JOURNAL OF VIROLOGY, Feb. 1984, p. 557–565
0022-538X/84/020557-09$02.00/0

Vol. 49, No. 2

Nucleotide Sequence of Noncoding Regions in Rous-Associated Virus-2: Comparisons Delineate Conserved Regions Important in Replication and Oncogenesis

DIANE BIZUB, RICHARD A. KATZ, AND ANNA MARIE SKALKA*

Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110

Received 15 August 1983/Accepted 18 October 1983

The nucleotide sequence of the regions flanking the long terminal repeat of Rous-associated virus-2 has been determined. The region analyzed spans the ends of the viral genome and includes the terminus of the *env* gene, the 3′ noncoding region, the 5′ noncoding region, and the beginning of the *gag* gene. These data have been compared with sequences available from other avian retroviruses. The comparisons reveal sections which are highly conserved and others which are quite variable. Sequence homologies within the conserved regions suggest details concerning the mode of origin of the *src*-transducing viruses. Included in the variable section is a region (XSR) found only in certain strains of Rous-derived virus. Its absence from other oncogenic viruses indicates that these sequences are not required to elicit disease.

Rous-associated virus-2 (RAV-2) is an avian leukosis virus initially isolated from stocks of Rous sarcoma virus (RSV) (7). When injected into one-day-old chickens, RAV-2 and other avian leukosis viruses cause a high incidence of bursal lymphomas after a period of 4 to 12 months. It is now known that these retroviruses, which lack an oncogene, induce such tumors by activating c-*myc*, the cellular homolog of the oncogene of MC29 virus (8, 20). This activation is presumed to depend on the transcription promoter or enhancer function encoded in the long terminal repeat (LTR) of a provirus which integrates in close proximity to c-*myc*. These proviruses are almost always defective; some contribute little more than an LTR at the c-*myc* locus (19, 20).

In a recent analysis of retroviral nucleotide sequences presumed to be required for oncogenicity, Tsichlis et al. (29) compared data obtained from a weakly oncogenic recombinant virus, NTRE-7, and its oncogenic exogenous and nononcogenic endogenous parents. These analyses showed that NTRE-7 inherited a segment which included the U_3 region of the LTR and an adjacent region of ca. 140 base pairs (called XSR) from the exogenous parent, the Prague strain of RSV (PR-RSV). It seemed most likely that the element responsible for oncogenicity in this case, as in the c-*myc* activation described above, was the promoter-enhancer function encoded in the U_3 region of the exogenous viral LTR. However, a contribution from the 3′ XSR sequences could not be excluded at that time.

Our laboratory has reported the nucleotide sequence of the RAV-2 LTR (14). In the present study we extended these analyses to include the 3′ and 5′ viral noncoding regions to determine if RAV-2 contains sequences similar to the XSR of NTRE-7. In our comparisons, we surveyed sequences from analogous regions of other retroviruses. The results show that the XSR sequence is not present in RAV-2 or several other oncogenic viral genomes studied, and thus it does not appear to be essential for oncogenicity. Our data also reveal homologies which suggest that the *src* oncogene could have been captured by RSV via a mechanism which included homologous recombination.

* Corresponding author.

MATERIALS AND METHODS

Clones. All clones used for sequencing were derived from λ RAV2-2. As previously reported (13), λ RAV2-2 was generated by inserting *Hin*dIII-digested RAV-2 covalently closed circular DNA containing one or two copies of the LTR into the λ cloning vector Charon 21A. Further subcloning involved insertion of the appropriate fragment from *Sal*I and *Bam*HI digested λ RAV2-2 DNA into plasmid (pBR322) and other phage (M13mp9) vectors. Two other RAV-2-derived clones which were utilized, mp2-R2.1(−) and mp2-R2.1(+), contain an *Eco*RI insert equivalent to a single LTR in either orientation (14).

DNA sequencing. The sequencing strategy is illustrated in Fig. 1. Sequencing by chemical cleavage was performed on both strands of the *Sal-Bam* pBR322 clone (pGJ14, provided by Grace Ju) as described by Maxam and Gilbert (18), starting about 320 base pairs away from the *Sal*I site. The dideoxy chain termination method was used with both the mp2 and mp9 M13 subclones (9, 21). Exonuclease III sequencing was done as described by Guo and Wu (6) by using replicative form I DNA from the mp9 subclone cut with *Sph*I.

RESULTS AND DISCUSSION

The region included in our sequence analysis of RAV-2 is indicated in the map in Fig. 1. We used recombinant DNA clones which originated from intracellular covalently closed circular viral DNA containing one copy of the LTR. The derived sequence (Fig. 2) runs from the end of the *env* gene through the adjacent noncoding region, the LTR, and another noncoding region which corresponds to the 5′ end of the virus. The analyses end in the sequences which encode the amino terminus of p19 in the viral *gag* gene. Table 1 lists the viral sequences we have compared and the sections included in subsequent figures and in the summary diagram (see Fig. 9).

The nucleotide and predicted amino acid sequence at the end of the *env* gene of RAV-2 and other retroviruses. Data from RAV-2 was compared with information available for two other Rous-derived viruses: a subgroup A Schmidt-Ruppin strain (SR-RSV) and, subgroup C, PR-RSV. Where relevant, data from another sarcoma virus, Y73, avian myeloblastosis virus (AMV), and the nononcogenic sub-

