JOURNAL OF VIROLOGY, Oct. 1984, p. 172–182
0022-538X/84/100172-11$02.00/0
Copyright © 1984, American Society for Microbiology
Vol. 52, No. 1

Nucleic Acid Sequences of the Oncogene v-*rel* in Reticuloendotheliosis Virus Strain T and Its Cellular Homolog, the Proto-Oncogene c-*rel*

KIRK C. WILHELMSEN, KATIE EGGLETON, AND HOWARD M. TEMIN*
McArdle Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706

Received 16 April 1984/Accepted 20 June 1984

Reticuloendotheliosis virus strain T (Rev-T) is a highly oncogenic replication-defective retrovirus which contains the oncogene v-*rel*. It is thought that Rev-T arose when a virus similar to Rev-A, the helper virus of Rev-T, infected a turkey and recombined with c-*rel* from that turkey. There is one large c-*rel* locus in the turkey genome which contains all of the sequences homologous to v-*rel* (K. C. Wilhelmsen and H. M. Temin, J. Virol. 49:521–529, 1984). We have sequenced v-*rel* and its flanking sequences, each of the regions of the c-*rel* locus from turkey that are homologous to v-*rel* and their flanking sequences, and the coding sequence for *env* and part of *pol* of Rev-A. The v-*rel* coding sequences can be translated into a 503-amino acid *env*-v-*rel*-out-of-frame-*env* fusion polypeptide. We have not detected any sequences in the Los Alamos or University of California-San Diego data bases that are more significantly related to the amino acid or nucleic acid sequence of v-*rel* than to the randomized sequence of v-*rel*. Comparison of Rev-A, Rev-T, and c-*rel* indicates that the v-*rel* sequences may have been transduced from the c-*rel* (turkey) locus by a novel mechanism. There are sequences in Rev-A and c-*rel* that are similar to splicing signals, indicating that the 5′ virus-*rel* junction of Rev-T may have been formed by cellular RNA splicing machinery. Eight presumed introns have presumably been spliced out of c-*rel* to generate v-*rel*. There are also short imperfect regions of homology between sequences at the boundaries of v-*rel* and sequences in Rev-A and c-*rel* (turkey), indicating that c-*rel* may have been transduced by homologous recombination. There are many differences between the amino acid sequences of the predicted translational products of v-*rel* and c-*rel* which may account for their difference in transformation potential. These sequence differences between v-*rel* and c-*rel* include 10 missense transitions, four missense transversions, and three places where Rev-T has a small in-frame deletion of sequences relative to c-*rel*. Most of the coding sequence differences between c-*rel* and v-*rel* are nonconservative amino acid changes.

Highly oncogenic retroviruses are retrovirus vectors that contain inserted sequences called oncogenes. Oncogenes are not necessary for virus replication, but they cause the transformation of infected target cells. With one exception (Rous sarcoma virus [RSV]), all highly oncogenic retroviruses are replication defective because their oncogenes are substituted for sequences that encode *trans*-acting viral replication functions.

Oncogenes are derived from normal cellular DNA sequences called proto-oncogenes. The mechanism by which these cellular DNA sequences are incorporated into retroviruses is unclear. It has been postulated that the incorporation of proto-oncogene sequences into a retrovirus to form a highly oncogenic retrovirus involves homologous or nonhomologous recombination (10, 16, 17, 21, 25, 27, 29, 31, 34).

Highly oncogenic retroviruses cause transformation of infected target cells shortly after infection as a result of expression of their oncogenes. In contrast, the normal expression of the related proto-oncogene sequences in situ in normal cells does not cause transformation. Therefore, there must be some qualitative or quantitative differences or both between the expression of oncogenes and proto-oncogenes. In one case, it was shown that there is a qualitative difference between the gene product of an oncogene and that of the related proto-oncogene that can account for their difference in transformation potential for NIH-3T3 cells (26). In another case, the difference in transformation potential for a rat cell line between an oncogene and the related proto-oncogene appears to be the result of a quantitative difference in the level of their expression (13, 30).

We have been studying the highly oncogenic retrovirus reticuloendotheliosis virus strain T (Rev-T), which contains the oncogene v-*rel*. Rev-T was first isolated from a turkey (28). It is thought that Rev-T arose when a virus similar to Rev-A infected a turkey and recombined with c-*rel* from that turkey. Analysis of molecular clones of proviruses of Rev-T and Rev-A (Rev-A is the nondefective helper virus of Rev-T) shows that in Rev-T, v-*rel* is substituted for most of *env* in Rev-A (see Fig. 1) (3, 8, 18). Rev-T also has a large deletion of sequences which encode much of *gag* and *pol* in Rev-A. This deletion is necessary for transformation of cells by v-*rel* (4). In Rev-T-infected cells, there are two viral RNA transcripts that contain v-*rel* sequences: a full-length genomic size transcript and a subgenomic RNA transcript which is similar in size to the subgenomic *env* mRNA in Rev-A (32; K. C. Wilhelmsen, unpublished data). The subgenomic v-*rel*-containing transcript in Rev-T-infected cells is thought to be translated to make the v-*rel* gene product. The positions of the subgenomic env mRNA splicing sequences in Rev-A are shown in Fig. 1 (32; Wilhelmsen, unpublished data).

To determine the nature of the v-*rel* gene product and to determine whether it is expressed as a fusion product with viral genes, we sequenced v-*rel* and its flanking sequences in Rev-T as well as the 3′ half of Rev-A. We wanted to determine (i) whether the acquisition of the *rel* sequences was by the recombination mechanism postulated for other oncogenes, (ii) whether all of the v-*rel*-specific sequences in Rev-T came from the previously identified c-*rel* turkey locus

* Corresponding author.

(37), and (iii) whether sequence differences exist that may explain the difference in transformation potential of v-*rel* in Rev-T and c-*rel* in situ. To answer these questions, we determined the nucleotide sequence of each of the regions of c-*rel* that are homologous to v-*rel*.

MATERIALS AND METHODS

Sources of DNA clones. Rev-A, Rev-T, and c-*rel* (turkey) clones have been described previously (3–5, 36, 37). Many subclones were made of c-*rel* sequences for this study. Details of the c-*rel* subclones can be obtained from us.

Nucleic acid sequence analysis. DNA sequencing was done by the method of Maxam and Gilbert (12). Data analysis was done by using the University of Wisconsin Genetics Computer Users Group programs (6) and the programs of Staden (23).

RESULTS

To predict the size, amino acid (AA) sequence, and transcriptional program of v-*rel* and the relationship of v-*rel* to *env*, we sequenced from 4.60 to 7.74 kilobase pairs (kbp) of Rev-A and from 2.91 to 4.83 kbp of Rev-T (see Fig. 2). To determine more precisely the relationship of v-*rel* to c-*rel*, we also sequenced the regions of c-*rel* that are homologous to v-*rel*. The sequencing strategies for Rev-A, Rev-T, and c-*rel* are shown in Fig. 2.

Rev-A. Figure 3 shows the sequence of part of Rev-A and the predicted protein translation product for two large overlapping open reading frames. No other open reading frames longer than 100 nucleotides were detected that also had a translational initiation codon. The reading frame that begins at 4,600 base pairs (bp) is presumed to be the 3′ coding sequence of the *pol* gene. The translational termination signal for *pol* is at 6,019 bp, which is 3′ to the predicted translational initiation codon for the *env* gene at 5,958 bp. The transcriptional termination signal for the *env* gene is at 7,704 bp. The position of the *env* RNA splice acceptor site has previously been mapped to 5,702 bp (Fig. 1). There are two sequences which are similar to the consensus sequence for RNA splice acceptors very close to 5,702 bp: one at 5,698 bp and another at 5,707 bp.

The long terminal repeats (LTRs) in spleen necrosis virus, another reticuloendotheliosis virus, have been sequenced previously and are bounded by 5-bp inverted repeats with the sequence AATGT (19). The 5′ end of the 3′ LTR of Rev-A begins at 7,731 bp. A polypurine tract, which is needed to prime plus-strand viral DNA synthesis, is found 5′ to the LTR between 7,718 and 7,730 bp.

Previous enzymological and immunological studies have indicated that reticuloendotheliosis viruses are more closely related to the C type murine retroviruses than to avian leukosis-sarcoma viruses (reviewed in reference 35). Comparison of the predicted amino acid sequences of the *pol* and *env* genes of Prague C strain of RSV, Moloney murine leukemia virus (MLV), and Rev-A support these studies (Fig. 4; other data not shown). The *pol* and *env* genes of MLV and Rev-A are much more closely related than are the genes of MLV and RSV or Rev-A and RSV. The *pol* gene products of Rev-A and MLV are more closely related than their *env* gene pruducts. The greatest homology of the *env* genes of MLV and Rev-A is at the 3′ end of the predicted polyproteins. The *env* polyproteins of MLV and Rev-A are cleaved into a leader peptide, a large peptide (gp70 and gp73, respectively), and a smaller peptide (p15 and gp22, respectively) (reviewed in reference 35). The sequences of the smaller peptides, p15 and gp22, are coded by the 3′ end of the *env* genes and are more closely related than are the larger gp70 and gp73 peptides.

