminal repeat (LTR) region of viral DNA appears to contain the sequences necessary for the initiation of transcription, termination of RNA and addition of the polyadenylate tail, and is intimately involved in integra-

tion, whichever model proves correct, the crucial difference(s) may reside within the LTR of the virus.

By T_1 oligonucleotide fingerprinting (3), hybridization (22), and analysis with restriction enzymes (25), the major differences detected between the ALVs and RAV-0 lie in the 3′ end of the RNA genome, and studies with in vivo recombinants suggest, but do not prove, that this is the region responsible for the difference in oncogenic potential between the ALVs and RAV-0 (6, 24, 29). I have cloned unintegrated circular RAV-0 DNA and shown that the DNA gives rise to replicating RAV-0 virus upon transfection. I have sequenced the LTR of this DNA and compared the RAV-0 sequence with the sequences of other members of the ASV-ALV group. It is no surprise that the LTR of RAV-0 is most closely related to the LTR of the inactive endogenous virus EV-1, but is related, although less closely, to the LTRs of the other ASV-ALV viruses. The RAV-0 LTR is significantly shorter than the LTR of RSV, being 278 (instead of 330) base pairs in length. The major differences between RAV-0 and the other replication-competent ASV-ALV viral DNAs lie in the U_3 segment of the LTR. There are minor differences in other noncoding regions: in U_5, the segment comprising the 5′ leader, and in the segment upstream from U_3, which corresponds more closely to the untranslated region of RSV between *env* and *src* than between *src* and the LTR. There are also changes in the end of the *env* gene, the most obvious of which appear to cause the RAV-0 *env* gene to terminate downstream of where the *env* gene terminates in RSV.

MATERIALS AND METHODS

Growth of virus; preparation and cloning of unintegrated RAV-0 DNA. RAV-0 virus was obtained from Harriet Robinson, and a high-titer stock was grown on line 15_B chicken embryo fibroblasts. One liter of virus was concentrated approximately 30-fold by centrifugation in a Beckman L19 rotor and used to infect two roller bottles of QT6 cells. Seventy-two hours after infection, viral DNA was prepared by the method of Hirt (15) and treated with RNase and pronase. The supercoiled forms of integrated viral DNA were enriched by a modification of the acid-phenol extraction procedure of Zasloff et al. (9, 30). The recovery of supercoiled viral DNA was monitored at various stages in the purification by gel electrophoresis, transfer to nitrocellulose, and hybridization with ^{32}P-labeled viral cDNA. Since mapping experiments did not reveal a restriction enzyme suitable for cloning in λ that cleaved the viral genome only once (25), the circular viral DNA was partially digested with *Eco*RI, which cuts circular RAV-0 DNA twice. This partially digested DNA was ligated to *Eco*RI-cleaved Charon 3A *lac* DNA. The resulting λ chimeras were packaged in vitro and plated on DP50 SupF. Approximately 0.1% of the plaques contained viral DNA. We have characterized a dozen clones, one of which appears to contain an entire copy of the viral genome and can be rescued by transfection of the cloned DNA into chicken cells. The clone (R8) has been mapped with restriction endonucleases and the transfer procedure of Southern (27), and it apparently contains a normal copy of the viral genome. R8 was cleaved to completion with *Hin*dIII and *Eco*RI and subcloned into pBR322. A clone was derived that begins in the *Hin*dIII site in *env* and ends in the *Eco*RI site near the end of *gag*. DNA from this clone was digested to completion with *Hin*dIII and treated with *Bal*31 (Bethesda Research Laboratories). The ends of the *Bal*-treated DNA were repaired with the Klenow fragment of *Pol*I and ligated to *Cla*I linkers (Collaborative Research, Inc.). A series of clones was derived with *Cla*I sites distributed every 200 to 300 base pairs in the region between *Hin*dIII and the LTR.

Sequence determination. DNA from the clones with *Cla*I linkers inserted was digested with *Cla*I and *Sst*I and labeled in the digestion mix with [α-^{32}P]dCTP (300 Ci/mmol) and the Klenow fragment of *Pol*I. The resulting fragments, labeled uniquely at the *Cla*I site, were fractionated on a 1% agarose gel, electroeluted, purified by chromatography on DEAE-Sephacel, and sequenced by the chemical procedures of Maxam and Gilbert (19). This sequence was confirmed by isolating a *Hin*dIII to *Sst*I fragment from R8, digesting to completion with *Hin*fI, labeling the resulting fragments with the Klenow fragment of *Pol*I, separating the labeled strands on a polyacrylamide gel and sequencing both separated strands, and by sequencing in both directions from the *Bst*EII site in the primer-binding site adjacent to the LTR.

RESULTS

Sequence of the RAV-0 LTR and adjacent regions. The RAV-0 LTR is 278 base pairs long, smaller than that of any other known replication-competent retrovirus. Figure 1 shows the sequence of the LTR and the adjacent regions. The RAV-0 LTR has the structural features that characterize all retrovirus LTRs (H. E. Varmus, R. Swanstrom, and D. Baltimore, *in* J. Coffin, N. Teich, H. E. Varmus, and R. Weiss, ed., *Molecular Biology of Tumor Viruses: RNA Tumor Viruses*, in press). The sequence TATA-TAA is present at a position 23 bases from the site where RNA was initiated (Fig. 1). The sequence usually associated with polyadenylation, AATAAA, is located 24 bases before the apparent site of polyadenylation. A polypurine tract (AGAGAGGGGA) is located immediately adjacent to the left end of the LTR. The sequence of the primer-binding site would suggest that, like Rous sarcoma virus (RSV) (5, 12), RAV-0 uses a tRNA trp primer. The ends of the RAV-0 LTR form a small imperfect inverted repeat (six of the first seven and seven of the first nine base pairs matched; see Fig. 1). In addition to these general structural features present in all retrovirus LTRs, the RAV-0 LTR shows considerable sequence homology with the