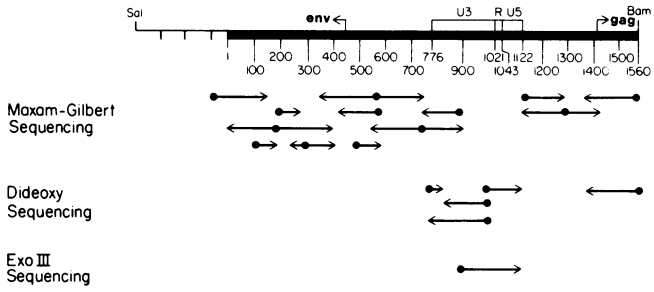
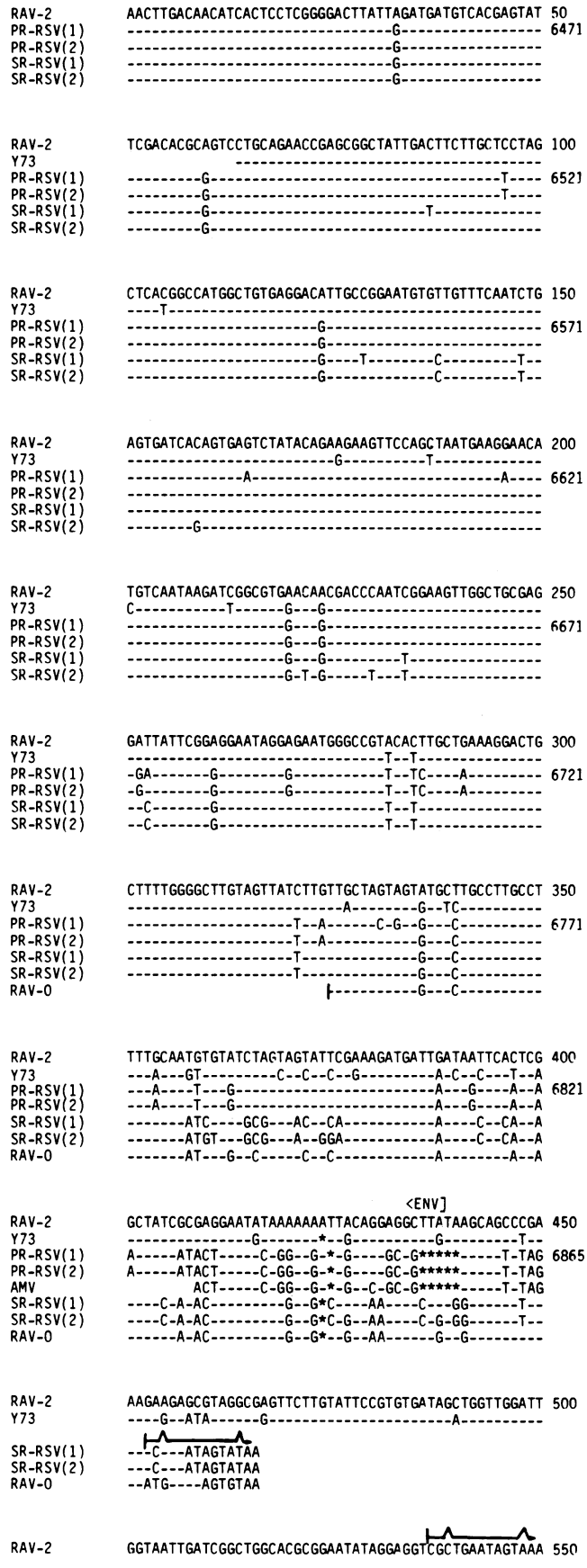


FIG. 1. Sequencing strategy. The figure shows a map of the ends of RAV-2 joined at the LTR as it occurs in clones derived from intracellular covalently closed circular viral DNA molecules. The heavy line shows the region sequenced, numbered as indicated from left to right. Results from three sequencing methods were combined (see the text). Arrows indicate the direction and origin of sections analyzed.

```
AACTTGACAACATCACTCCTCGGGGACTTATTAGATGATGTCACGAGTAT 50
TCGACACGCAGTCCTGCAGAACCGAGCGGCTATTGACTTCTTGCTCCTAG 100
CTCACGGCCATGGCTGTGAGGACATTGCCGGAATGTGTTGTTTCAATCTG 150
AGTGATCACAGTGAGTCTATACAGAAGAAGTTCCAGCTAATGAAGGAACA 200
TGTCAATAAGATCGGCGTGAACAACGACCCAATCGGAAGTTGGCTGCGAG 250
GATTATTCGGAGGAATAGGAGAATGGGCCGTACACTTGCTGAAAGGACTG 300
CTTTTGGGGCTTGTAGTTATCTTGTTGCTAGTAGTATGCTTGCCTTGCCT 350
TTTGCAATGTGTATCTAGTAGTATTCGAAAGATGATTGATAATTCACTCG 400
GCTATCGCGAGGAATATAAAAAAATTACAGGAGGCTTATAAGCAGCCCGA 450
                    <ENV]
AAGAAGAGCGTAGGCGAGTTCTTGTATTCCGTGTGATAGCTGGTTGGATT 500
GGTAATTGATCGGCTGGCACGCGGAATATAGGAGGTCGCTGAATAGTAAA 550
CTTGTAGACTTGGCTACAGCATAGAGTATCTTCTGTAGCTCTGATGACTG 600
CTAGGAAATAATGCTACGGATAATGTGGGGAGGGCAAGGCTTGCGAATCG 650
  [DIRECT REPEAT
GGTTGTAACGGGCAAGGCTTGACTGAGGGGACAATAGCATGTTTAGGCGA 700
AAAGCGGGGCTTCGGTTGTACGCGGTTAGGAGTCCCCTCAGGATATAGTA 750
```

```
 DIRECT REPEAT][   PPT   ][U3
GTTTCGCTTTTGCATAGGGAGGGGGAAATGTAGTCTTATGCAATACTCTT 800
                  [   IR   ]
GTAGTCTTGCAACATGCTTATGTAACGATGAGTTAGCAACATGCCTTATA 850
AGGAGAGAAAAAGCACCGTGCATGCCGATTGGTGGGAGTAAGGTGGTATG 900
ATCGTGGTATGATCGTGCCTTGTTAGGAAGGCAACAGACGGGTCTAACAC 950
GGATTGGACGAACCACTGAATTCCGCATTGCAGAGATATTGTATTTAAGT 1000
                CAP
          U3]|     REPEAT        [U5
GCCTAGCTCGATACAATAAACGCCATTTGACCATTCACCACATTGGTGTG 1050
          +/-
CACCTGGGTTGATGG C/T CGGACCGTTGATTCCCTGACGACTACGAGC 1096
  +/-                    U5][      PBS       ]
AC C/A TGCATGAAGCAGAAGGCTTCATTTGGTGACCCCGACGTGATCG 1142
        [   IR   ]
TTAGGGAATAGTGGTCGGCCACAGGCGGCGTGGCGATCCTGTCCTCATCC 1192
GTCTCGCTTATTCGGGGAGCGGACGATGACCCTAGTAGAGGGGGCTGCGG 1242
CTTAGGAGGGCAGAAGCTGAGTGGCGTCGGAGGGAGCCCTACTGCAGGGG 1292
GCCAACATACCCTACCGAGAACTCAGAGAGTCGTTGGAAGACGGGAAGGA 1342
AGCCCGACGACTGAGCGGTCCACCCCAGGCGTGATTCCGGTTGCTCTGCG 1392
TGATTCCGGTCGCCCGGTGGATCAAGCATGGAAGCCGTCATAAAGGTGAT 1442
                [GAG>
TTCGTCCGCGTGTAAGACCTATTGCGGGAAAACCTCTCCTTCTAAGAAGG 1492
AAATAGGGGCTATGTTGTCCCTGTTACAAAAGGAAGGGTTGCTTACGTCC 1542
CCCTCAGACTTATATTCC                                  1560
```