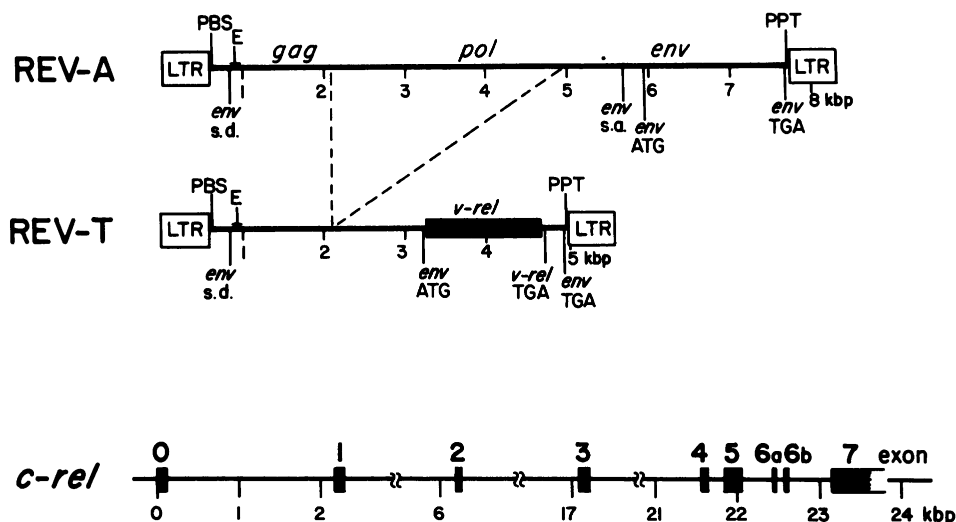


FIG. 1. Structures of Rev-A, Rev-T, and c-*rel* (turkey). The positions of the *cis*-acting sequences in Rev-T and Rev-A that are required for virus replication are labeled as follows: LTR for long terminal repeat; PBS for primer binding site; E for encapsidation sequences; and PPT for polypurine tract (32). Also indicated are the relative positions of the *gag*, *pol*, and *env* coding sequences, and the v-*rel* substitution. The dotted lines indicate the position of the large deletion in Rev-T relative to Rev-A. The positions of the splicing control sequences for *env* and v-*rel* are indicated by *env* s.d. for the env transcript splice donor and *env* s.a. for the env transcript splice acceptor (32; Wilhelmsen, unpublished data). The positions of the translational initiation and termination codons for *env* and the *rel*-*env* fusion product are shown (see Fig. 3 and 5 for details). The black bars on the c-*rel* map indicate regions of c-*rel* that are homologous to v-*rel*. Sequences similar to consensus RNA splice acceptor and donors were found at the 5′ and 3′ boundaries, respectively, of each of the regions of c-*rel* that are homologous to v-*rel*, except at the 3′ boundary of region 7 (see Table 2 for details). The maps of Rev-A and Rev-T have marks at 1-kbp intervals. The map of c-*rel* is discontinuous, but the correct coordinates are indicated with marks at 1-kbp intervals for each segment.