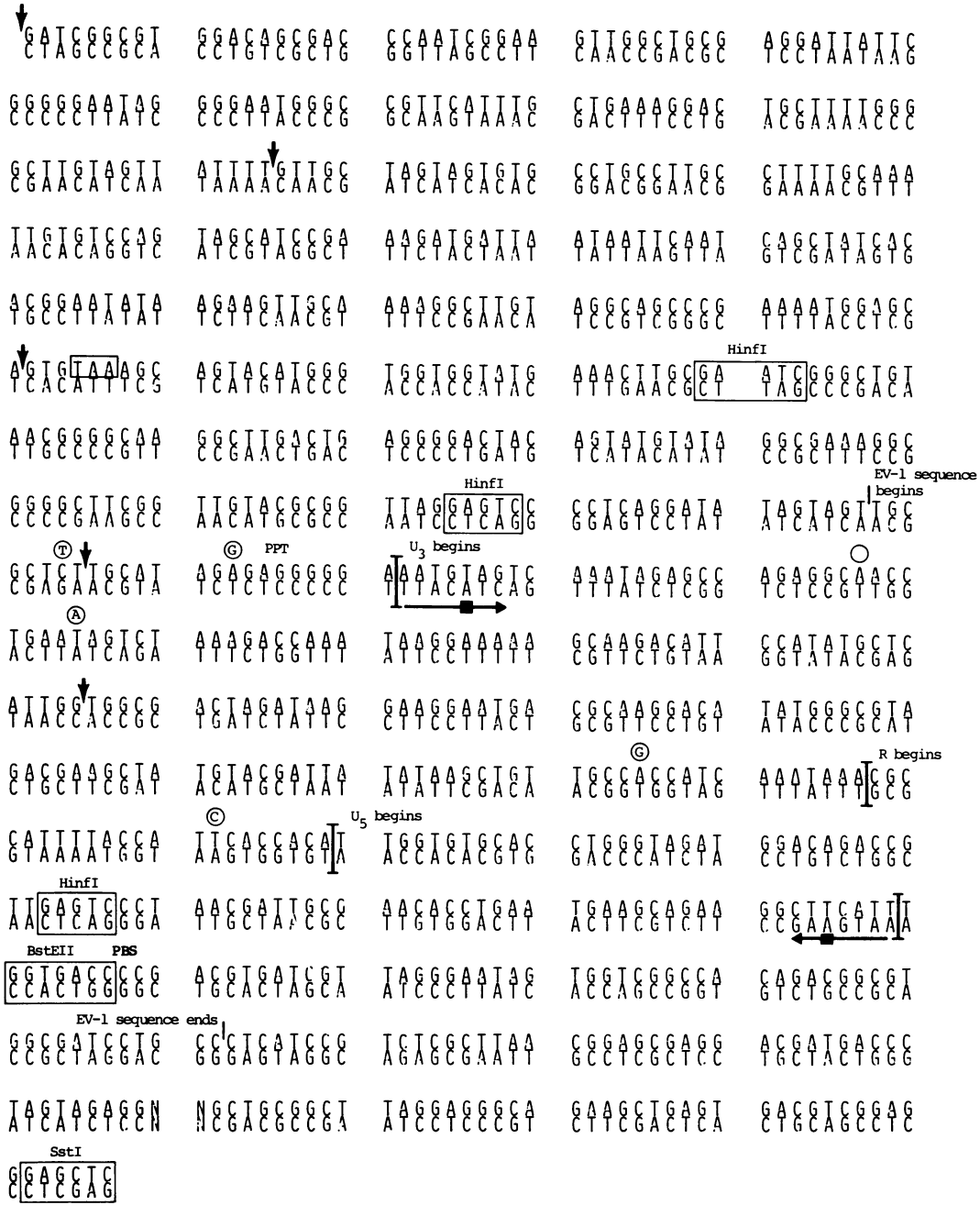

FIG. 1. Sequence of the RAV-0 LTR and adjoining regions. The sequence of the RAV-0 LTR and the adjacent regions was determined by the methods of Maxam and Gilbert (19). A series of subclones was prepared by inserting *Cla*I linkers (vertical arrows) after digestion with *Bal*31 exonuclease for various times. These subclones provided additional sequencing start sites in the regions in and around the RAV-0 LTR. Prominent structural features such as the beginning and end of the U_3, R, and U_5 regions, the polypurine tract (PPT), the primer-binding site (PBS), and the end of the *env* gene are marked on the sequence (the TAA codon at the end of *env* is boxed). The inverted repeats at the ends of the LTR are marked by arrows under the sequence. Differences between the RAV-0 sequence and the published sequence of EV-1 (5) are shown as different bases given above the RAV-0 sequence. The available EV-1 sequence is shorter than the sequence I have derived for RAV-0, and the extent of the EV-1 sequence is denoted by a pair of bars above the RAV-0 sequence. An A-T pair present in the RAV-0 sequence and absent in EV-1 is denoted by an open circle.