FIG. 2. Nucleotide sequence of the ends of the RAV-2 genome. The sequence begins 152 base pairs downstream of the start of the gp37 section in the envelope gene (*env*). *Env* ends at nucleotide 438. The region from *env* to the start of the LTR includes a section homologous to one first identified as a direct repeat (DR) on either side of *src* in SR-RSV (3). Its limits (nucleotides 654 to 765) are indicated. A polypurine tract (PPT) (nucleotides 766 to 776) lies between the DR section and the beginning of the LTR. The U_3 region of the LTR begins at nucleotide 777 and ends at nucleotide 1021. Between the end of U_3 and the beginning of U_5 is a 21-nucleotide sequence which is repeated at either end of the viral RNA genome (R). The U_5 region begins at nucleotide 1043 and ends at nucleotide 1122. The ambiguities at positions 1066 and 1099 indicate differences obtained in analysis of two M13 clones which were derived from the same fragment inserted in opposite orientations (14). At position 1066, the (+) orientation showed a C, whereas the (−) showed a T; at position 1099 the (+) orientation was a C, whereas the (−) was an A. These differences were confirmed, and we presume they were the result of random mutation within the insert. An IR region in the LTR is located between nucleotides 777 and 791 and between nucleotides 1108 and 1122. The cap site for viral mRNA is located at position 1022, and the $tRNA^{trp}$ primer binding site (PBS) is located from positions 1123 to 1140. The *gag* gene begins at position 1420.



group E virus Rous-associated virus-0 (RAV-0) are also included (Fig. 3). Orientation is made possible by comparison with PR-RSV, which has been sequenced in its entirety (22). The RAV-2 sequence starts 152 base pairs downstream from the start of gp37. At the 5′ end there are few single base differences among the sequences compared (Fig. 3). Most of these are in the third position of the codons and do not affect the amino acid sequence (Fig. 4). None of the differences relate to subgroup specificity since this information is included in the gp85 region, encoded upstream of gp37 in the *env* gene (1). The gp37 protein is believed to be responsible for anchoring gp85 in the viral envelope, and it has been proposed that an extended stretch of hydrophobic amino acids near its carboxyl end may be involved (11). This stretch is highly conserved in the viruses compared (Fig. 4), but the region following and extending to the end of gp37 is quite variable in both nucleotide and amino acid sequences. Thus, it appears that neither length nor sequence fidelity in this region is important for protein function. This variability continues into the adjacent noncoding region considered in the following section.

Comparisons of RAV-2 and other retroviral sequences in the region between *env* and the LTR. In addition to published sequences, these comparisons and those in our summary (see Fig. 9) include information on the Bryan strain of RSV (BH-RSV) which was made available to us before publication (17a). Comparisons (Fig. 5) show that the region between *env* and the LTR may be divided into two sections. Immediately following *env* is a section (RAV-2, ca. nucleotides 450 to 650) which is variable in both length and sequence. In some cases (RAV-0, AMV), the section is virtually absent; in other cases (e.g., Fujinami sarcoma virus [FSV]) the sequence is completely different from that of RAV-2. In FSV, it seems possible that the stretch of nucleotides corresponding to RAV-2 nucleotides 518 to 630 originates from the cellular *fps* locus (23). In RSV, the origin of the variable region is unclear (28). It is, therefore, noteworthy that the sarcoma virus Y73, isolated at a different time and from a different geographical location than RAV-2 (12), shows a striking similarity in all of the 3′ region sequenced, including this variable section. The similarity shown between this region of RAV-2 and BH-RSV is not unexpected since they were derived from the same original stocks of RSV (7).

The variable section in this region is followed by a section (RAV-2, nucleotides 654 to 765) that is highly conserved. It includes a stretch of ca. 100 nucleotides first identified as a direct repeat on either side of the *src* gene of SR-RSV (3) and an 11-base pair polypurine tract (PPT) presumed to play a role in plus strand strong stop DNA synthesis (10, 26).

Comparison of information from RAV-2 and that available for c-*src* (28), BH-RSV, and SR-RSV (27) reveals similarities which suggest a mechanism of origin for *src*-transducing viruses (Fig. 6). We note that sequence homology between RAV-2 and BH-RSV starts at a region 63 base pairs down-

FIG. 3. The nucleotide sequence of the end of the RAV-2 *env* gene compared with similar regions in several other retroviruses. Numbering for RAV-2 is as indicated in Fig. 1 and 2; for PR-RSV, it is as established by Schwartz et al. (22). The end of the RAV-2 *env* coding region is noted. Dashes indicate similarity with RAV-2 data. The differences are as indicated. Asterisks show deletions. The arrows show a 13-base pair homology between SR-RSV and RAV-2. It is broken at positions where homology is incomplete. References for sequence data are listed in Table 1.

TABLE 1. Retroviral sequences compared[a]

Virus	Oncogene	Envelope (*env*)	Variable (V)	DR	LTR	Leader	Source or reference
RAV-2	Leukosis	X	X	X	X	X	This work
BH-RSV	*src*		X	X			(17a)
Y73	*yes*	X	X	X	X	X	(16)
PR-RSV (1)	*src*	X	S	S	X	X	(22; cloned)
PR-RSV (2)	*src*	X					(22; cDNA)
PR-RSV (3)	*src*					X	(4)
SR-RSV (1)	*src*	X[b]	S	S	X[b]		(27)
SR-RSV (2)	*src*	X[c]				X[c]	(2)
AMV	*myb*	S	X	X	S		(17)
FSV	*fps*		X	X	S	X	(23)
RAV-0	Endogenous	X	X	X	S		(29)

[a] X, Sequences included in Fig. 4 to 7. S, Information included in Fig. 9.
[b] See reference 24.
[c] See reference 26.