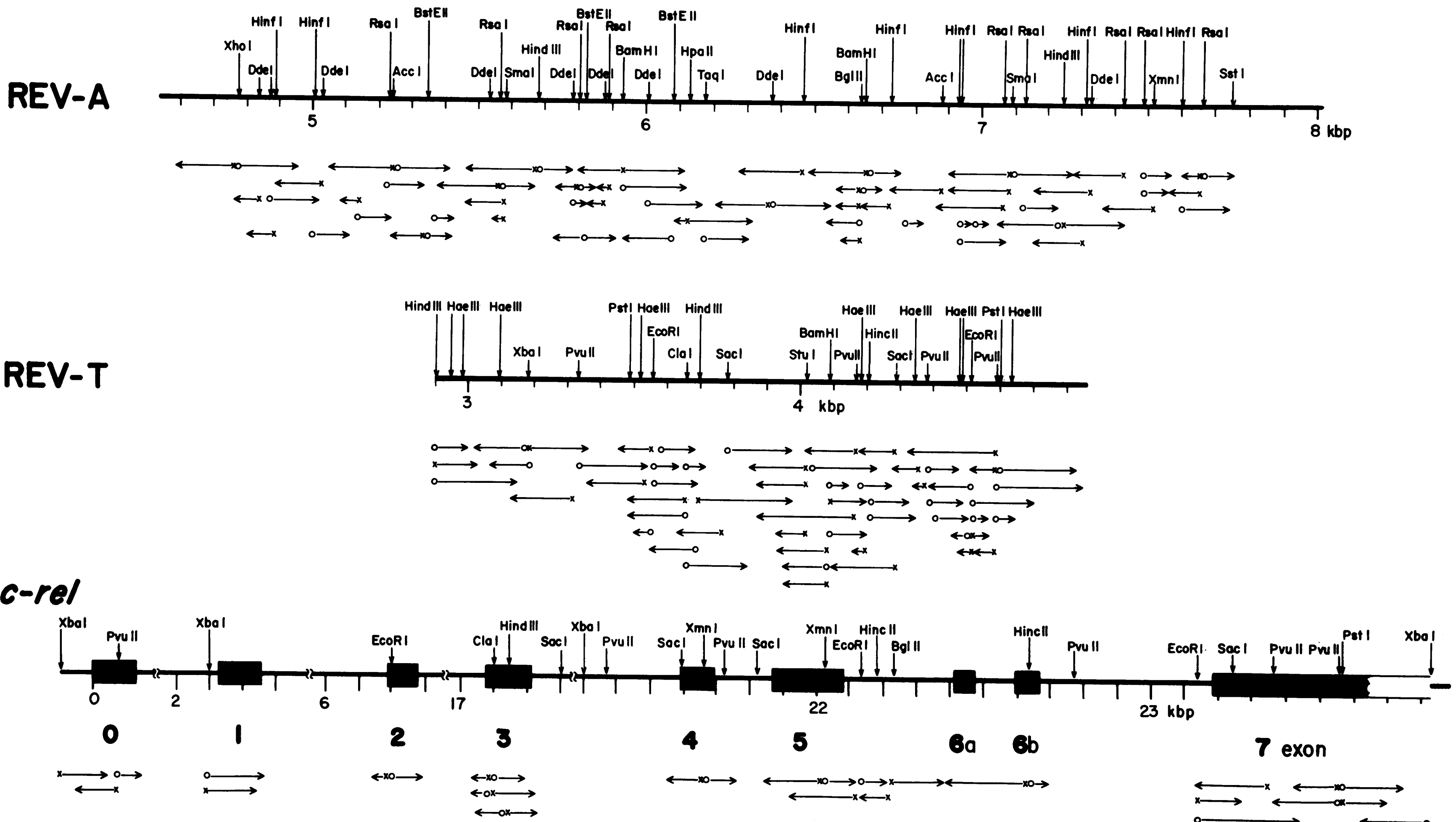


FIG. 2. Sequencing strategies for Rev-A, Rev-T, and c-*rel*. Restriction enzyme cleavage site maps for Rev-A, Rev-T, and c-*rel* are shown. The coordinates of each map are the same as in Fig. 1, with marks at 100-bp intervals. Only the restriction enzyme cleavage sites relevant to the sequencing are shown in the map of Rev-A. The map of c-*rel* is discontinuous, but appropriate coordiates with marks at 100-bp intervals are indicated. The bars in the map of c-*rel* indicate the size and position of the putative exons of c-*rel*. The open reading frame of c-*rel* extends beyond the *Xba*I site that is 3′ to the 3′-most region of c-*rel* that is homologous to v-*rel*. The horizontal arrows below each map indicate the fragments sequenced. Arrows have a ○ or an × to indicate that either the + or − strand, respectively, was sequenced.

```
4600  AAAGGGGAACTGTTGACGGCGGGGGAAAAGGCTATCAAAAACGCCCCCGAAATTCTAGCCTTGCTGACCGCGGTCTGGCTACCTAAACGGGTAGCTGTGATGCATTGCAAGGGACACCAG
      LysGlyGluLeuLeuThrAlaGlyGluLysAlaIleLysAsnAlaProGluIleLeuAlaLeuLeuThrAlaValTrpLeuProLysArgValAlaValMetHisCysLysGlyHisGln
      pol

4720  AAAGACGATGCACCCACATCAACTGGTAACCGACGGGCAGATGAGGTGGCTCGAGAAGTAGCTATACGTCCTTTGAGCACCCAAGCCACCATCTCCGATGCACCGGATATGCCAGATACC
      LysAspAspAlaProThrSerThrGlyAsnArgArgAlaAspGluValAlaArgGluValAlaIleArgProLeuSerThrGlnAlaThrIleSerAspAlaProAspMetProAspThr

4840  GAAACCCCTCAGTATTCGCATGTGGAGGAGGCTCTAGGACATCGACTGAGGGGGACCAAAGACTCTGCTGGGTGGTGGCACCTTCCTGACGGGCGCCTCTTGCTTCCCAGAGCAGTGGGG
      GluThrProGlnTyrSerHisValGluGluAlaLeuGlyHisArgLeuArgGlyThrLysAspSerAlaGlyTrpTrpHisLeuProAspGlyArgLeuLeuLeuProArgAlaValGly

4960  AGAAAGGTCCTAGAGCAAACTCACAGAGCCACACATTTGGGTGAATCAAAATTGACTGAGTTAGTTAGAAAACACTATCCTATCTGTGGGATTTATCGAGCAGCTAGAGATATCACCACC
      ArgLysValLeuGluGlnThrHisArgAlaThrHisLeuGlyGluSerLysLeuThrGluLeuValArgLysHisTyrProIleCysGlyIleTyrArgAlaAlaArgAspIleThrThr

5080  CGCTGTGTGGCATGTGCCCAGGTAAACCCTAGAGCAGCTCCCGTGGAGAAAGGGCTCAATTCCCGCATCAGAGGTGCGGCGCCAGGGGAACACTGGGAGGTGGATTTCACAGAAATGATA
      ArgCysValAlaCysAlaGlnValAsnProArgAlaAlaProValGluLysGlyLeuAsnSerArgIleArgGlyAlaAlaProGlyGluHisTrpGluValAspPheThrGluMetIle

5200  ACAGCCAAAGGGGGGTATAAATACCTGCTTGTACTGGTAGACACATTCTCGGGCTGGGTAGAGGCATATCCAGCAAAAAGAGAAACCTCCCAAGTGGTGATTAAGCATCTAATTCTTGAT
      ThrAlaLysGlyGlyTyrLysTyrLeuLeuValLeuValAspThrPheSerGlyTrpValGluAlaTyrProAlaLysArgGluThrSerGlnValValIleLysHisLeuIleLeuAsp

5320  ATTATCCCCAGGTTTGGGTTACCGGTCCAGATCGGGTCCGACAATGGGCCGGCCTTTGTGGCAAAAGTGACACAGCAGTTGTGTGAGGCCCTAAATGTCTCCTGGAAGCTCCATTGTGCA
      IleIleProArgPheGlyLeuProValGlnIleGlySerAspAsnGlyProAlaPheValAlaLysValThrGlnGlnLeuCysGluAlaLeuAsnValSerTrpLysLeuHisCysAla

5440  TACCGACCTCAAAGCTCGGGACAGGTGGAAAGGATGAACCGAACCCTAAAGAAAGCCATCGCTAAATTGGAGGATAGAGACAGGAGGGGACTGGGTCTCCCTCCTCCCTCAGGCTTTGCT
      TyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrLeuLysLysAlaIleAlaLysLeuGluAspArgAspArgArgGlyLeuGlyLeuProProProSerGlyPheAla

5560  CCGGGCACGGTGTACCCCGGGAGGGAGGGATTGTCACCCTTCGAGATATTATATGGTCTCAAGCCTCCTGTAGTTCCCCGTGTAGGATGTGACAAGCTTGCCAGCATAACCAACCAAACC
      ProGlyThrValTyrProGlyArgGluGlyLeuSerProPheGluIleLeuTyrGlyLeuLysProProValValProArgValGlyCysAspLysLeuAlaSerIleThrAsnGlnThr

5680  TTGCTTAAGTCCCTACAGGCCCTCCAGGCTACTAGGTCTCTGGCTCGGGCAGCTGCGCGACCAACTGCCCCAGAAAGAAGCTCAGCAAGACCGTACCCCACTGTCCCCAACCTGGTGACC
      LeuLeuLysSerLeuGlnAlaLeuGlnAlaThrArgSerLeuAlaArgAlaAlaAlaArgProThrAlaProGluArgSerSerAlaArgProTyrProThrValProAsnLeuValThr

5800  TCGTTCTTCGTTAAGAAGCACGACTTCCAGCAGTTGGGGCCACGGTGGGACGGACCCTACACTGTAGTCCTAAGTACCCCCACCGCGGTAAAGGCCGCTGGGAAGACCCCGTGGATCCAC
      SerPhePheValLysLysHisAspPheGlnGlnLeuGlyProArgTrpAspGlyProTyrThrValValLeuSerThrProThrAlaValLysAlaAlaGlyLysThrProTrpIleHis

5920  TACTCTCGACTCAAGAAAGCTCCTGACAACCAAGAAGAATGGACTGTCTCACCGACCTCCGATCCACTGAGGGTAAAGTTGACCAGGCGGGCAAAACCCTAATTCTTCTTGTGGTTTGGTGG
      TyrSerArgLeuLysLysAlaProAspAsnGlnGluGluTrpThrValSerProThrSerAspProLeuArgValLysLeuThrArgArgAlaLysProEnd pol TERMINATION
                           env INITIATION  MetAspCysLeuThrAspLeuArgSerThrGluGlyLysValAspGlnAlaGlyLysThrLeuIleLeuLeuValValTrpTrp

6042  GGGTTTGGGACCACTGCCGAGGGTCACCCCTTGCAGCAACTTTGGGAACTGCCTTGTGACTGCTCCGGGGGATATGTCTCCCCCGACCTACCTATTACACCAACTCCCTCGATTGCGGTA
      GlyPheGlyThrThrAlaGluGlyHisProLeuGlnGlnLeuTrpGluLeuProCysAspCysSerGlyGlyTyrValSerProAspLeuProIleThrProThrProSerIleAlaVal
      env

6162  GCTTCACCGCTACCTGACTTACGGGTCTGGTTGCAGGGTAGTTGGGGCTGGGGAGGGGGATTTAGACAACAGTGGGAGTGTGTGTTTAAACCCAAGATCATACCCTCTGTGCAGGAGCAG
      AlaSerProLeuProAspLeuArgValTrpLeuGlnGlySerTrpGlyTrpGlyGlyGlyPheArgGlnGlnTrpGluCysValPheLysProLysIleIleProSerValGlnGluGln

6282  CCAGGGCCCTGCGAATGCCTCACGATAGCCACCCAAATGCATTCCACCTGTTACGAAAAGGCTCAGGAATGCACCCTCCTGGGAAAAACTTATTTTACTGCCATCCTACAGAAAACAAAG
      ProGlyProCysGluCysLeuThrIleAlaThrGlnMetHisSerThrCysTyrGluLysAlaGlnGluCysThrLeuLeuGlyLysThrTyrPheThrAlaIleLeuGlnLysThrLys

6402  CTAGGTTCGTATGAAGACGGGCCTAATAAACTGCTCCAAGCCTCTTGCACGGGAATCTGGGAAACCAGTATGTTGGGACCCCGCTGCCCCTGTGTATGTCTCGATGGGGGCGGTCCCACT
      LeuGlySerTyrGluAspGlyProAsnLysLeuLeuGlnAlaSerCysThrGlyIleTrpGluThrSerMetLeuGlyProArgCysProCysValCysLeuAspGlyGlyGlyProThr

6522  GACAGATTCGGAAGAATCTGTGCGGAAGGACTGGAGGAAATCATCAGGCACAGCTACCCCTCCGTACAGTATCACCCTTTAGCCCTGCCCCGACCAAGAGGAGTAGATCTGGATCCCCAG
      AspArgPheGlyArgIleCysAlaGluGlyLeuGluGluIleIleArgHisSerTyrProSerValGlnTyrHisProLeuAlaLeuProArgProArgGlyValAspLeuAspProGln

6642  ACGTCTGACATACTGGAAGCTACTCACCAGGTCCTTAACGCCACTAATCCTCAGCTAGCAGAGAACTGCTGGCTTTGCATGACTCTTGGAACTCAATCCCCGCAGCCATCCCGGCGAATG
      ThrSerAspIleLeuGluAlaThrHisGlnValLeuAsnAlaThrAsnProGlnLeuAlaGluAsnCysTrpLeuCysMetThrLeuGlyThrGlnSerProGlnProSerArgArgMet

6762  GCAATGTCACTCTCGATGGAAATTGCAGTCTTAGCCTCCCTTTCGGGTGCAACCCACCGGGTCAATAGATGTCAACTGCTATGCAGGGAAGCAGACAATAGGACTGGTATACCCGTAGGG
      AlaMetSerLeuSerMetGluIleAlaValLeuAlaSerLeuSerGlyAlaThrHisArgValAsnArgCysGlnLeuLeuCysArgGluAlaAspAsnArgThrGlyIleProValGly

6882  TATGTTCATTTCACTAACTGCACTAGCATCCAAGAGTCTCTAACGAGACGAGTCATATATGAAATCTTACGAGACTATGTCCTCCACCGGGTCATGTATTTGTGTGTGGAACAACATGCC
      TyrValHisPheThrAsnCysThrSerIleGlnGluSerLeuThrArgArgValIleTyrGluIleLeuArgAspTyrValLeuHisArgValMetTyrLeuCysValGluGlnHisAla

7002  TACACGGCGCTCCCTAATAAATGGATAGGGTTGTGCATACTGGCATCAATCGTACCCGACATGAGCATAATACCCGGGGAAGAGCCTATCCCACTCCCATCCATCGAGTACACCGCTGGG
      TyrThrAlaLeuProAsnLysTrpIleGlyLeuCysIleLeuAlaSerIleValProAspMetSerIleIleProGlyGluGluProIleProLeuProSerIleGluTyrThrAlaGly

7122  CGTCATAAGAGGGCAGTCCAGTTTATCCCCCTGCTTGTAGGTCTAGGGATTACAGGGGCTACACTTGCTGGTGGAACTGGACTTGGGGTTTCCGTTCACACTTATCACAAGCTTTCTAAT
      ArgHisLysArgAlaValGlnPheIleProLeuLeuValGlyLeuGlyIleThrGlyAlaThrLeuAlaGlyGlyThrGlyLeuGlyValSerValHisThrTyrHisLysLeuSerAsn

7242  CAATTGATTGAAGATGTCCAGGCTCTTTCAGGGACCATCAATGACCTACAGGACCAGATTGACTCCCTAGCTGAGGTTGTCTTACAAAATAGAAGGGGGTTAGACCTATTGACTGCCGAA
      GlnLeuIleGluAspValGlnAlaLeuSerGlyThrIleAsnAspLeuGlnAspGlnIleAspSerLeuAlaGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuThrAlaGlu

7362  CAAGGAGGAATATGTCTCGCACTCCAGGAGAAGTGCTGTTTTTACGCCAACAAGTCGGGTATAGTACGTGACAAGATCCGGAAACTCCAGGAGGACCTTCTCGCGAGGAAACGTGCACTG
      GlnGlyGlyIleCysLeuAlaLeuGlnGluLysCysCysPheTyrAlaAsnLysSerGlyIleValArgAspLysIleArgLysLeuGlnGluAspLeuLeuAlaArgLysArgAlaLeu

7482  TACGACAACCCCCTGTGGAACGGCTTGAACGGCTTCCTTCCATATTTGCTACCCTCGTTGGGCCCCCTGTTTGGGCTCATATTGTTCCTGACCCTCGGCCCGTGCATCAGGAAGACCCTG
      TyrAspAsnProLeuTrpAsnGlyLeuAsnGlyPheLeuProTyrLeuLeuProSerLeuGlyProLeuPheGlyLeuIleLeuPheLeuThrLeuGlyProCysIleArgLysThrLeu

7602  ACTCGCATTATCCATGACAAAATTCAGGGCAGTAAAAATCCTCGCATTAGTCCCGCAGTACAAGCCACTCCCAACAGAGATGGATACCCTAGGTCAATGGTTTGACCAGAACATACAAGA
      ThrArgIleIleHisAspLysIleGlnGlySerLysAsnProArgIleSerProAlaValGlnAlaThrProAsnArgAspGlyTyrProArgSerMetValEnd  env TERMINATION

7722  GCAGTGGGGAATGTGGGAGGGAGCTCT
               LTR
```

FIG. 3. Sequence of the 3′ coding sequence of Rev-A. The DNA sequence of the 3′ half of the *pol* gene, the *env* gene, and the 3′ nontranslated sequences before the LTR of Rev-A are shown. The predicted AA sequences of part of *pol* and all of *env* are shown below the DNA sequence. The sequence has been given the initial starting coordinate of 4600, which corresponds to the coordinate system in Fig. 1 and 2.