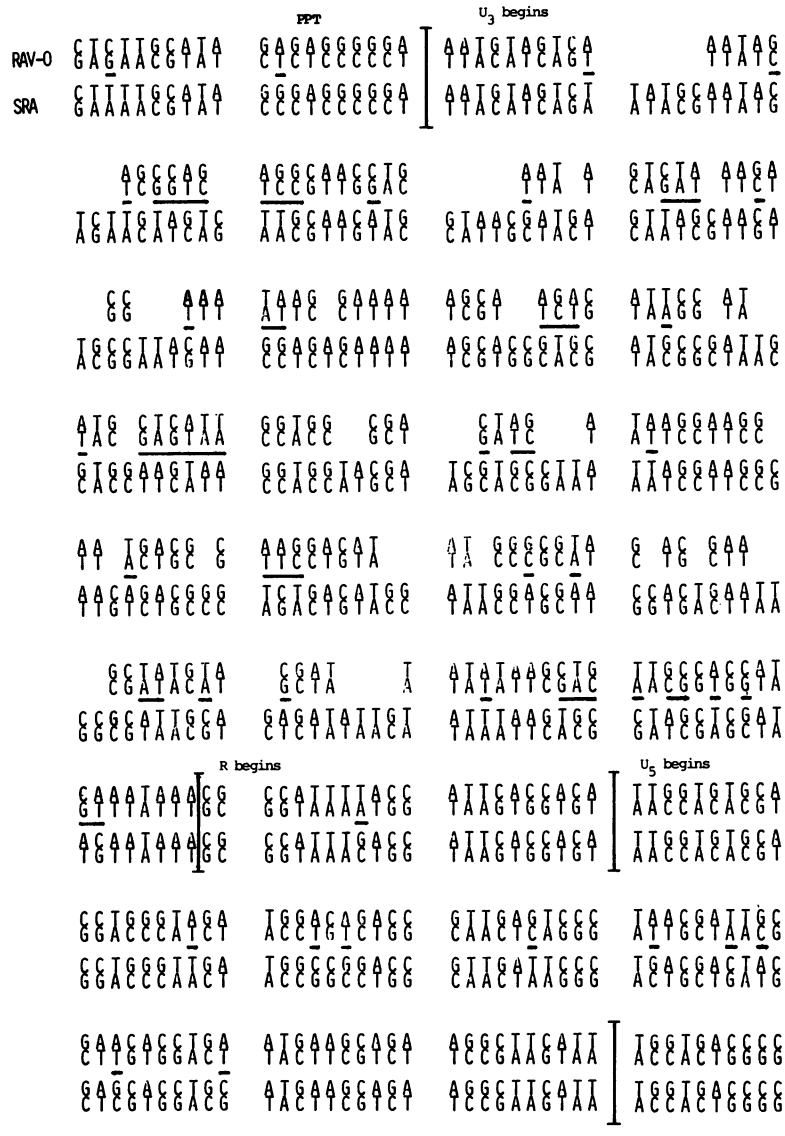

FIG. 2. Comparison of the RAV-0 LTR and the LTR of the SR-A strain of RSV (28). The LTR of RAV-0 is significantly smaller than the LTR of the other members of the ASV-ALV group. The differences reside principally in the U_3 region. I have aligned the U_3 regions wherever it was possible, leaving gaps in the RAV-0 sequences at positions where the sequences present in SR-A are obviously absent in RAV-0. In some small regions there are no obvious homologies. These regions are indicated by small bars between the RAV-0 and SR-A sequences. The distribution of bars between the sequences and of gaps within the sequences shows directly the disposition of the differences in the two sequences. The more modest changes in U_5 and R are also indicated in this same fashion. This complete sequence of the RAV-0 U_5 is in perfect agreement with a partial sequence of RAV-0 "strong stop" DNA (Swanstrom, unpublished data). PPT designates the polypurine tract.

LTRs of the other members of the ASV-ALV group. The RAV-0 LTR is strikingly similar to the LTR of EV-1 (16), a defective endogenous virus, and is less closely related to the replication-competent members of the ASV-ALV group. There are only four differences between the LTR of RAV-0 and the LTR of EV-1. Two are transitions, one is a transversion, and the fourth is an additional A-T base pair in the RAV-0 sequence (Fig. 1). This additional A-T pair, near the left end of U_3, is part of a sequence which gave, when the sequence was derived in

one direction, a serious compression artifact, and I initially misread the sequence. Since this misreading of the RAV-0 sequence gave a sequence identical to that published for EV-1, it is possible that EV-1 and RAV-0 are identical in this region and that there is an error in the published sequence of EV-1 (16). The published sequence of EV-1 includes about 60 base pairs to the right of U_5; these are identical in RAV-0. About 25 base pairs of sequence from the left of the EV-1 U_3 have been published; there are two changes from RAV-0, both transitions (Fig. 1). The simplest explanation of this similarity is that EV-1, and the other endogenous chicken viruses, arose by germ line infection of a virus very closely related to RAV-0 (10, 17, 18, 25).

Because the sequences of the EV-1 LTR and the RAV-0 LTR are nearly identical, the relationship between the RAV-0 LTR and the RSV LTR is very similar to the relationship between the LTRs of RSV and EV-1 (16). The RAV-0 LTR is 55 base pairs shorter than the RSV LTR, and the sequences could be aligned as though the RAV-0 LTR had suffered numerous small deletions or, conversely, as though the RSV LTR had acquired numerous small insertions (Fig. 2). Even when the two sequences are aligned optimally, there are still small regions of nonhomology. These could be explained as deletions from the RSV LTR, insertions into the RAV-0 LTR, or simple mutations. The T_1 oligonucleotide used by Tsichlis and Coffin (29) to distinguish the endogenous and exogenous viruses (C^X and C^n, respectively) derive from the region of U_3 immediately adjacent to R. The relevant sequences were GCCACCATCAATAACG in RAV-0, which gave rise to the T_1 oligonucleotide 08, and GATACAATAACG in Schmidt-Ruppin A (SR-A), which gave rise to the T_1 oligonucleotide C (see Fig. 2).