stream from the termination site of the RAV-2 *env* gene. The homology includes the end of the coding sequence of the BH-RSV *src* gene. The first seven nucleotides in this region (G^G_AAGGTC) are repeated at the 3′ edge of a 39-base pair sequence, 0.9 kilobase downstream from c-*src* in what appears to be an intron region picked up in v-*src*. It has been proposed that some sort of abnormal splicing event (illustrated by the dashed lines connecting c-*src* and BH-RSV in Fig. 6) linked the end of c-*src* to the 5′ edge of this 39-base pair region (28). We suggest that the capture of *src* sequences in BH-RSV might have involved recombination at the 3′ edge of this 39-base pair region within the 7-base pair homology with RAV-2. An alternative possibility, that this region of homology with v-*src* was picked up by RAV-2 through recombination with BH-RSV, seems unlikely, since the seven-base pair homology is also present in Y73 (see Fig. 5), an independent transforming virus which has not been passaged with RAV-2. As with general recombination between viral genomes, the recombination between RAV-2 and c-*src* might be facilitated by incorporation of a c-*src*-containing transcript into virus particles along with the viral genome(s). Plausible mechanisms by which this may occur have been proposed (5, 25). They suggest, as a starting point, that a provirus is integrated into host DNA upstream of the locus of the gene which will eventually be transduced. Subsequent deletion of a stretch of DNA between the two, including the end of the provirus and the beginning of the cellular gene, leads to a fusion of viral and cellular transcription units. Transcription of this fused region, driven now by the left-end viral LTR, can produce an RNA product containing cellular gene exons and some viral sequences, including those required for packaging within a virion. The second recombination event could occur during reverse transcription in subsequent infection by this virion (15). We suggest that in the case of BH-RSV this second crossover was facilitated by the short region of homology in the cellular sequence. The resulting provirus would contain the cellular (*src*) gene flanked on either side by retroviral sequences. It is possible that, in some cases, the chromosomal deletions which fuse viral and chromosomal transcription units also involve recombination between short stretches of homology (30).

Figure 6 shows a second stretch of nucleotides at the end of *src* with interesting partial homology. This stretch begins immediately adjacent to the seven-base pair homology with c-*src* and consists of 13 nucleotides which are repeated in the end of the *src* (9 of 13) and *env* (11 of 13) genome of SR-RSV. By analogy with BH-RSV, this suggests that capture of *src* by SR-RSV may have involved a recombination event at the end of *env* of a helper virus. The acquisition of the XSR and

```
RAV-2       NLTTSLLGDLLDDVTSIRHAVLQNRAAIDFLLLAHGHGCEDIAGMCCFN
Y73                                 ----------------------------
PR-RSV(1)   ------------------------------------------V-------
PR-RSV(2)   ------------------------------------------V-------
SR-RSV(1)   ------------------------------------------V-------
SR-RSV(2)   ------------------------------------------V-------

RAV-2       LSDHSESIQKKFQLMKEHVNKIGVNNDPIGSWLRGLFGGIGEWAVHLLK
Y73         ---------R--------------DS-----------------------
PR-RSV(1)   ----------------K-------DS---------I-------------
PR-RSV(2)   ------------------------DS-----------------------
SR-RSV(1)   ------------------------DS-----------------------
SR-RSV(2)   ---Q---------------------DS-L---------------------

RAV-2       GLLLGLVVILLLVVCLPCLLQCVSSSIRKMIDNSLGYREEYKKITGGL
Y73         --------------------V---------N--FS----C--LQEACKQPERGI
PR-RSV(1)   -----------L--------F---------NS-IN-HT--R-MQ--AV
PR-RSV(2)   --------------------F---------NS-IN-HT--R-MQ--AV
SR-RSV(1)   --------------------I-CGN-----N--IS-HT----LQKAYGQPESRIV
SR-RSV(2)   --------------------MLCGNR----N--IS-HT----LQKACGQPESRIV
RAV-0                ----------I---------N--IS-HT----LQKACRQPENGAV
```

FIG. 4. Amino acid sequence at the end of *env* as predicted from the nucleotide sequence data. The single-letter amino acid letter code is used. Dashes indicate homology; only substitutions are noted. The boxed region identifies a 27-amino acid long hydrophobic region thought to function as a membrane anchor for the viral glycoprotein complex (11).