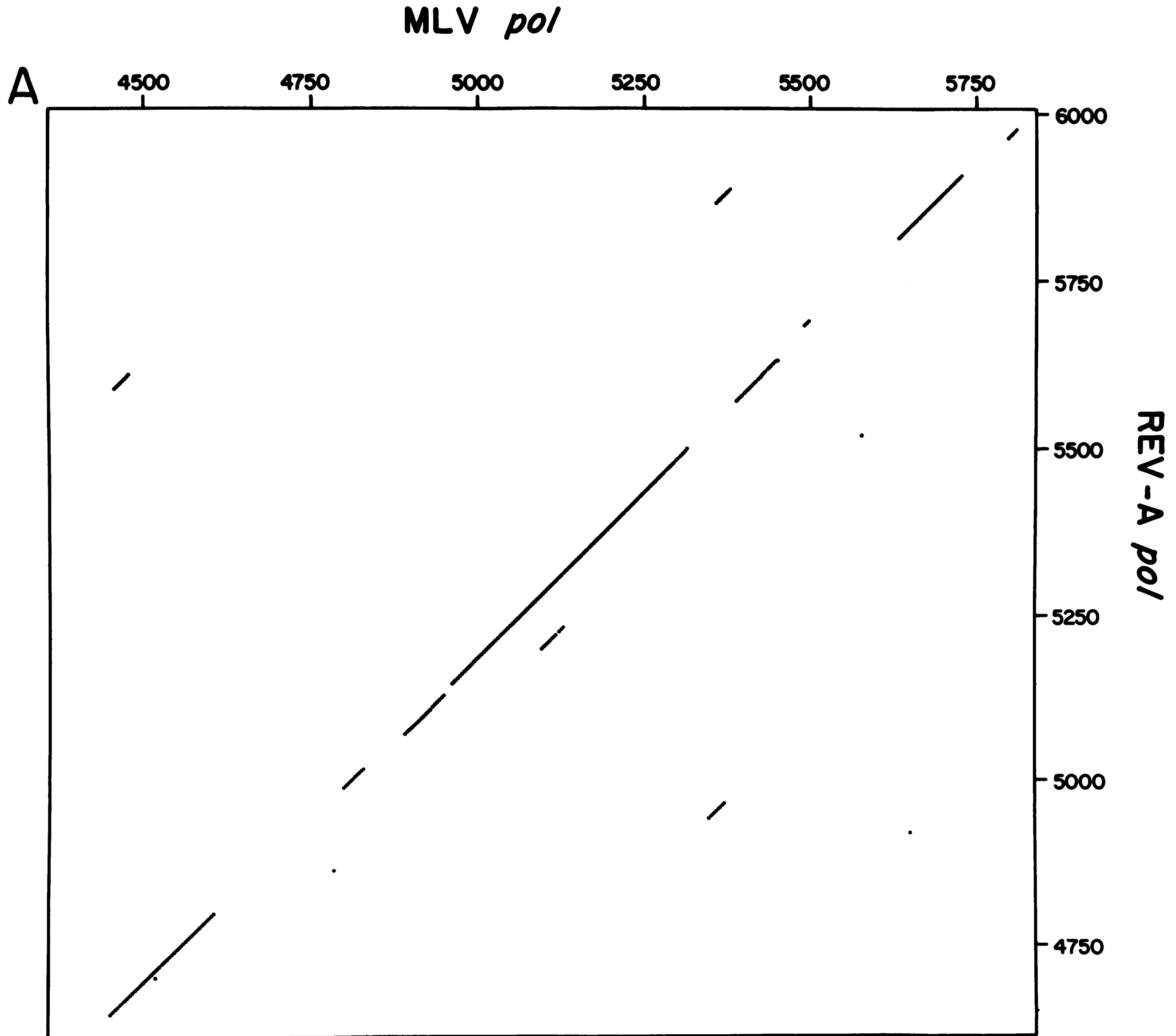
Codons of amino acids that are potential glycosylation sites (AsnXxxThr or AsnXxxSer) in the *pol* gene were found at 5,419, 5,482, and 5,677 bp. Codons of amino acids that are potential glycosylation sites in the *env* gene were found at 6,864, 6,903, and 7,416 bp.

Rev-T. Figure 5 shows the sequence of Rev-T between 2,898 and 4,832 bp. Comparison of the Rev-T sequence with the sequence of the *env* region of Rev-A shows that the v-*rel* sequences in Rev-T are between approximately 3,255 and 4,675 bp, confirming previous reports (3, 8, 18).

Both the 5′ and 3′ sequences that flank v-*rel* have multiple transitions, transversions, and frame shift mutations relative to Rev-A. The *env* RNA splice acceptor(s), if they functioned as they do in Rev-A, would be at 2,953 or 2,962 bp or both in Rev-T. One of these projected splice acceptors, at 2,962 bp, contains a G to A transition and now has a sequence inconsistent with the consensus sequence for RNA splice acceptors.

One reading frame is open for the entire length of v-*rel*. It has the putative env transcriptional initiation codon in phase upstream approximately 35 bp (at 3,221 bp) and a translational termination codon downstream of v-*rel* approximately 55 bp in the *env* coding sequences but in a different reading frame from the putative *env* gene. No other translational initiation signals are found 3′ to the major subgenomic RNA splice acceptor and 5′ to the *env* transcriptional initiation signal. The *env*-v-*rel*-out-of-frame-*env* fusion polypeptide is predicted to be 503 AAs long. Two other reading frames were also found which could be translated to generate peptides longer than 50 AAs. One is from upstream of 2,898 to 3,126 bp and begins in the same reading frame as the putative *pol* gene but ends in a different reading frame than *pol* due to the multiple frame shift mutations relative to Rev-A between 2,985 and 3,070 bp in Rev-T, and another is a small open reading frame within the v-*rel* sequences which has a translational initiation codon at 4,312 bp and a translational termination codon 56 AAs later at 4,483 bp.

The 503-AA rel fusion polypeptide appears to be the



transforming product of Rev-T, since mutations which interfere with translation initiated at the *env* start signal or are in the coding sequences of the rel fusion polypeptide inhibit transformation by Rev-T (3, 4). The *rel* gene product is unusually rich in serine (9.1%) and proline (8.7%) compared with other proteins and oncogenes (see reference 24). Charged residues account for 23% of the total AAs, and only 15% of the charged residues are in the carboxy 30% of the predicted rel fusion polypeptide. Hydrophobic residues are evenly distributed throughout the predicted polyprotein. The codons of AAs that are potential glycosylation sites (AsnXxxThr or AsnXxxSer) are at 3,380, 3,734, and 4,535 bp in Rev-T.

Large nucleic acid and peptide data bases have been searched to determine whether there are nucleic acid or peptide sequences that are related to the v-*rel* coding sequences (personal communications from C. Burks at Los Alamos National Laboratory, Los Alamos, N.M., and R. Doolittle at the University of California at San Diego). No sequences were found that were more significantly related to v-*rel* than to the best alignment of independent randomizations of the v-*rel* sequence.

c-*rel*. All of the regions of c-*rel* that are homologous to v-*rel* have been sequenced by using the sequencing strategy shown in Fig. 2. The sequence of the regions of c-*rel* that are homologous to v-*rel* has been aligned next to the sequence of Rev-T in Fig. 5. All of the v-*rel*–specific sequences in Rev-T have homologous sequences in c-*rel*, indicating that the c-*rel* locus previously identified in turkey DNA is the source of all of the v-*rel*–specific sequences in Rev-T. Table 1 is a summary of the differences between the coding sequences of v-*rel* and c-*rel*. The sequence differences between v-*rel* and c-*rel* include: 10 missense transitions, four missense transversions, three places where Rev-T has a small in-frame