There is a large open reading frame ending 165 bases upstream of the LTR that is likely to encode *env*. Not only is this the only sizable open reading frame (see Fig. 4), but there is also considerable homology with the proposed *env* gene of the SR-A (8) and Prague C (Pr-C) (D. Schwartz, personal communication) strains of RSV. When these sequences are compared in the region which corresponds to the carboxy terminus of *env*, a divergence is found. The Pr-C *env* gene is longer than that of SR-A, and the sequence of the RAV-0 *env* gene, which is closer to that of SR-A than to that of Pr-C, is the longest of the three. The inferred amino acid sequence of these three *env* genes is given in Fig. 3; that of the DNA is given in Fig. 4.

We have sequenced the region between *env* and *src* in SR-A and found significant differences between our data (J. Sorge and S. Hughes, unpublished data) and those published by Czer-

RAV-0	ILEU	GLY	VAL	ASP	SER	ASP	PRO	ILEU	GLY	SER	TRP	LEU	ARG
PRC	"	"	"	"	"	"	"	"	"	"	"	"	"
SRA	"	"	"	"	"	"	LEU	"	"	"	"	"	"
	GLY	LEU	PHE	GLY	GLY	ILEU	GLY	GLU	TRP	ALA	VAL	HIS	LEU	LEU
	"	ILEU	"	"	"	"	"	"	"	"	"	"	"	"
	"	LEU	"	"	"	"	"	"	"	"	"	"	"	"
	LYS	GLY	LEU	LEU	LEU	GLY	LEU	VAL	VAL	ILEU	LEU	LEU	LEU	VAL
	"	"	"	"	"	"	"	"	"	"	"	"	"	"
	"	"	"	"	"	"	"	"	"	"	"	"	"	"
	VAL	CYS	LEU	PRO	CYS	LEU	LEU	GLN	ILEU	VAL	SER	SER	SER	ILEU
	"	"	"	"	"	"	"	"	PHE	"	"	"	"	"
	"	"	"	"	"	"	"	"	MET	LEU	CYS	GLY	ASN	ARG
	ARG	LYS	MET	ILEU	ASN	ASN	SER	ILEU	SER	TYR	HIS	THR	GLU	TYR
	"	"	"	"	"	SER	"	"	ASN	"	"	"	"	"
	"	"	"	"	"	ASN	"	"	GLU	LYS	PRO	HIS	GLY	ILEU
	LYS	LYS	LEU	GLN	LYS	ALA	CYS	ARG	GLN	PRO	GLU	ASN	GLY	ALA	VAL
	ARG	"	MET	"	GLY	GLY	ALA	VAL							

FIG. 3. Inferred amino acid sequences of the ends of the *env* genes of SR-A, Pr-C, and RAV-0. The implied amino acid sequences of the *env* genes of SR-A (8), Pr-C (Schwartz, personal communication), and RAV-0 have been aligned by maximizing homology at the nucleic acid and protein levels. The nucleic acid sequences from the region at the end of *env* are compared in Fig. 4.