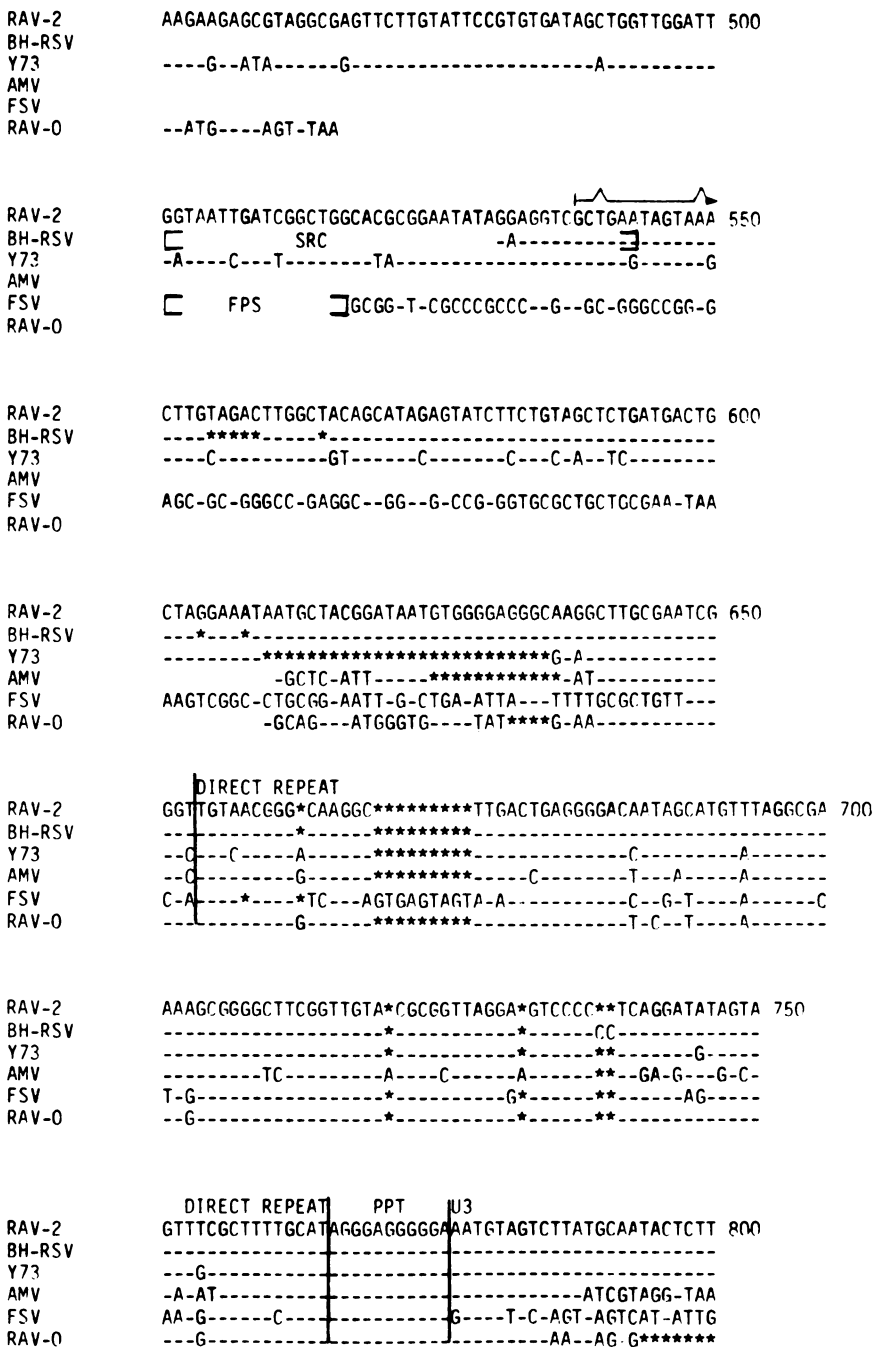


FIG. 5. Comparison of RAV-2 and other retroviral sequences in the region between *env* and the LTR. Symbols follow the convention established in Fig. 3. Blank spaces or entire lines (e.g., AMV, FSV, RAV-0) indicate lack of sequence which may be interpreted as extended deletions compared with the RAV-2 sequence. The arrow over nucleotides 537 to 550 is explained in the legend to Fig. 3. The brackets enclose the coding sequences for the oncogenes *src* (see Fig. 6) and *fps*.

the direct repeat (DR) regions between the *env* and *src* genes of SR-RSV (see Fig. 9) could be the result of subsequent interactions during passage of the virus.

Comparison of the LTR regions of RAV-2 and other retroviruses. In the course of sequence analysis of the RAV-2 noncoding regions, many fragments were generated which spanned the LTR. Data generated from analysis of these fragments revealed errors in our previously reported sequence (14) which we now believe to be attributable mainly to misreadings of the C and T lanes in the dideoxy analyses used in those studies. The data shown in Fig. 7 serve to correct these errors as well as for comparison of the RAV-2 LTR with other Rous-derived and the Y73 retroviral LTR sequences. There is extensive homology among the LTRs of these viruses. Moreover, sequence conservation is absolute in the regions of the PPT and at both ends of the LTR in a section which includes the long (12 of 15 base pair) inverted repeat (IR). Thus, we conclude that variations are not tolerated in these sites believed to be important in plus strand DNA synthesis (PPT) and the integration of viral

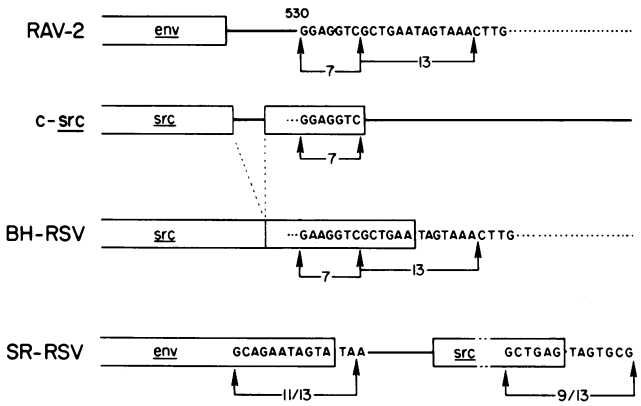


FIG. 6. Sequence homologies between RAV-2, c-*src*, BH-RSV, and SR-RSV. The sequence shown for RAV-2 starts 62 base pairs downstream from the end of *env*. The first seven base pairs are also found in the c-*src* locus, at the edge of the region also present in v-*src*. Recombination within the seven-base pair region would generate the sequence shown for BH-RSV, in which the following two codons and the termination signal are derived from the noncoding region of the viral genome. Nucleotide sequences from the seven-base pair region to the 3′ termini of RAV-2 and BH-RSV are very similar (cf. Fig. 9). Adjacent to the 7-base pair region is a 13-base pair sequence in RAV-2 and BH-RSV, which is repeated (11 of 13) in the end of the *env* and *src* (9 of 13) genes of SR-RSV.