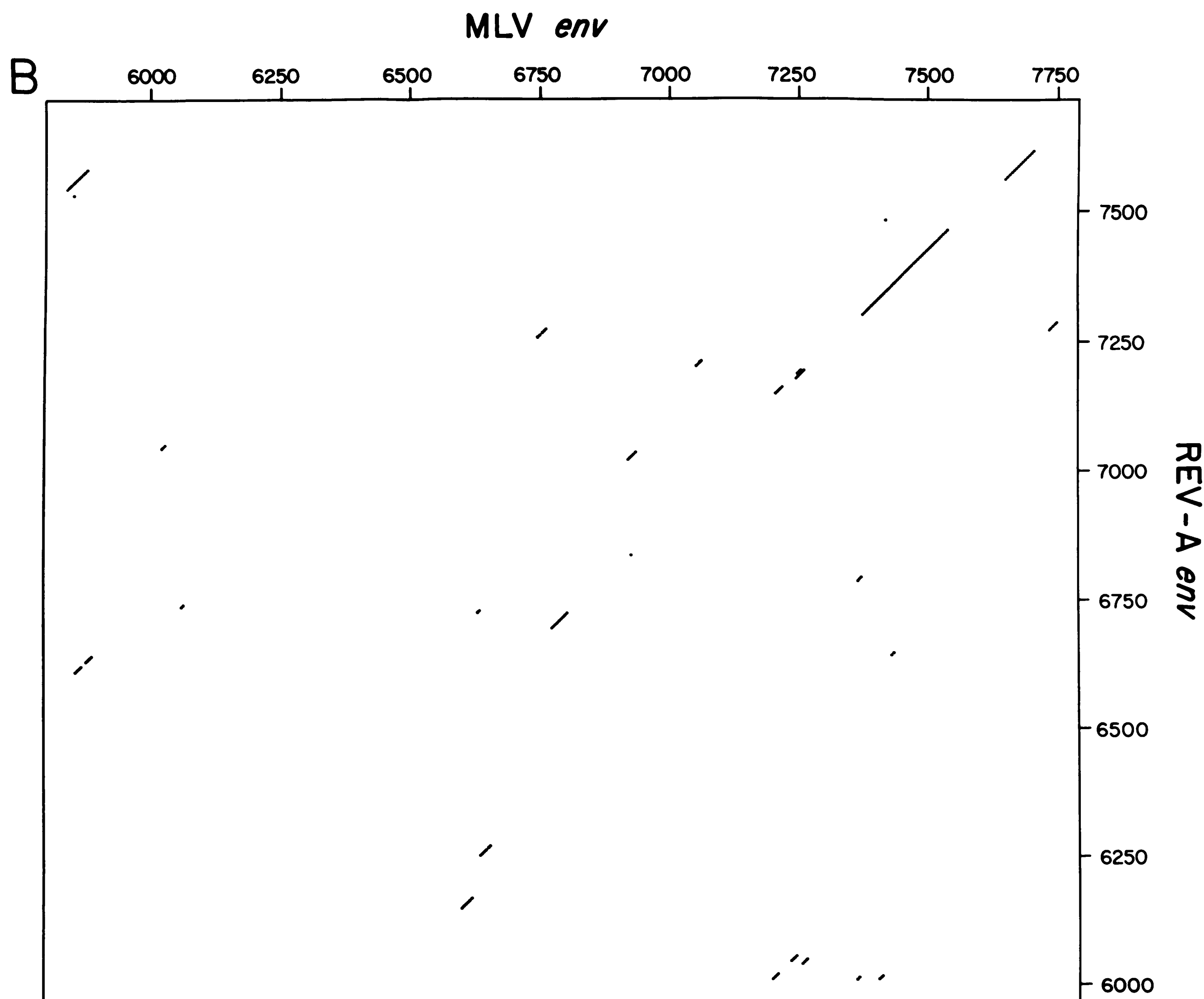


FIG. 4. Comparison of MLV and Rev-A *pol* and *env* AA sequences. Dot matrix comparisons of the predicted AA sequences of the *pol* and *env* genes of MLV and the *pol* and *env* genes of Rev-A are shown. (A) Comparison of the *pol* genes; (B) comparison of the *env* genes. In both A and B, the Rev-A sequences are on the ordinate and the MLV sequences are on the abscissa. Only the 3′ parts of the Rev-A *pol* sequences were available for this comparison. Each dot represents a window in which at least 7 of 20 AAs are identical. The coordinates at the edge of each matrix are the nucleotide coordinates used in Fig. 3 for Rev-A and by Shinnick et al. (21) for MLV.

```
REV-A                                                          T                         G
REV-T                             AAGCTTGCCAGCATAACCAACCAAACCTCGCTTAAGTCCCTACAGGCCCTCCAAGCTACTAGGTCTCTGGCTC  2980
REV-A    XX      A                 C                 XX  XXXXXX        C  TT                   T        T
REV-T GGGCAGCGCTGCGCGCCAACTGCCCCAGAAAGAAGCTCAGTAAGACCGTACCCCACTGTTCCCACTGTTCCAACCTGGTGATCTCGCCTTCGTTAGGAAGCACGACTACCAGCAGTCGGG  3100
REV-A                               A   A                  C                   A     T          A
REV-T GCCACGGTGGGACGGACCCTACACTGTAGTCCTCAGTGCCCCCACCGCGGTAAAGGTCGCTGGGAAGACCCCGTGGGTCCACCACTCTCGACTCTAGAAAGCTCCTGACAACCAAGAAGA  3220
REV-A      G       G        C           AAGTT  C AGGCGGGCA   CCC AATT TT TTGT GTTT  T GGG TT GGG CCAC GCCGAG GTCACCCC TGCAG
           Cys     Asp      Ser         LysValAspGlnAlaGlyLysThrLeuIleLeuLeuValValTrpTrpGlyPheGlyThrThrAlaGluGlyHisProLeuGln
      env INITIATION
c-rel  GCTCTAAAT A AA GTT TGTT  GTT T A [                                     T            C
                                  .Stop                                  Met
REV-T ATGGACTTTCTCACCAACCTCCGATTCACTGAGGGTATCTCAGAGCCCTACATTGAAATATTTGAACAACCCAGGCAAAGGGGTACGCGTTTCAGATATAAATGTGAAGGAAGATCAGCT  3340
      MetAspPheLeuThrAsnLeuArgPheThrGluGlyIleSerGluProTyrIleGluIlePheGluGlnProArgGlnArgGlyThrArgPheArgTyrLysCysGluGlyArgSerAla
      rel POLYPROTEIN INITIATION
c-rel                exon 0                    ][                  exon 1                         T
REV-T GGTAGCATTCCAGGAGAACACAGTACTGACAACAACAAGACATTCCCATCCATACAGATCCTAAACTATTTTGGAAAAGTCAAAATAAGAACTACATTGGTAACAAAGAACGAACCCTAC  3460
      GlySerIleProGlyGluHisSerThrAspAsnAsnLysThrPheProSerIleGlnIleLeuAsnTyrPheGlyLysValLysIleArgThrThrLeuValThrLysAsnGluProTyr
c-rel                       A                                     G      ][                                   G
                           .Asp                                   Arg .
REV-T AAGCCACACCCTCACGATCTAGTTGGAAAAGGCTGCAGAGATGGCTACTATGAAGCAGAGTTTGGGCCCGAACGGCAAGTCTTGTCTTTTCAGAATTTGGGAATTCAATGTGTGAAGAAA  3580
      LysProHisProHisAspLeuValGlyLysGlyCysArgAspGlyTyrTyrGluAlaGluPheGlyProGluArgGlnValLeuSerPheGlnAsnLeuGlyIleGlnCysValLysLys
c-rel                exon 2                     ][         C   C
REV-T AAAGACCTGAAAGAATCAATTTCTTTGCGAATCTCAAAGAAAATCAATCCCTTTAATGTGCCTGAGGAACAGTTGCATAACATCGATGAGTACGATCTCAACGTTGTCCGCCTCTGTTTC  3700
      LysAspLeuLysGluSerIleSerLeuArgIleSerLysLysIleAsnProPheAsnValProGluGluGlnLeuHisAsnIleAspGluTyrAspLeuAsnValValArgLeuCysPhe
c-rel                exon 3                                        ][
REV-T CAAGCTTTCCTCCCTGATGAACATGGCAACTACACATTGGCTCTTCCTCCTTTGATTTCCAACCCAATCTATGACAACAGAGCTCCCAACACGGCAGAACTGAGGATTTGTCGTGTGAAT  3820
      GlnAlaPheLeuProAspGluHisGlyAsnTyrThrLeuAlaLeuProProLeuIleSerAsnProIleTyrAspAsnArgAlaProAsnThrAlaGluLeuArgIleCysArgValAsn
c-rel                exon 4                           ][
REV-T AAGAACTGTGGAAGTGTAAAGGGAGGAGATGAAATTTTTCTTCTGTGTGACAAAGTTCAAAAAGATGACATAGAGGTCAGATTTGTCTTGGGCAACTGGGAGGCAAAGGGCTCCTTCTCC  3940
      LysAsnCysGlySerValLysGlyGlyAspGluIlePheLeuLeuCysAspLysValGlnLysAspAspIleGluValArgPheValLeuGlyAsnTrpGluAlaLysGlySerPheSer
c-rel                                                  A                 exon 5            A          A
                                                       Arg                                             .Asp
REV-T CAAGCTGATGTTCATCGCCAGGTCGCAATTGTATTTAGAACACCGCCGTTCCTCGGAGACATCACAGAACCCATCACGGTGAAGATGCAGTTACGAAGGCCTTCAGACCAGGCAGTCAGT  4060
      GlnAlaAspValHisArgGlnValAlaIleValPheArgThrProProPheLeuGlyAspIleThrGluProIleThrValLysMetGlnLeuArgArgProSerAspGlnAlaValSer
c-rel                                 ][ T  A                           exon 6a   T          ][
                                         SerTyr                                   .Leu
REV-T GAACCAGTGGATTTCAGATATTTACCAGATGAAGAGGATCCGTCTGGCAACAAAGCAAAAAGGCAAAGATCAACACTGGCTTGGCAAAAACCCATACAGGACTGCGGATCAGCTGTGACA  4180
      GluProValAspPheArgTyrLeuProAspGluGluAspProSerGlyAsnLysAlaLysArgGlnArgSerThrLeuAlaTrpGlnLysProIleGlnAspCysGlySerAlaValThr
                                              ATT                                                  GCG
c-rel                exon 6b                  Ile     ][          exon 7                           Ala
REV-T GAGAGGCCAAAAGCGGCTCCTATCCCCACTGTCAACCCTGAAGGAAAGCTGAAGAAAGAACCAAATATGTTTTCACCTACGCTGATGCTGCCTGGGCTAGGAACACTGAGCTCCAGTCAG  4300
      GluArgProLysAlaAlaProIleProThrValAsnProGluGlyLysLeuLysLysGluProAsnMetPheSerProThrLeuMetLeuProGlyLeuGlyThrLeuSerSerSerGln
c-rel
REV-T ATGTACCCTGCATGCAGCCAGATGCCCACCCAGCCTGCGCAGCTTGGCCCTGGGAAGCAGGACACACTCCATTCCTGCTGGCAGCAGCTGTACAGCCCCTCCCCTTCAGCCAGCAGCCTG  4420
      MetTyrProAlaCysSerGlnMetProThrGlnProAlaGlnLeuGlyProGlyLysGlnAspThrLeuHisSerCysTrpGlnGlnLeuTyrSerProSerProSerAlaSerSerLeu
                                                                         CCACGA
c-rel    A          A                                                    TyrHisAsp                    G
         Met.       Asn                                                                               Asp
REV-T CTCAGCTTGCACTCACACAGCAGCTTCACAGCGGAAGTGCCTCAGCCTGGTGCTCAGGGCAGTAGCTCTCTCCCGGCCTATAACCCACTGAACTGGCCTGATGAGAAGAATTCCAGTTTT  4540
      LeuSerLeuHisSerHisSerSerPheThrAlaGluValProGlnProGlyAlaGlnGlySerSerSerLeuProAlaTyrAsnProLeuAsnTrpProAspGluLysAsnSerSerPhe
c-rel                                                  A                                                              G
                                                       Asp                                                            Ala
REV-T TACAGGAATTTTGGCAACACACATGGGATGGGAGCAGCGTTGGTGTCAGCTGCAGGCATGCAGAGTGTTTCCAGTAGCAGCATCGTCCAGGGCACTCATCAGGCCAGTGCCACTACTGCA  4660
      TyrArgAsnPheGlyAsnThrHisGlyMetGlyAlaAlaLeuValSerAlaAlaGlyMetGlnSerValSerSerSerIleValGlnGlyThrHisGlnAlaSerAlaThrThrAla
                                                                                                                 CTGTACGACAA
                                                                                                                 LeuTyrAspAsn
REV-A  CA GG G  AT   T                   T                          A  A                     G A      C
       GlnGlyGlyIleCysLeuAlaLeuGlnGluLysCysCysPheTyrAlaAsnLysSerGlyIleValArgAspLysIleArgLysLeuGlnGluAspLeuLeuAlaArgLysArgAla Pro
c-rel          A      GAAAC  A GACATGA CTGCA CAG C CAA TTTG  AAGTATAC  AGAT  T A T TAAGCA CCACAGGCA CAG TCCA  AG T CCTG AACATGT
               Asn    GluThrAsnAspMetAsnCysThrMetLeuAsnPheGluLysTyrThrGlnMetLeuAsnValSerAsnHisArgGlnGlnLeuHisGlnValProAlaThrCys
REV-T AGCATCATGACCATGCCTCGCACTCCAGGAGAAGTGCCGTTTTTACGCCAACAAGTCGGGTATCGTTCGTGACAAGATCCGGAAACTCCAAGGGGACCTTGTCGCGAGGAAACGTGCACC  4780
      SerIleMetThrMetProArgThrProGlyGluValProPheLeuArgGlnGlnValGlyTyrArgSerEnd  rel POLYPROTEIN TERMINATION

REV-A                                             C       env CONTINUES
        LeuTrpAsnGlyLeuAsnGlyPheLeuProTyrLeuLeuProSerLeu
c-rel   TCC   CAGC  C  GCA CAC    C  AG TCACAACCAAA GTAGCT  c-rel CONTINUES
      ProProGlySerAlaTrpGlnHisSerLeuMetSerGlnProAsnValAla
REV-T CCCTGTGGAACGGCTTGAACGGCTTCCTTCCATATTTGCTACCCTTGTTG  env CODING SEQUENCE CONTINUES
```