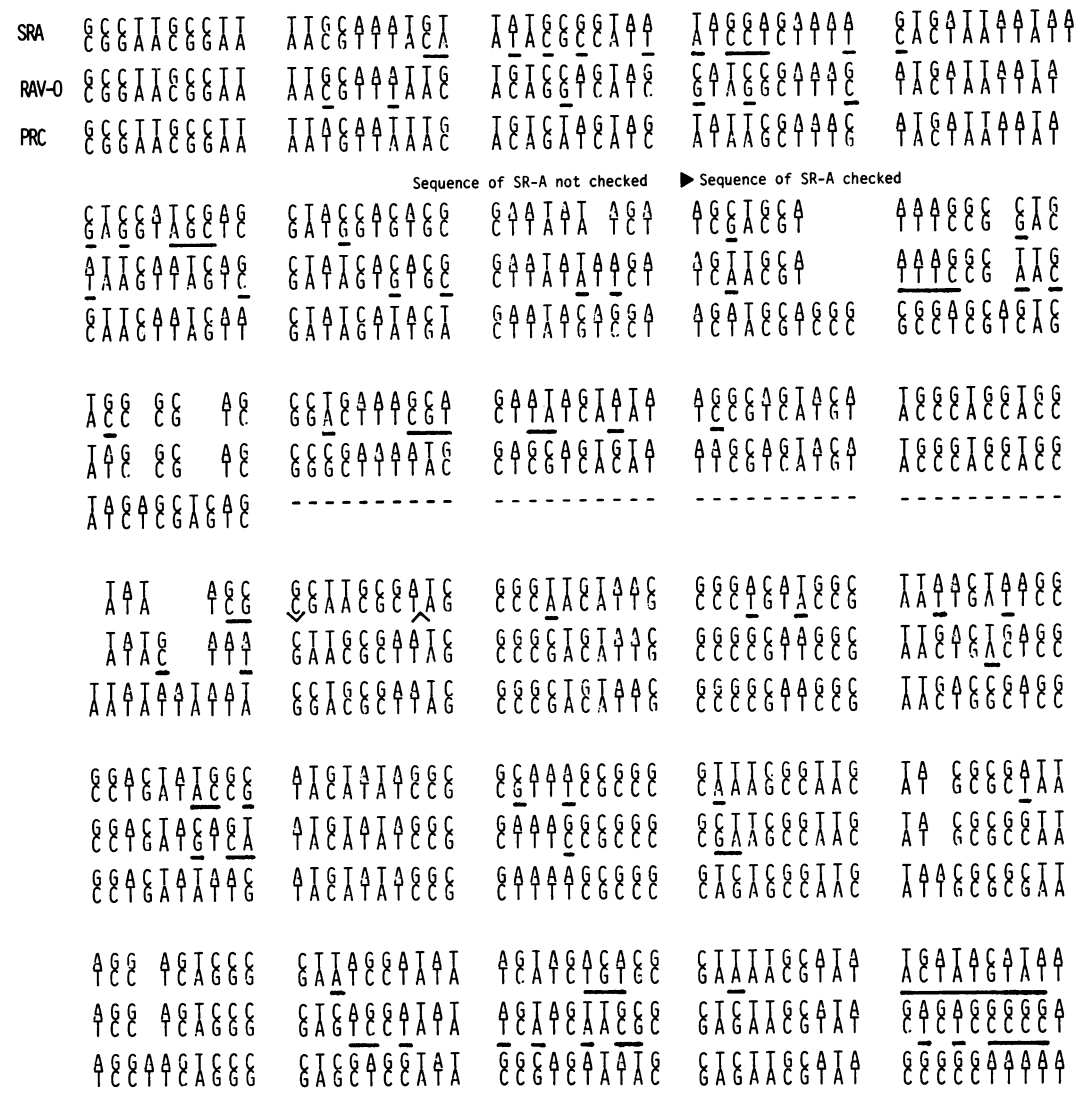

FIG. 4. Comparison of the sequences from *env* to the LTR of RAV-0 with sequences from *env* toward *src* in SR-A and Pr-C. The region of RAV-0 from the end of *env* to the LTR was aligned with the end of *env* and the intergenic region between *env* and *src* of both Pr-C (Schwartz, personal communication) and SR-A (8). We have checked a portion of the SR-A sequence in this region, using subclones derived from the same λ clone (9) sequenced by Czernilofsky et al. (8) and found significant differences in the region between *env* and *src* (J. Sorge and S. Hughes, unpublished data). The SR-A sequence given in the figure is a composite of the published data and our unpublished revision. The portion not checked is indicated in the figure, to the left of the arrowhead; those that have been checked are to the right of the arrowhead. Since a portion of the sequence found in RAV-0 and SR-A near the end of the *env* gene is apparently deleted in the Pr-C sequence (Schwartz, personal communication), I have indicated this apparent deletion by a gap in the Pr-C sequence filled in by a dotted line. Smaller deletions (usually a single base pair) are shown as gaps. Lack of homology between SR-A and RAV-0 and Pr-C and RAV-0 are indicated by bars between the RAV-0 and SR-A and the RAV-0 and Pr-C sequences.

nilofsky et al. (8) (Fig. 4). Although we did not sequence *env* in SR-A, it is possible that some of the differences between the SR-A and RAV-0 sequences are due to errors. Such errors cannot, however, account for the significant differences between the Pr-C and RAV-0 *env* sequences (Fig. 4).

In the case of SR-A, the region of homology continues beyond *env* and extends for about 150 nucleotides, which are apparently noncoding, diverging near the point in the RAV-0 sequence where the polypurine tract adjacent to the LTR begins and in SR-A about halfway between *env* and *src* (Fig. 4). A similar comparison with Pr-C

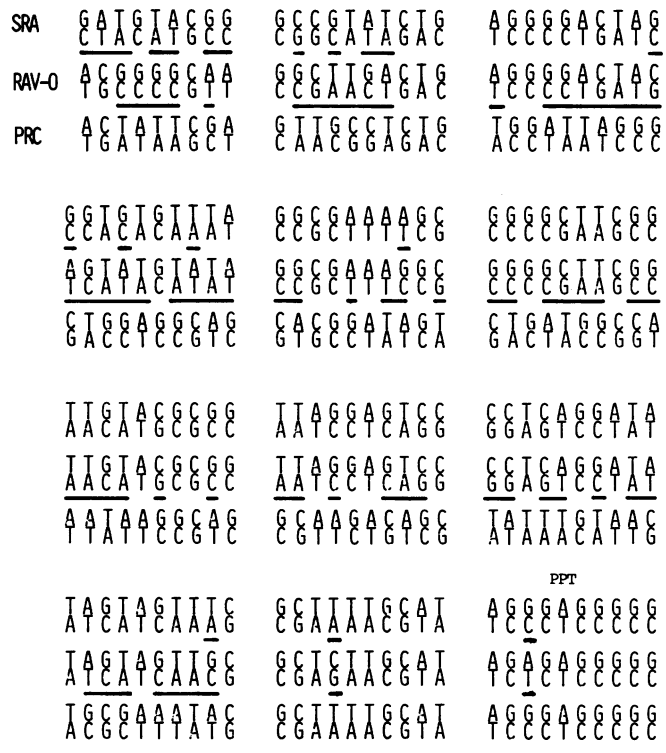

FIG. 5. Comparison of the sequences between *env* and the LTR of RAV-0 with the sequences between *src* and the LTR of Pr-C and SR-A. The regions upstream of the LTR are compared for the three viruses. Lack of homology is indicated as a bar between the sequences being compared. The polypurine tract is marked (PPT). The divergence of Pr-C from the other two viruses is clearly evident by an examination of the frequency and the extent of the barred region between the RAV-0 and Pr-C sequences.

reveals an apparent deletion, extending from the end of *env* for about 40 bases. Beyond this "deletion" the homology resumes, and there is a polypurine tract in Pr-C between *env* and *src* at a site corresponding in relative position to the polypurine tract in RAV-0 which lies adjacent to the LTR.