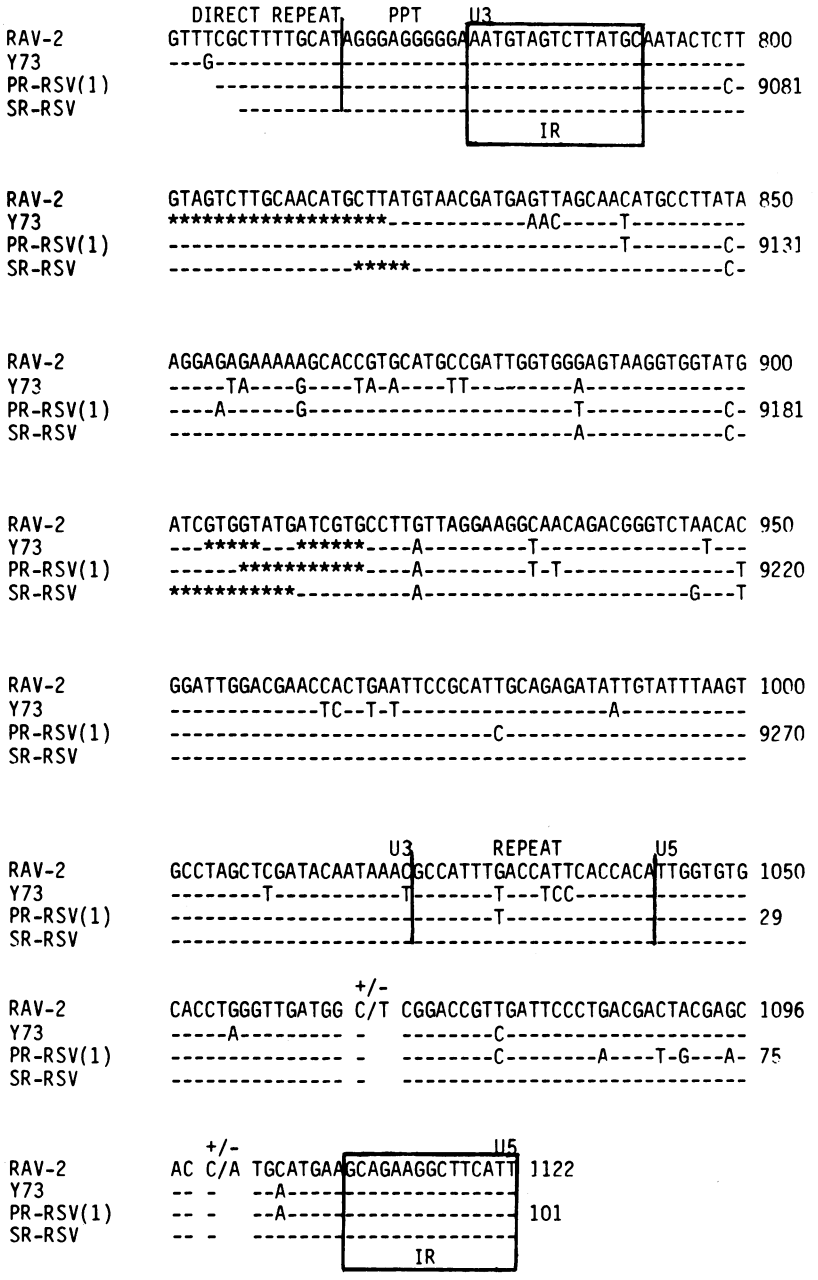


FIG. 7. Comparison of the LTR sequence of RAV-2 with three other retroviruses. Symbols follow the convention established in Fig. 3. In RAV-2, the U_3 of the LTR begins at nucleotide 777 and ends at nucleotide 1021. REPEAT is a 21-base pair sequence between U_3 and U_5 which is repeated in the RNA at either end of the viral genome. The U_5 begins at nucleotide 1043 and ends at nucleotide 1122. IR indicates the 15-base pair inverted repeats, located at the beginning of U_3 and at the end of U_5, which are highly conserved in this set. The U_3 regions of the LTRs listed represent one of the three possible types identified in Fig. 9.

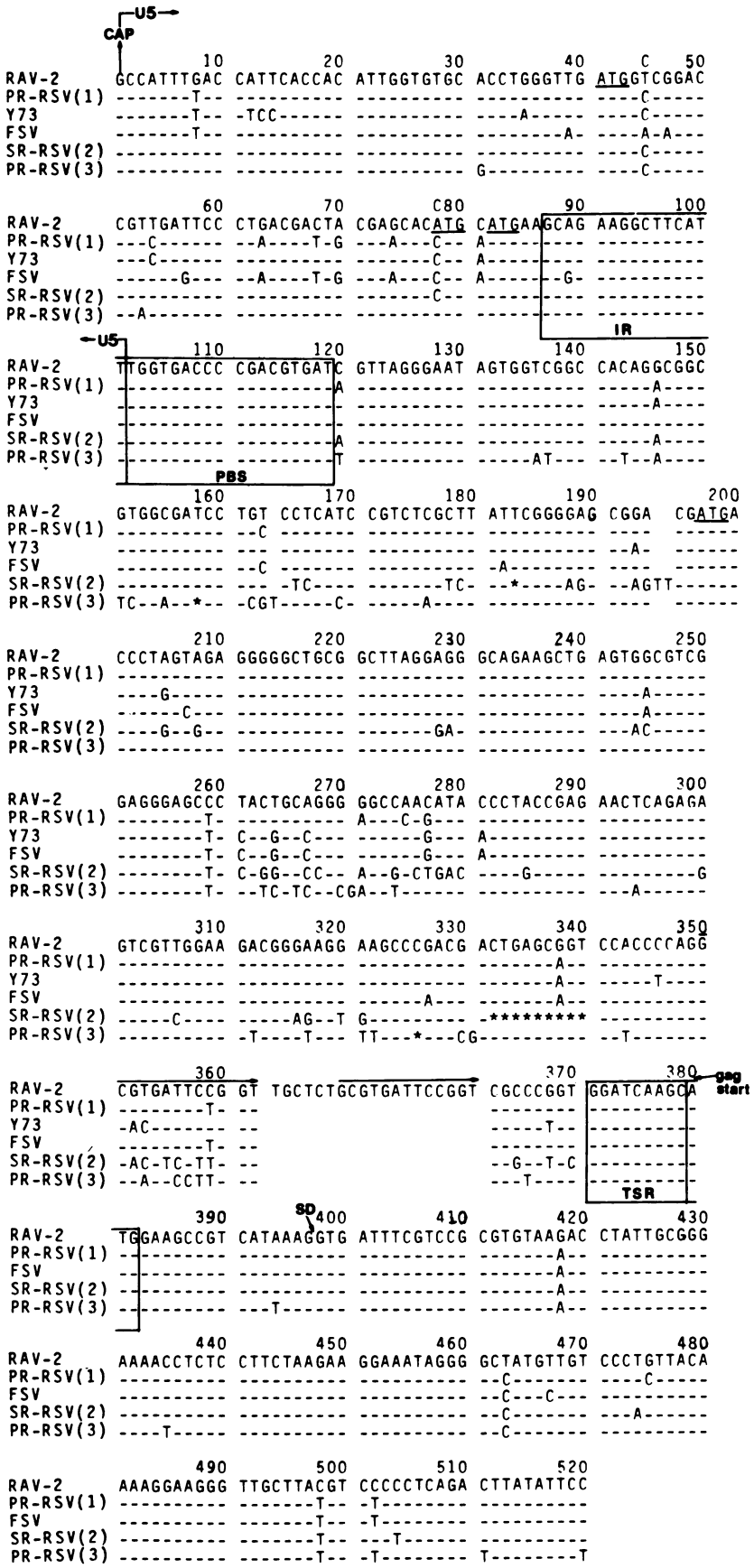


FIG. 8. Nucleotide sequence comparison of the RAV-2 leader with other avian retrovirus leader sequences. For this comparison, numbering corresponds to that established for PR-RSV (22); the initiating G residue at the cap site is denoted nucleotide 1. This corresponds to RAV-2 nucleotide 1022 in our sequence data (Fig. 1, 2, 6, and 7). The four nonfunctional ATG initiation codons within the leader region are underlined. IR is as described in the legend to Fig. 7. PBS, $tRNA^{trp}$ primer binding site; TSR, translation start region; arrows, a direct repeat; SD, splice donor site.