FIG. 5. Comparison of the Rev-T sequence with Rev-A and c-*rel*. The sequence of Rev-T is shown with the AA sequence of the predicted *rel* protein. Some of the sequences of Rev-A and c-*rel* have been put above the Rev-T sequence as indicated. The positions where the sequences of Rev-A or c-*rel* are the same as Rev-T are blank. The positions where they differ from Rev-T are indicated with the letter of the nucleotide. An × indicates that there is a nucleotide present in Rev-T that is not present in Rev-A. Short horizontal lines indicate a position where either c-*rel* or Rev-A has nucleotides not present in Rev-T. All three reading frames are closed 5′ to the transcriptional initiation codon for the predicted rel polyprotein of Rev-T. 5′ and 3′ to the transcriptional initiation codon for *env* there is another open reading which appears to code for the *pol* gene (Fig. 3). The translations of the *pol* and *env* genes of Rev-A are shown in Fig. 3. The symbol][indicates the presumed recombination sites at the boundaries of regions of c-*rel* that are homologous to v-*rel*. When the boundary position is ambiguous, these marks have been placed at a position consistent with the cleavage site in a consensus RNA splice donor and acceptor (see Fig. 6). The first nucleotide of the translational initiation codon for the predicted env and rel polypeptides has been given the coordinate 3221, which corresponds to 3.220 kbp on the map of Rev-T in Fig. 2; 0.034 kbp on the map of c-*rel* in Fig. 2; and 5.958 kbp on the map of Rev-A in Fig. 2. The termination of the predicted rel polypeptide corresponds to 4.732 kbp in the map of Rev-T in Fig. 2 and 7.432 kbp in the map of Rev-A in Fig. 2; 4855 corresponds to 23.73 kbp in the map of c-*rel* in Fig. 2.

TABLE 1. Differences between v-*rel* and c-*rel*

Position	Type of difference and position	Sequence in:	
		v-*rel*	c-*rel*
5′ end in Rev-T		*env* fusion	
	Missense differences		
	transitions		
	3306	C Thr	T Met
	3492	G Gly	A Asp
	3537	A Gln	G Arg
	3995	G Gly	A Arg
	4100	C Pro	T Ser
	4152	C Pro	T Leu
	4440	G Ser	A Asp
	4529	A Asn	G Asp
	4596	G Gly	A Asp
	4655	A Thr	G Ala
	Transversions		
	4053	C Ala	A Asp
	4104	C Ser	A Tyr
	4427	T Leu	A Met
	4471	C Thr	A Asn
	Insertions and deletions[a]		
	4332	CTGAAG	CTGATTAAG
		LeuLys	LeuIleLys
	4289	CTGAGC	CTGGCGAGC
		LeuLys	LeuAlaLys
	4401	TATAAC	TACCACGATAAC
		TyrAsn	TyrHisAspAsn
3′ end		Out-of-frame *env* fusion	
	Silent differences		
	Transitions		
	3219	T	C
	3351	C	T
	3480	A	G
	3553	T	C
	3558	T	C
	4049	G	A
	4165	C	T

[a] The positions of the deletions and insertions are ambiguous.

deletion of sequences relative to c-*rel*, and seven silent transitions. Most of the coding sequence differences between c-*rel* and v-*rel* are nonconservative AA changes.

The boundaries between the regions of c-*rel* that are homologous to v-*rel* have been placed in Fig. 5 at the cleavage site in an RNA transcript which would be used by the cell RNA splicing machinery. Table 2 shows the flanking sequences of the regions of c-*rel* that are homologous to v-*rel* and the predicted cleavage site that would be made in an RNA transcript by the cell RNA splicing machinery. Each of the boundaries of regions of c-*rel* that are homologous to v-*rel*, except for the 3′ boundary of exon 7, is consistent with their being exon-intron boundaries. Furthermore, the reading frame 5′ or 3′ or both to many of the regions of c-*rel* that are homologous to v-*rel* has a termination codon in the sequences in the close vicinity of the presumed exon-intron boundaries. Within a few rounds of replication, retroviruses splice out introduced intervening sequences (20, 22). Therefore, it is very likely that in the c-*rel* locus, the regions that are homologous to v-*rel* are exons and the intervening regions between these homologous sequences are introns.

Transduction of c-*rel*. Figure 6 shows the nucleotide sequences at the 5′ and 3′ junctions of Rev-A, Rev-T, and c-*rel*. At one site in the 5′ junction sequences between c-*rel* and the parental retrovirus, there is in Rev-T the remnant of a possible splice donor and acceptor, AGGG. c-*rel* at this site has the sequence TTTTGTTTTGTTAAG/G, which is similar to the consensus sequence for a splice acceptor (Y_{11}NYAG/G) (14). A plausible, presumably cryptic, splice donor can also be identified at the recombination site in Rev-A (at 3,255 bp in Fig. 5 and at 5,991 bp in Fig. 3). The sequence in Rev-A at the junction is AGG/GTAA, which is similar to the consensus sequence for a splice donor (MG/GTRT). It is possible that Rev-T was formed by the use of the cellular RNA splicing machinery to generate a recombinant between the parental virus and c-*rel*. This process would be analogous to the loss of the sequences between the regions of c-*rel* that are homologous to v-*rel*. However, both the 5′ and 3′ sequences contain small regions of homology between Rev-A, Rev-T, and c-*rel*. Therefore, it is possible that homologous recombination was involved in the formation of Rev-T.

DISCUSSION

Rev-A. We have identified two large open reading frames between 4.6 and 7.74 kpb in Rev-A which presumably encode the 3′ part of the *pol* gene product and the *env* polyprotein. The predicted amino acid sequences are more related to the predicted amino acid coding sequence for the *pol* and *env* genes of MLV than to those of the Prague C strain of RSV. It has been shown previously that there is little sequence homology (9) and little antigenic relationship (28) between reticuloendotheliosis viruses and avian leukosis-sarcoma viruses. Reverse transcriptase from avian leuko-

TABLE 2. Nucleic acid sequences of junctions of regions of c-*rel* that are homologous to v-*rel*

Region	Flanking sequences of homologous regions	
	5′ end	3′ end
Exon 0	TTGTTTTGTTTTAAG/G.	.CAG/GTAATA
Exon 1	CTCTTATTCTTGTAG/A.	.GTC/GTGAGT
Exon 2	TCATGTGTACTTCAG/T.	.ATG/GTGAGT
Exon 3	CCTTTTGCTGTCCAG/T.	.ACA/GTAAGT
Exon 4	CCTGTTTATTTCCAG/G.	.AAG/GTAACG
Exon 5	TTTCTAATGAACTAG/A.	.AGG/GTATGG
Exon 6a	CTCCTTTCATTATAG/A.	.GTG/GTAAGG
Exon 6b	TTTCTTCTGTCGCAG/G.	.AAG/GTAAAT
Exon 7	CTTTCCTTTTCACAG/A.	.ATGAACATGG
RNA splice consensus[a]	Acceptor YYYYYYYYYYYNYAG/G.	Donor .MAG/GTRAGT[b]

[a] See reference 14.
[b] M is A or C; R is A or G; Y is C or T; N is A or G or C or T.

sis-sarcoma viruses is antigenically unrelated to reticuloendotheliosis virus reverse transcriptase (1). In contrast, reverse transcriptase from reticuloendotheliosis virus is antigenically related to reverse transcriptase of murine leukemia viruses (2).