It is also possible to compare the sequences which lie between *src* and the LTR of the RSVs with the region between *env* and the LTR of RAV-0. Again, RAV-0 is more closely related to SR-A than to Pr-C (Fig. 5). There is in SR-A a large direct repeat which flanks *src* (8). This repeat was found to be homologous to part of the region between *env* and the LTR of RAV-0. A direct comparison of the sequences between *src* and the LTR of SR-A with the region between *env* and the LTR of RAV-0 reveals a region of homology beginning at the LTR of about 120 base pairs in length; the last 40 base pairs (approximately) adjacent to *env* in RAV-0 do not have a clear homolog in the region between *src* and the LTR of SR-A. In Pr-C, the region of homology is much shorter; only about the first 20 base pairs adjacent to the LTR are found in the region between *src* and the LTR of Pr-C. On the other side of the LTR, the untranslated leader to the right of U_5 is very similar in RAV-0 and Pr-C, although there are differences (Fig. 6). It is possible that this region is important as the recognition signal for packaging RNA into virions (26) and that the conservation of the sequence in this region reflects this requirement.

DISCUSSION

Although RAV-0 is an endogenous virus, it replicates reasonably well, and the small RAV-0 LTR thus contains all of the information required in a retrovirus LTR. Since the RAV-0 LTR is very similar to the LTR of EV-1, a defective endogenous virus which is transcribed at very low levels (13), it is unlikely that the poor transcription of EV-1 is due to a defect in the EV-1 LTR. The few changes between the EV-1 and RAV-0 LTRs are not in regions like the "TATAA" box (2), which we can recognize as being involved in transcription. In addition, EV-

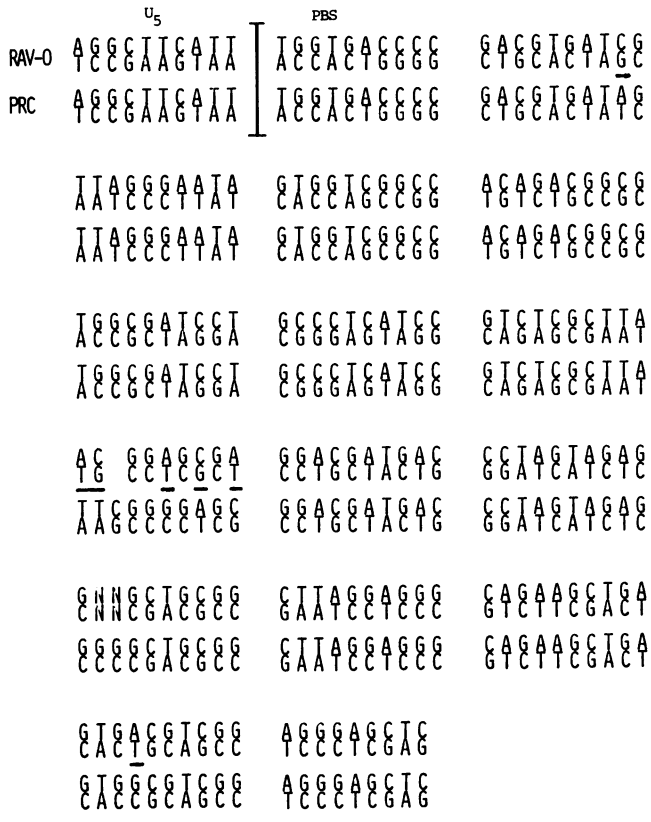

FIG. 6. Comparison of the sequence of the untranslated leader sequences between U_5 and *gag* in Pr-C and RAV-0. Most of the regions between U_5 and *gag* in Pr-C and RAV-0 are shown for comparison. Lack of homology is shown by a bar between the sequences.

1 becomes transcriptionally active after azacytidine treatment, which suggests that, under proper circumstances, the EV-1 LTR can function to initiate transcription (11). These data taken together with earlier studies suggest either that there may be contextual constraints on expression of proviruses (4) or that DNA modification may play a major role in regulation of expression (11), or both.

Even though the RAV-0 is a competent replicating virus, it does not, despite a clear homology with the ALV viruses, induce any neoplasia in domestic chickens. Previous studies suggested, but did not prove, that the crucial difference in oncogenic potential resides within the LTR (6, 24, 29). Both hybridization and T_1 oligonucleotide fingerprinting experiments show that the major differences between the ALVs and RAV-0 are in the region at the 3′ end of viral RNA, and my sequence analysis has demonstrated that the differences lie mostly within the U_3 region. The ALVs can induce leukosis by integrating next to the cellular oncogene *c-myc* and enhancing *c-myc* transcription (14, 21, 22a, 23).

RAV-0 does not induce leukosis in domestic chickens; therefore, at some level it does not duplicate an event or process produced by ALV. There are several possibilities, some more likely than others. It is possible the RAV-0 cannot infect the appropriate target cell, although this is unlikely since subgroup E ALVs are oncogenic (6, 24). It is also possible, though unlikely, that RAV-0 activates *c-myc* but kills the target cell. It seems more likely either that RAV-0 does not integrate near *c-myc* or that RAV-0 proviruses integrated near *c-myc* do not enhance transcription to a level sufficient for transformation. Differences in the U_3 region of the LTRs of ALV and RAV-0 could alter the "specificity" of integration, if there is any specificity in retrovirus integration. Alternatively, the signals for regulation of RNA transcription that reside in the U_3 region of the LTR could be responsible for changes in the ability of RAV-0 to activate adjacent genes such as *c-myc*. Comparing the U_3 regions of ALV and RAV-0 does not resolve these issues, although it is tempting to speculate that whatever the RAV-0 LTR lacks in terms of

sequence may be responsible for the concomitant loss of oncogenic potential. The fact that the differences between the ALV and RAV-0 U_3s are distributed over the whole U_3 region makes a direct test of this hypothesis difficult.