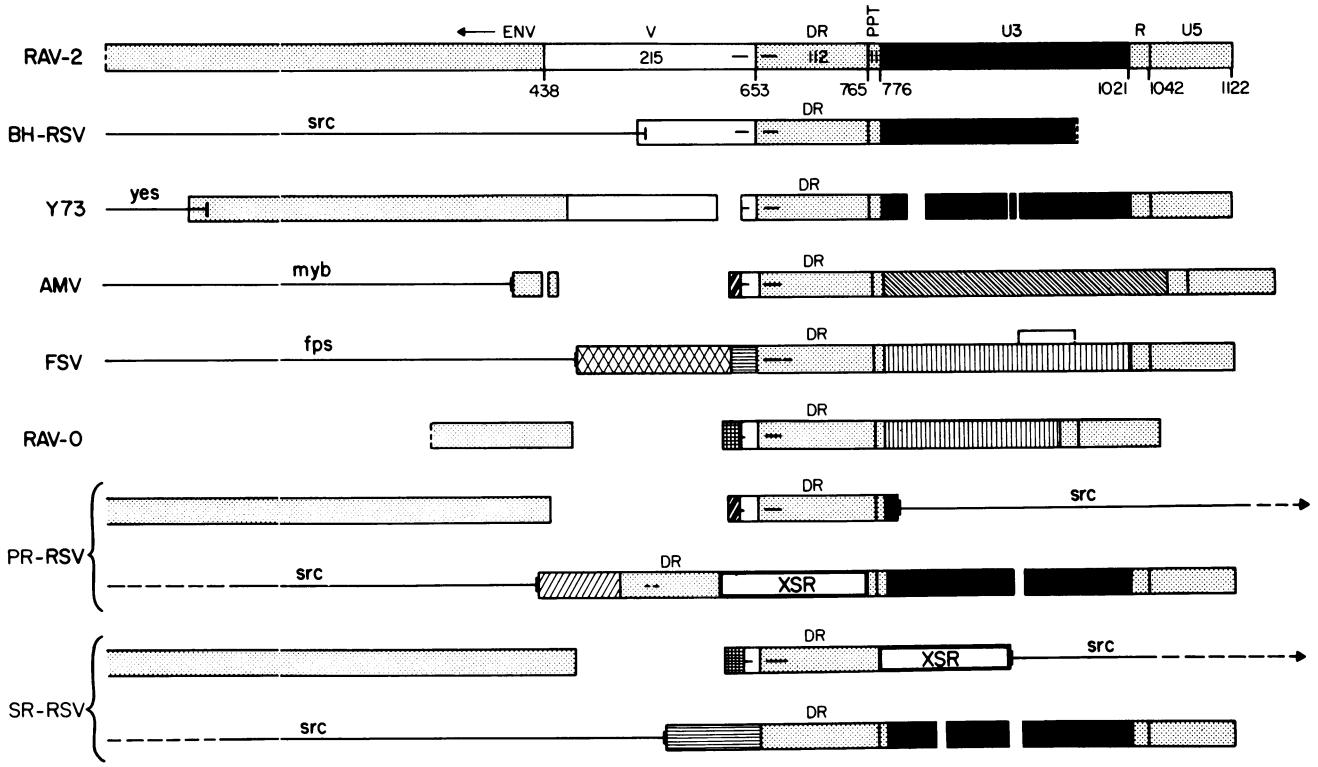


FIG. 9. Summary of relationships between terminal regions of several avian retroviral DNAs. The various regions are arranged in blocks within each section as defined in the text. Homologous regions within each section are identified by stipples. Nonhomologous sequences within the V and U_3 regions are distinguished by various patterns. The bracket over the U_3 region of the FSV sequence shows the approximate location of a partial direct repeat which accounts in large part for the size difference between it and the related U_3 region of RAV-0 (23). The XSR regions of the PR-RSV and SR-RSV strains are labeled. Open areas in the V and U_3 sections indicate deletions relative to RAV-2. Oncogene-containing segments are symbolized by straight lines, and they are also labeled. In BH-RSV the overlap in coding sequence of *src* with the V region is also shown. The PR-RSV and SR-RSV genomes have been arranged so that the analogous sections fall in the appropriate columns. Since sequences corresponding to the V and DR sections are repeated on either side of *src*, these are represented by overlapping lines from both directions; linear maps of PR-RSV and SR-RSV can be constructed by joining the *src* regions at the dashes. Adjacent dashes in the V and DR blocks show a short (12-base pair) internal direct repeat which is partially conserved in each genome as indicated. Vertical dashes at the right ends of the RAV-2 and RAV-0 maps and the left end of the BH-RSV map show the endpoints for which sequence data are available. Numbers within RAV-2 boxes show their lengths in nucleotides. Numbers below the RAV-2 map provide landmarks for comparison with the sequence data in Fig. 3, 5, 6, and 7.

DNA into the genome (IR). The prevalence of deletions, particularly in the stretch corresponding to RAV-2 nucleotides 901 to 917, suggests regions where sequence is not essential for function. There are additional single base pair changes scattered throughout the LTR whose significance we cannot yet assess.

Comparison of the 5′ leader sequences of RAV-2 and other retroviruses. Analysis of this region, which starts at the mRNA cap site and includes the U_5 section of the LTR, is shown in Fig. 8. Here again we observe an overall similarity, with several regions absolutely conserved. Those conserved include the IR region already noted and an adjacent 18-base pair sequence corresponding to the binding site for the tRNA primer (PBS) of minus strand viral DNA synthesis. A third region of conservation was observed immediately upstream from the ATG at position 380. Five potential ATG initiation codons are present within the leader region of RAV-2 at positions 41, 78, 82, 197, and 380. The ATG at position 380 corresponds to the *gag* gene initiation codon (22). Information from separate studies in our laboratory (unpublished) suggests that the adjacent conserved region (TSR) may be important for efficient mRNA translation. Many single base changes are scattered throughout the noncoding leader, some (e.g., positions 260 to 285) suggesting areas where sequence variation may not hinder function. One change in RAV-2, which appears to have been generated by a duplication of 13 nucleotides starting at position 350, has also been observed by Schwartz et al. in some genomes of PR-RSV virus (22). Some differences are also noted in the region downstream of the ATG initiation site which encodes the start of the *gag* gene. However, most of these are single base changes in the third codon position and thus do not affect coding.

Summary of comparisons. The diagram in Fig. 9 presents a summary of the relationships we have noted among the genomes analyzed. Here the regions have been arranged in blocks to emphasize their analogy. The one exception is the region from the end of *env* to the DR, which we denote as V. Its extreme variability and, in some cases, absence suggest that no essential function is encoded there.