The coding sequences of the *pol* and *env* genes in both Rev-A and MLV overlap by ca. 60 nucleotides. The coding sequences of *pol* and *env* overlap by ca. 120 nucleotides in the Prague C strain of RSV, but unlike Rev-A and MLV, the translational initiation codon for the env polyprotein is 5′ to the RNA cleavage site in the *gag* coding sequences. Rev-A and MLV have RNA splice acceptors ca. 260 and 280 nucleotides, respectively, upstream of the translational initiation codon for *env*. The organization of *env* and *pol* transcription and translation in Rev-A is additional evidence that reticuloendotheliosis viruses are more closely related to murine leukemia viruses than to avian leukosis-sarcoma viruses. Previously, $tRNA^{Pro}$ has been identified as the primer associated with virion RNA of reticuloendotheliosis viruses and murine leukemia viruses (15, 33). In contrast, avian leukosis-sarcoma viruses have $tRNA^{Trp}$ primers (7).

Avian leukosis-sarcoma virus virions contain reverse transcriptase of the form α-β. The β peptide has an endonuclease activity and can be cleaved by a virus-coded protease into an α peptide and p32 which has endonuclease activity (35). In murine type C viruses, all virion reverse transcriptase activity is separable from peptides which have endonuclease activity, but as in avian leukosis-sarcoma viruses, both peptides are believed to be cleaved from a precursor polypeptide (11). In avian leukosis-sarcoma viruses, the endonuclease activity associated with p32 and reverse transcriptase is from the carboxy end of the β peptide. Therefore, it is likely that the endonuclease activity is coded by sequences at the 3′ end of the *pol* gene. Further, it is likely that at least some of the Rev-A sequence we have determined codes for a virion-associated endonuclease. Infectious integration-defective mutants of spleen necrosis virus, a reticuloendotheliosis virus, have been made by mutating within the portion of *pol* that we have sequenced (A. Panganiban, personal communication). It is possible that these integration-defective mutants have lost the endonuclease activity without loss of reverse transcriptase activity.

Rev-T. We have predicted that *v-rel* is expressed as an *env*-v-*rel*-out-of-frame-*env* fusion polypeptide with 503 AAs. We have not detected in large computer searches any sequences that are significantly related to v-*rel*.

Recently, Stephens et al. (24) reported the sequence of v-*rel* based on the sequence of a recombinant DNA clone that contained the 5′ 30% of v-*rel* and a provirus obtained by Chen et al. (3) that is known to be infectious but is defective for transformation. Their sequence and ours are the same.

The predicted rel fusion polypeptide is predicted to be expressed from the previously identified subgenomic RNA transcript (Wilhelmsen, unpublished data). It was expected that the subgenomic transcript was spliced by using the same signals as *env*. Sequence analysis has allowed us to identify

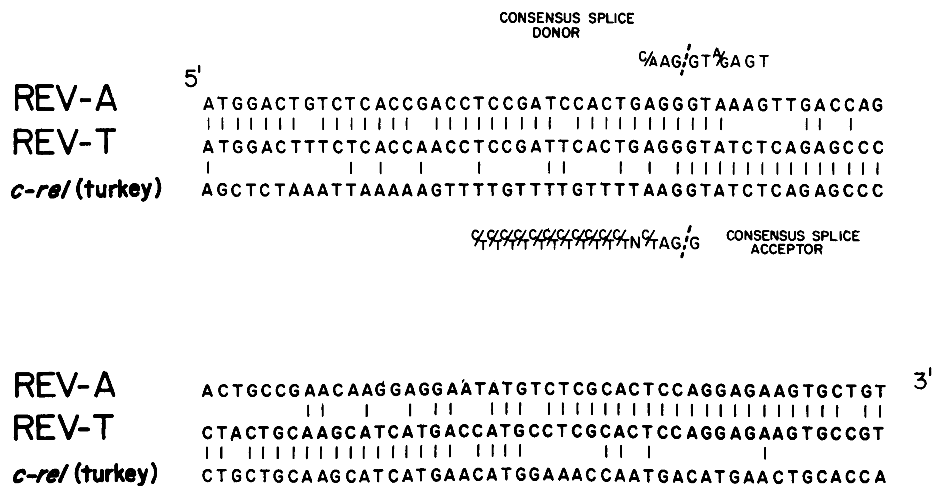


FIG. 6. Nucleic acid sequences of Rev-A, Rev-T, and c-*rel* (turkey) at the junction of c-*rel* with the ancestral virus of Rev-T. The upper three sequences in this figure show the 5′ junction sites used to form Rev-T from c-*rel* and an ancestral virus. Rev-A is believed to be similar to the ancestral virus of Rev-T. Above and below the sequences of Rev-A, Rev-T, and c-*rel* (turkey) are consensus sequences for RNA splice donors and acceptors, respectively. The cleavage sites used in RNA splicing are indicated by a dotted slash in the consensus sequences. Sequence identities between Rev-A and Rev-T or Rev-T and c-*rel* (turkey) are indicated with vertical lines. The lower sequences in this figure show the 3′ junction site used to form Rev-T from c-*rel* (turkey) and an ancestral virus.

the position of the *env* RNA splice sequences in Rev-T and to identify a significant mutation that might affect their usage by Rev-T. Chen and Temin have previously determined that the expression of the transformation phenotype is dependent on a large deletion of Rev-T relative to Rev-A (4). The use of a splice acceptor other than the *env* splice acceptor by Rev-T may also be necessary for the proper level of expression of v-*rel*. Other oncogenes are expressed from spliced subgenomic RNA transcripts (36). Whether an oncogene is expressed from a spliced or an unspliced RNA may be important in determining the level of its expression (G. Tarpley and H. M. Temin, manuscript in preparation).

c-*rel*. There are nine regions in the c-*rel* (turkey) locus that are homologous to v-*rel*. We previously reported that there are eight regions of c-*rel* that are homologous to v-*rel* (37). The sequence data presented here indicate that region 6 is divided by DNA sequences that are not homologous to v-*rel* (Fig. 1). All of the v-*rel* sequences could have come from the one c-*rel* locus of turkeys we identified.

Proto-oncogenes have been found that are colinear with oncogenes. Other proto-oncogenes have a few small regions of DNA (putative introns) between regions of DNA homologous to oncogenes (putative exons). c-*rel* is like other complex proto-oncogenes which have many exons that are separated by introns. v-*src* is unusual because it is derived from cellular sequences of two transcriptional domains (27).

We have previously determined that c-*rel* is transcribed at low levels in many tissues. Analysis of c-*rel* mRNA from chicken spleen cells indicates that more of the c-*rel* locus is transcribed than the sequences of c-*rel* that are homologous to v-*rel* (5).

Expression of the v-*rel* sequences by Rev-T in appropriate target cells changes the phenotype of those cells from normal controlled growth to uncontrolled growth. There are three models for explaining why expression of v-*rel* in target cells causes transformation: (i) the products of v-*rel* and c-*rel* are functionally similar but there is a quantitative difference in the level of expression of c-*rel* and v-*rel*; (ii) the gene products of c-*rel* and v-*rel* are qualitatively different in function; or (iii) there are both qualitative and quantitative differences between c-*rel* and v-*rel* expression.

For some oncogenes, it has been possible to distinguish in a qualified manner between the quantitative and qualitative difference models. The modification of the 3'-most coding sequences of the proto-oncogene c-*src* has previously been implicated as the only possible significant structural difference between v-*src* that is transforming and c-*src* that is nontransforming (27). c-*src* may be different qualitatively or quantitatively or both from v-*src*. The *ras* oncogenes in the T24 bladder carcinoma cell line v-Ha-*ras* and the v-Ki-*ras* gene products have been shown to have a qualitative difference from c-Ha-*ras* for transformation of NIH-3T3 cells due to the substitution of one AA (for example, see reference 26). The coding sequence differences between the oncogene v-*fos* and the proto-oncogene c-*fos* do not affect their ability to transform a rat cell line (13, 30). v-*fos* is postulated to transform cells because it is expressed in a quantitatively different manner.

There are many differences between v-*rel* expression by Rev-T and c-*rel* expression in situ that could cause a qualitative difference in the function of their gene products. The 5' end of the c-*rel* gene product differs from the 5' end of the v-*rel* gene product, which is a fusion with *env* sequences. The v-*rel* gene product has out-of-frame-*env* coding sequences substituted for the 3' coding sequences of c-*rel*. Fourteen-base-pair changes have been identified that would cause the v-*rel* and c-*rel* gene products to contain different AAs. v-*rel* also has three very small in-frame deletions relative to c-*rel*.

The codon differences and the in-frame deletion differences may be of no consequence and may be due to sequence differences between the c-*rel* locus in turkey 7 and the specific locus from which v-*rel* was acquired. We have previously observed dimorphism of the c-*rel* locus in turkeys, but there may be other alleles of c-*rel* that are more similar to v-*rel* (36, 37).