Do host factors influence the oncogenic potential of the avian viruses, differentiating between the oncogenic ALVs and RAV-0? A preliminary report from Weiss and Frisby suggests that host factors may play a key role; they have found RAV-0 to be tumorigenic in Sonnerat's jungle fowl (R. Weiss and D. Frisby, *in* D. F. Yolm, ed., *10th International Symposium for Comparative Research on Leukemia and Related Diseases*, in press). Although the sample size was small, about half of the RAV-0-infected birds died of apparent lymphoid leukosis, and several more died from "wasting disease." Even in Sonnerat's jungle fowl, however, the RAV-0-induced tumors appeared on average significantly later, and at a lower frequency, than tumors induced by the subgroup E recombinant RAV-60.

It is not yet clear why RAV-0 is oncogenic in these particular birds and not in ordinary chickens, but Sonnerat's jungle fowl completely lack RAV-0-related endogenous viruses, and RAV-0 replicates to a very high titer in these birds. Whether one of these factors, or others, as yet undiscovered, distinguishes the oncogenic susceptibility of chickens and Sonnerat's jungle fowl, it is clear the differences between the hosts, as well as differences between the viruses, can profoundly alter the incidence of tumorigenesis. This implies that a host factor, or factors, is involved. It would be particularly interesting to learn whether the RAV-0-induced tumors present in Sonnerat's jungle fowl are the result of *c-myc* activation by RAV-0.

Aside from the LTR sequences, the most obvious differences between the RSVs which have been sequenced and the region of RAV-0 DNA I have sequenced are at the very end of the *env* gene. If the sequences at the ends of the *env* genes of Pr-C (Schwartz, personal communication), SR-A (8), and RAV-0 are compared, there is a clear homology that breaks down at the very end of the *env* gene. All three *env* genes appear to terminate in different places, giving primary translation products which are very different at their carboxy termini. The dramatic differences in predicted amino acid sequence suggests that the extreme carboxy end of gp37 is probably not functional.

I was surprised to find a complete homolog of the region between *env* and the LTR of RAV-0 (excluding the polypurine tract) in the region between *env* and *src* of the SR-A strain of ASV. Beyond the apparent deletion at the end of *env*, Pr-C also contains very similar sequences. In Pr-C there is even a polypurine tract present at the same position relative to *env* as the polypurine tract which is adjacent to the LTR in RAV-0. If we assume that RAV-0 never contained *src*, then this homology presumably is the result of the way *src* was acquired by the ancestor of the RSV viruses. This antecedent virus presumably had considerable homology to RAV-0, although, unless there were additional recombination events, it had a U_3 region similar to that of the oncogenic members of the ALV-ASV group.

It is unlikely that, when *src* was first acquired, it was simply inserted into the untranslated region between *env* and the LTR, since in SR-A the *src* gene is flanked by an essentially complete copy of sequence which in RAV-0 lies between *env* and the LTR. The idea that *src* was not merely inserted into the region between *env* and the LTR is reinforced by the presence of part of U_3 upstream of *src* in the Pr-C and B77 strains of RSV (Mardon, personal communication; Swanstrom, personal communication). It is likely that two viral genomes together with cellular *src* were required to create SR-A. Through integration or the actions of reverse transcriptase or both, c-*src* was apparently added to the end of a viral genome with the concomitant loss of U_3. To complete the virus, a second copy of the sequence found between *env* and the LTR of RAV-0 was then added to the end of *src* together with U_3; almost certainly a second viral genome was the donor of this segment.

The region between the end of U_5 and the beginning of *gag* appears to be highly conserved between Pr-C RSV and RAV-0. The primer-binding site is identical, which suggests that RAV-0, like ASV, uses tRNA trp as a primer for DNA synthesis. The conservation of the rest of this region argues that there is some functional constraint on what is believed to be a noncoding region. Experiments with a mutant virus with a 150-base-pair deletion in this region imply that the region is important for efficient packaging of RNA into virions (26), a likely explanation for the conservation of these sequences.

ACKNOWLEDGMENTS

I thank Ann Mutschler for expert technical assistance and my colleagues, especially J. Sorge, G. Payne, and J. Fiddes, for helpful discussions.

This research was supported by a Public Health Service grant from the National Institutes of Health.

ADDENDUM IN PROOF

Additional sequencing in the region of SR-A which encodes the carboxy terminus of gp37 does not agree with the data published by Czernilofsky et al. (8). The new DNA sequence (J. Sorge, unpublished data) predicts that the carboxy terminus of SR-A gp37 is quite similar to the carboxy terminus of RAV-0 gp37.