When organized in this block fashion, the similarity in Rous-derived genomes is immediately apparent. An obvious difference is the XSR sequence; its absence in RAV-2, BH-RSV, and other oncogenic viral genomes indicates that it is not essential for oncogenicity. Another striking feature is the strong conservation of several sections in the noncoding region of all the viral genomes. These include the DR and adjacent PPT and the R, U_5, and adjacent leader regions (not shown). The functions encoded in the conserved regions appear to be compatible with at least three types of U_3

sections. These include the Rous/Y73 type, a RAV-0/FSV type, and a third typified by AMV. The interchangeability of the conserved sections and different U_3 regions is consistent with the notion that they represent separate functional units which are recognized by different viral or cellular protein complexes. For the U_3 region, known to contain important transcriptional control signals, interaction with cellular proteins involved in mRNA synthesis must be critical. The U_5 and leader regions contain translational signals, those required for packaging of viral genomic RNA and perhaps other viral functions. Thus, this region must interact with distinct sets of molecules, including those involved in the translational machinery of the hosts, as well as viral structural proteins encoded in the *gag* gene. Conservation at the ends of the LTR and adjacent PPT and primer binding site probably reflects selection imposed by requirement for reverse transcription and integration, reactions which involve interaction with yet another set of molecules, including the viral *pol* gene product and perhaps associated factors. There is, at present, no information concerning a possible function for the DR region, and thus it is not possible to identify the complexes with which it may interact. However, its strong conservation among the genomes shown suggests that its presence is essential and that its function may be revealed through site directed mutational analysis.

ACKNOWLEDGMENTS

We are grateful to T. Lerner and H. Hanafusa for providing unpublished data on BH-RSV and to C. Van Beveren for critical review of the manuscript.

LITERATURE CITED

1. **Coffin, J.** 1982. Structure of the retroviral genome, p. 308. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
2. **Czernilofsky, A. P., A. D. Levinson, H. E. Varmus, J. M. Bishop, E. Tischer, and H. Goodman.** 1983. Corrections to the nucleotide sequence of the *src* gene of Rous sarcoma virus. Nature (London) **301**:737–738.
3. **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.
4. **Darlix, J. L., M. Zuker, and P. F. Spahr.** 1982. Structure-function relationship of Rous sarcoma virus leader RNA. Nucleic Acids Res. **17**:5183–5196.
5. **Goldfarb, M. P., and R. A. Weinberg.** 1981. Generation of novel, biologically active Harvey sarcoma virus via apparent illegitimate recombination. J. Virol. **38**:136–150.
6. **Guo, L.-H. and R. Wu.** 1982. New rapid methods for DNA sequencing based on exonuclease III digestion followed by repair synthesis. Nucleic Acids Res. **10**:2065–2084.
7. **Hanafusa, H.** 1975. Avian RNA tumor viruses, p. 49–90. *In* F. F. Becker (ed.), Cancer, vol. 2. Plenum Publishing Corp., New York.
8. **Hayward, W. S., B. G. Neel, and S. M. Astrin.** 1981. Activation of a cellular *onc* gene by promoter insertion in ALV-induced lymphoid leukosis. Nature (London) **290**:475–481.
9. **Heidecker, G., J. Messing, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
10. **Hishinuma, F., P. J. DeBona, S. Astrin, and A. M. Skalka.** 1981. Nucleotide sequence of acceptor site and termini of integrated avian endogenous provirus *ev*-l: integration creates a 6 bp repeat of host DNA. Cell **23**:155–164.
11. **Hunter, E., E. Hill, M. Hardwick, A. Bhown, D. E. Schwartz, and R. Tizard.** 1983. Complete sequence of the Rous sarcoma virus *env* gene: identification of structural and functional regions of its product. J. Virol. **46**:920–936.
12. **Itohara, S., K. Hirata, M. Inoue, M. Hatsuoka, and A. Sato.** 1978. Isolation of a sarcoma virus from a spontaneous chicken tumor. Gann **69**:825–830.
13. **Ju, G., L. Boone, and A. M. Skalka.** 1980. Isolation and characterization of recombinant DNA clones of avian retroviruses: size heterogeneity and instability of the direct repeat. J. Virol. **33**:1026–1033.
14. **Ju, G., and A. M. Skalka.** 1980. Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. Cell **22**:379–386.
15. **Junghans, R. P., L. R. Boone, and A. M. Skalka.** 1982. Products of reverse transcription in avian retrovirus analyzed by electron microscopy. J. Virol. **43**:544–554.
16. **Kitamura, N., A. Kitamura, K. Toyoshima, Y. Hirayama, and M. Yoshida.** 1982. Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. Nature (London) **297**:205–208.
17. **Klempnauer, K.-H., T. J. Gonda, and J. M. Bishop.** 1982. Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene. Cell **31**:453–463.

17a. **Lerner, T. L., and H. Hanafusa.** 1984. DNA sequence of the Bryan high-titer strain of Rous sarcoma virus: extent of *env* deletion and possible geneological relationship with other viral strains. J. Virol. **49**:549–556.

18. **Maxam, A. M., and W. Gilbert.** 1977. A new method for DNA sequence analysis. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
19. **Neel, B. J., and W. S. Hayward.** 1981. Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. Cell **23**:323–334.
20. **Payne, G. S., S. A. Courtneidge, L. B. Crittenden, A. M. Fadly, J. M. Bishop, and H. E. Varmus.** 1981. Analyses of avian leukosis virus DNA and RNA in bursal tumors: viral gene expression is not required for maintenance of tumor state. Cell **23**:311–322.
21. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
22. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–863.
23. **Shibuya, M., and H. Hanafusa.** 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other sarcoma viruses. Cell **30**:787–795.
24. **Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus.** 1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. U.S.A. **78**:124–128.
25. **Swanstrom, R., R. C. Parker, H. E. Varmus, and J. M. Bishop.** 1983. Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **80**:2519–2523.
26. **Swanstrom, R., H. E. Varmus, and J. M. Bishop.** 1982. Nucleotide sequence of the 5′ noncoding region and part of the *gag* gene of Rous sarcoma virus. J. Virol. **41**:535–541.
27. **Takeya, T., R. A. Feldman, and H. Hanafusa.** 1982. DNA sequence of the viral and cellular *src* gene of chickens. 1. Complete nucleotide sequence of an *Eco*RI fragment of recovered avian sarcoma virus which codes for gp37 and pp60^{src}. J. Virol. **44**:1–11.
28. **Takeya, T., and H. Hanafusa.** 1983. Structure and sequence of the cellular gene homologous to the mechanism for generating the transforming virus. Cell **32**:881–890.
29. **Tsichlis, R. N., L. Donehower, G. Hager, N. Zeller, R. Malavarca, S. Astrin, and A. M. Skalka.** 1982. Sequence comparison in the crossover region of an oncogenic avian retrovirus recombinant and its nononcogenic parent: genetic regions that control growth rate and oncogenic potential. Mol. Cell. Biol. **2**:1331–1338.
30. **Van Beveren, C., F. van Straaten, T. Curran, R. Muller, and I. M. Verma.** 1983. Analysis of FBJ-MuSV provirus and c-*fos* (mouse) gene reveals that viral and cellular *fos* gene products have different carboxy termini. Cell **32**:1241–1255.