Alternatively, it is possible that c-*rel* expression is very closely regulated in different cell populations and that transformation by Rev-T is the result of inappropriate expression of a gene functionally equivalent to c-*rel*. v-*rel* is expressed by transcription promoted by the LTR promoter and by translation initiated and terminated in viral sequences. It is plausible that, if c-*rel* expression is regulated at the level of transcription or translation, v-*rel* expression has escaped regulation. c-*rel* expression could also be regulated by its position on a chromosome. Since Rev-T integrates at random in the genome, v-*rel* may be released from normal regulation.

The models for acquisition of proto-oncogenes by retroviruses (35) include: integration followed by deletion between viral and cell DNA sequences; virus integration 5' to the target sequences with a large read-through transcript being spliced; and RNA-RNA recombination. The structural relationships of Rev-A, Rev-T, and c-*rel* do not allow discrimination between the various models for acquisition of sequences by retroviruses. Since there are plausible splice sequences in both c-*rel* and Rev-A at the 5' junction, it is most likely that v-*rel* was acquired by the use of the cellular RNA splicing machinery and then homologous recombination at the 3' junction. However, examination of the consensus sequences for RNA splice donors and acceptors aligned with respect to their cleavage sites shows that the alignment of actual RNA splice donors and acceptors has a high probability of having a short region of homology. There are short regions of homology at the 5' and 3' junctions of the proto-oncogene c-*rel* and Rev-A, a virus similar to the ancestor of Rev-T. The short regions of homology at the junctions may indicate that homologous recombination was responsible for the acquisition of v-*rel*.

Other highly oncogenic retroviruses do not appear to have used RNA splicing to acquire their oncogenes. At least one highly oncogenic retrovirus has short regions of homology at the virus-oncogene boundary when the sequences of helper virus or related virus, the oncogene, and the proto-oncogene are compared (30). Other highly oncogenic retroviruses have little if any sequence homology among helper virus, oncogene, and proto-oncogene at the virus-oncogene boundary (10, 16, 17, 21, 25, 27, 29, 31, 34).

ACKNOWLEDGMENTS

We thank Ann Troup and Susan Hellenbrand for excellent technical assistance, R. Doolittle and C. Burks for assistance in comparing the sequence of v-*rel* with other known sequences, and C. Miller, N. Panganiban, B. Sugden, B. Sylla, and G. Tarpley for helpful comments on the manuscript.

This investigation was supported by Public Health Service research grants CA-07175 and CA-22443 from the National Cancer Institute. K.C.W. is a predoctoral trainee supported by Public Health Service training grant CA-09135 from the National Insitutes of Health. H.M.T. is an American Cancer Society Research Professor.

LITERATURE CITED

1. **Bauer, G., and H. M. Temin.** 1980. Radioimmunological comparison of the DNA polymerases of avian retroviruses. J. Virol. **33:**1046–1057.
2. **Bauer, G., and H. M. Temin.** 1980. Specific antigenic relationships between the RNA-dependent DNA polymerases of avian reticuloendotheliosis viruses and mammalian type C retroviruses. J. Virol. **34:**168–177.
3. **Chen, I. S. Y., T. W. Mak, J. J. O'Rear, and H. M. Temin.** 1981. Characterization of reticuloendotheliosis strain T DNA and isolation of a novel variant of reticuloendotheliosis virus strain T by molecular cloning. J. Virol. **40:**800–811.
4. **Chen, I. S. Y., and H. M. Temin.** 1982. Substitution of 5′ helper sequences into non-rel portion of reticuloendotheliosis virus strain T suppresses transformation of chicken spleen cells. Cell **31:**111–120.
5. **Chen, I. S. Y., K. C. Wilhelmsen, and H. M. Temin.** 1983. Structure and expression of *c-rel*, the cellular homolog to the oncogene of reticuloendotheliosis virus strain T. J. Virol. **45:**104–113.
6. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–396.
7. **Harada, F., R. C. Sawyer, and J. E. Dahlberg.** 1975. A primer ribonucleic acid for initiation of the vitro Rous sarcoma virus deoxyribonucleic acid synthesis. Nucleotide sequence and amino acid acceptor activity. J. Biol. Chem. **250:**3487–3497.
8. **Hu, S. S. F., M. M. C. Lai, T. C. Wong, R. S. Cohen, and M. Sevoian.** 1981. Avian reticuloendotheliosis virus: characterization of genome structure by heteroduplex mapping. J. Virol. **37:**899–907.
9. **Kang, C.-Y., and H. M. Temin.** 1973. Lack of sequence homology among RNAs of avian leukosis-sarcoma viruses, reticuloendotheliosis viruses, and chicken endogenous RNA-directed DNA polymerase activity. J. Virol. **12:**1314–1324.
10. **Kempnauer, K.-H., T. J. Gonda, and J. M. Bishop.** 1982. Nucleotide sequence of the retroviral leukemia gene v-*myb* and its cellular progenitor v-*myb*: the architecture of a transduced oncogene. Cell **31:**453–463.
11. **Kopchick, J. J., J. Harless, B. S. Geisser, R. Killam, R. R. Hewitt, and R. B. Arlinghaus.** 1981. Endodeoxyribonuclease activity associated with Rauscher murine leukemia virus. J. Virol. **37:**274–283.
12. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65:**499–560.
13. **Miller, A. D., T. Curran, and I. M. Verma.** 1984. *c-fos* protein can induce cellular transformation: a novel mechanism of activation of a cellular oncogene. Cell **36:**51–60.
14. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10:**459–472.
15. **Peters, G. G., and C. Glover.** 1980. Low-molecular-weight RNAs and initiation of RNA-directed DNA synthesis in avian reticuloendotheliosis virus. J. Virol. **33:**708–716.
16. **Reddy, E. P., R. K. Reynolds, D. K. Watson, R. S. Schultz, J. Lautenberger, and T. S. Papas.** 1983. Nucleotide sequence analysis of the proviral genome of avian myelocytomatosis virus (MC29). Proc. Natl. Acad. Sci. U.S.A. **80:**2500–2504.
17. **Reddy, E. P., M. J. Smith, and S. A. Aaronson.** 1981. The complete nucleotide sequence of Moloney murine sarcoma virus genome. Science **214:**445–450.
18. **Rice, N. R., R. R. Hiebsch, M. A. Gonda, H. R. Bose, Jr., and R. V. Gilden.** 1982. Genome of reticuloendotheliosis virus: characterization by use of cloned proviral DNA. J. Virol. **42:**237–252.
19. **Shimotohno, K., S. Mizutani, and H. M. Temin.** 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. Nature (London) **285:**550–554.
20. **Shimotohno, K., and H. M. Temin.** 1982. Loss of intervening sequences in genomic mouse α-globin DNA inserted in an infectious retrovirus vector. Nature (London) **299:**265–268.
21. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293:**543–548.
22. **Sorge, J., and S. H. Hughes.** 1982. Splicing of intervening sequences introduced into an infectious retrovirus vector. J. Mol. Appl. Genet. **1:**547–559.
23. **Staden, R.** 1980. A new method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res. **8:**3673–3694.
24. **Stephens, R. M., N. R. Rice, R. R. Hoebsch, H. R. Bose, Jr., and R. V. Gilden.** 1983. Nucleotide sequence of v-*rel*: the oncogene of reticuloendotheliosis virus. Proc. Natl. Acad. Sci. U.S.A. **80:**6229–6233.
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. **Tabin, C. J., S. M. Bradley, C. I. Bargmann, R. A. Weinberg, A. G. Papageorge, E. M. Scolnick, R. Dhar, D. R. Lowy, and E. H. Chang.** 1982. Mechanism of activation of a human oncogene. Science **300:**143–149.
27. **Takeya, T., and H. Hanafusa.** 1983. Structure and sequence of the cellular gene homologous to the RSV *src* gene and the mechanism for generating the transforming virus. Cell **32:**881–890.
28. **Theilen, G. H., R. F. Zeigel, and M. J. Twiehaus.** 1966. Biological studies with RE virus (strain T) that induces reticuloendotheliosis in turkeys, chickens, and Japanese quails. J. Natl. Cancer Inst. **37:**731–738.
29. **Van Beveren, C., J. A. Galleshaw, V. Jonas, A. J. M. Berns, R. F. Doolittle, D. J. Donoghue, and I. M. Verma.** 1981. Nucleotide sequence and formation of the transforming gene of a mouse sarcoma virus. Nature (London) **289:**258–262.
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.
31. **Van Beveren, C., F. van Straaten, J. A. Galleshaw, and I. D. Verma.** 1981. Nucleotide sequence of the genome of a murine sarcoma virus. Cell **27:**97–107.
32. **Watanabe, S., and H. M. Temin.** 1983. Construction of a helper cell line for avian reticuloendotheliosis virus cloning vectors. Mol. Cell. Biol. **3:**2241–2249.
33. **Waters, L. C.** 1975. Transfer RNAs associated with the 70S RNA of AKR murine leukemia virus. Biochem. Biophys. Res. Commun. **65:**1130–1136.
34. **Watson, D. K., E. P. Reddy, P. H. Duesberg, and T. S. Papas.** 1983. Nucleotide sequence analysis of the chicken c-*myc* gene reveals homologous and unique coding regions by comparison with the transforming gene of avian myelocytomatosis virus MC29, *gag-myc*. Proc. Natl. Acad. Sci. U.S.A. **80:**2146–2150.
35. **Weiss, R., N. Teich, H. Varmus, and J. Coffin (ed.).** 1982. The molecular biology of tumor viruses, part 3: RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Wilhelmsen, K. C., I. S. Y. Chen, and H. M. Temin.** 1983. The organization of *c-rel* in chicken and turkey DNAs, p. 43–56. *In* T. E. O'Connor and F. J. Rauscher, Jr. (ed.), Oncogenes and retroviruses: evaluation of basic findings and clinical potential. Alan R. Liss, Inc., New York.
37. **Wilhelmsen, K. C., and H. M. Temin.** 1984. Structure and dimorphism of c-*rel* (turkey), the cellular homolog to the oncogene of reticuloendotheliosis virus strain T. J. Virol. **49:**521–529.