LITERATURE CITED

1. **Bishop, J. M.** 1981. Enemies within: the genesis of retrovirus oncogenes. Cell **23**:5–6.
2. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
3. **Coffin, J., M. Champion, and F. Chabot.** 1978. Nucleotide sequence relationships between the genomes of an endogenous and an exogenous avian tumor virus. J. Virol. **28**:972–991.
4. **Cooper, G. M., and H. M. Temin.** 1976. Lack of infectivity of the endogenous avian leukosis virus-related genes in the DNA of uninfected chicken cells. J. Virol. **17**:422–430.
5. **Cordell-Stewart, B., J. M. Taylor, W. Rhode, H. M. Goodman, and J. M. Bishop.** 1974. Transfer RNAs of Rous sarcoma virus and the initiation of DNA synthesis by viral reverse transcriptase, p. 117–135. *In* A. Kaplan (ed.), Viral transformation and endogenous viruses, Proceedings of the 1974 Abraham Flexner Symposium. Academic Press, Inc., New York.
6. **Crittenden, L. B., W. S. Hayward, H. Hanafusa, and A. M. Fadly.** 1980. Induction of neoplasms by subgroup E recombinants of exogenous and endogenous avian retroviruses (Rous-associated virus type 60). J. Virol. **33**:915–919.
7. **Crittenden, L. B., R. L. Witter, and A. M. Fadly.** 1979. Low incidence of lymphoid tumors in chickens continuously producing endogenous virus. Avian Dis. **23**:646–653.
8. **Czernilofsky, A. P., A. D. Levinson, H. E. Varmus, J. M. Bishop, E. Tischer, and H. M. Goodman.** 1980. Nucleotide sequence of an avian sarcoma virus oncogene (*src*) and proposed amino acid sequence for gene product. Nature (London) **287**:198–203.
9. **DeLorbe, W., P. Luciw, H. M. Goodman, H. E. Varmus, and J. M. Bishop.** 1980. Molecular cloning and characterization of avian sarcoma virus circular DNA molecules. J. Virol. **36**:50–61.
10. **Frisby, D., R. A. Weiss, M. Roussel, and D. Stehelin.** 1979. The distribution of endogenous chicken retrovirus sequences in the DNA of galliform birds does not coincide with avian phylogenetic relationships. Cell **17**:623–634.
11. **Groudine, M., R. Eisenman, and H. Weintraub.** 1981. Chromatin structure of endogenous retroviral genes and activation of an inhibitor of DNA methylation. Nature (London) **292**:311–317.
12. **Harada, F., R. C. Sawyer, and J. C. Dahlberg.** 1975. A primer ribonucleic acid for initiation of *in vitro* Rous sarcoma virus deoxyribonucleic acid synthesis. J. Biol. Chem. **250**:3487–3497.
13. **Hayward, W. S., S. B. Braverman, and S. M. Astrin.** 1980. Transcriptional products and DNA structure of endogenous avian proviruses. Cold Spring Harbor Symp. Quant. Biol. **44**:1111–1121.
14. **Hayward, W. S., B. G. Neel, and S. Astrin.** 1981. Activation of a cellular *onc* gene by promoter insertion in ALV-induced lymphoid leukosis. Nature (London) **290**:475–480.
15. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
16. **Hishinuma, F., P. J. DeBona, S. Astrin, and A. M. Skalka.** 1981. Nucleotide sequence of the acceptor site and termini of the integrated avian endogenous provirus *ev-1*: integration creates a 6 bp repeat of host DNA. Cell **23**:155–165.
17. **Hughes, S. H., F. Payvar, D. Spector, G. Payne, R. T. Schimke, H. Robinson, J. M. Bishop, and H. E. Varmus.** 1979. Heterogeneity of genetic loci in chickens: analysis of endogenous viral and non-viral genes by cleavage of DNA with restriction endonucleases. Cell **18**:347–359.
18. **Hughes, S., K. Toyoshima, J. M. Bishop, and H. E. Varmus.** 1981. Organization of the endogenous proviruses of chickens: implications for origin and expression. Virology **108**:189–207.
19. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with bare-specific chemical cleavages. Methods Enzymol. **65**:499–560.
20. **Motta, J. V., L. B. Crittenden, H. G. Purchase, U. A. Stone, W. Okazaki, and R. L. Witter.** 1975. Low oncogenic potential of avian endogenous RNA tumor virus infection or expression. J. Natl. Cancer Inst. **55**:685–689.
21. **Neel, B. G., W. S. Hayward, H. C. Robinson, J. Fang, and S. M. Astrin.** 1981. Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. Cell **23**:323–334.
22. **Neiman, P., S. Das, D. MacDonnel, and C. McMillin-Helsel.** 1977. Organization of the shared and unshared sequences in the genomes of chicken endogenous and sarcoma viruses. Cell **11**:321–329.

22a. **Payne, G. S., J. M. Bishop, and H. E. Varmus.** 1982. Multiple arrangements of viral DNA and an activated host oncogene (*c-myc*) in bursal lymphomas. Nature (London) **295**:209–214.

23. **Payne, G. S., S. A. Courtneidge, L. B. Crittenden, A. M. Fadly, J. M. Bishop, and H. E. Varmus.** 1981. Analysis of avian leukosis virus DNA and RNA in bursal tumors: viral gene expression is not required for maintenance of the tumor state. Cell **23**:311–322.
24. **Robinson, H. L., M. N. Pearson, D. W. DeSimone P. N. Tsichlis, and J. M. Coffin.** 1980. Subgroup-E avian-leukosis-virus-associated disease in chickens. Cold Spring Harbor Symp. Quant. Biol. **44**:1133–1142.
25. **Shank, P., S. Hughes, and H. E. Varmus.** 1981. Restriction endonuclease mapping of the DNA of Rous associated virus O reveals extensive homology in structure and sequence with avian sarcoma virus DNA. Virology **108**:177–188.
26. **Shank, P. R., and M. Linial.** 1980. Avian oncovirus mutant (Se21Q1b) deficient in genomic RNA: characterization of a deletion in the provirus. J. Virology **36**:450–456.
27. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
28. **Swanstrom, R., W. DeLorbe, J. M. Bishop, and H. E. Varmus.** 1981. Sequencing of cloned unintegrated avian sarcoma virus DNA reveals direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. U.S.A. **78**:124–128.
29. **Tsichlis, P. N., and J. M. Coffin.** 1980. Role of the C region in relative growth rates of endogenous and exogenous avian oncoviruses. Cold Spring Harbor Symp. Quant. Biol. **44**:1123–1132.
30. **Zasloff, M., G. D. Ginder, and G. Felsenfeld.** 1978. A new method for the purification of covalently closed circular DNA molecules. Nucleic Acids Res. **5**:1139–1152.