Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 6229–6233, October 1983
Biochemistry

Nucleotide sequence of v-*rel*: The oncogene of reticuloendotheliosis virus

(leukemia/tumor virus/transforming protein)

ROBERT M. STEPHENS*, NANCY R. RICE*†, RONALD R. HIEBSCH*, HENRY R. BOSE, JR.‡, AND RAYMOND V. GILDEN§

*Litton Bionetics, Inc.-Basic Research Program, National Cancer Institute-Frederick Cancer Research Facility, Frederick, MD 21701; ‡Department of Microbiology, University of Texas at Austin, Austin, TX 78731; and §Program Resources, Inc., National Cancer Institute-Frederick Cancer Research Facility, Frederick, MD 21701

Communicated by D. Carleton Gajdusek, July 14, 1983

ABSTRACT **The nucleotide sequence of v-*rel*, the oncogene carried by reticuloendotheliosis virus (REV), has been determined. The defective transforming genome REV arose through the insertion of v-*rel* (1,415 nucleotides) into the *env* gene of the helper virus REV-A. The predicted *rel* protein (503 amino acids) employs the REV-A *env* initiator and terminates within the p20E region of *env*. Because there are no natural antisera that detect the REV transforming protein, this nucleotide sequence provides the first step toward its isolation and characterization. The predicted protein is clearly distinct from all other reported transforming proteins but may be very distantly related to members of the *src* family.**

Reticuloendotheliosis virus (REV), a type C retrovirus, causes acute leukemia in young chickens. The virus is a complex of a replication-competent helper virus called REV-A and a replication defective genome called REV, which is responsible for transformation (1). REV proviral DNA contains some sequences found in the helper virus, including long terminal repeats (LTRs) and parts of the *gag*, *pol*, and *env* genes, but it also contains an additional segment of about 1.5 kilobases (kb) not found in REV-A (2–6). This segment, termed v-*rel*, is the presumptive *onc* gene of the virus. Like the other known *onc* genes, v-*rel* appears to have been derived from a normal cellular gene, for related sequences have been found in DNA from uninfected avian species (5–9). How v-*rel* relates to the other *onc* genes has been unclear. No hybridization has been detected between REV and nucleic acids of other oncogenic viruses (refs. 5 and 8; M. Shibuya and H. Hanafusa, personal communication), and because to date there has been no antiserum that detects the v-*rel* protein (10, 11), its relatedness to other transforming proteins has been unknown. We report here the complete DNA sequence of v-*rel*. The protein predicted from this sequence is clearly distinct from all of the transforming proteins published to date but does appear to be very distantly related to members of the *src* family.

MATERIALS AND METHODS

The cloning of *Sac* I fragments of REV and REV-A proviral DNA into λgt*WES*·λB has been described (6). For the purpose of sequence analysis various subclones of these DNA fragments were constructed in pBR322 or pUC8 (or both), both of which have been modified to include a *Sac* I site. A Charon 4A clone of REV-T (5) was a gift of I. Chen. Its internal 0.88-kb *Eco*RI fragment was subcloned into pBR322 and was used to confirm the sequence across the *Sac* I sites within v-*rel*.

Restriction fragments of DNA were labeled at their 3′ termini by using [α-^{32}P]dNTPs (Amersham, 3,000 Ci/mmol; 1 Ci = 3.7×10^{10} Bq) and the Klenow fragment of *Escherichia coli* DNA polymerase I (12). End-labeled fragments were then digested with the appropriate restriction enzyme and separated by agarose gel electrophoresis. The sequence was determined according to the method of Maxam and Gilbert (12), by using reactions for G, G+A, C, C+T, and A+C. Over all areas the reported sequence is the result of multiple independent determinations. Each area has been read at least four times and in all areas both DNA strands have been read. Examples of the sequence analysis strategy are shown in Fig. 1.

RESULTS AND DISCUSSION

Nucleotide Sequence of v-*rel*. The nucleotide sequence of v-*rel* and adjoining regions of REV-A-related proviral DNA is shown in Fig. 2 together with the predicted amino acid sequence of the *rel* protein. Junctions of v-*rel* with helper sequences were precisely localized by comparison with DNA sequence derived from REV-A clones. v-*rel* consists of 1,415 nucleotides.

As shown in Fig. 3*A*, there is only one reading frame of v-*rel* that is open for any significant length. This open frame begins slightly upstream from the 5′ boundary of v-*rel* and extends through the entire transforming region to a termination codon (TGA) located within REV-A-derived *env* sequences 58–60 bases beyond the 3′ terminus of v-*rel*. Both $p90^{gag-yes}$ of Y73 (15) and the protein predicted by the v-*myb* sequence of avian myeloblastosis virus (16, 17) also terminate within *env*, but the proteins specified by v-*mos*, v-*sis*, v-*fes*, v-*fps*, and v-*myc* are all predicted to terminate within the transforming region itself (18–22). One possible explanation of the arrangement found in REV is that the c-*rel* gene that was inserted into REV-A was incomplete, lacking its 3′ terminus and hence its usual terminator. This is consistent with the recent finding of sequences toward the 3′ end of c-*rel* that do not hybridize with v-*rel* (9).

Downstream from the 5′ terminus of v-*rel*, there is no ATG codon within the open reading frame for almost 800 nucleotides. However, there is an in-frame ATG immediately upstream from v-*rel*, within REV-A-derived sequences. Furthermore, the sequence surrounding this codon, A-G-A-A-T-G-G, agrees with that of the consensus initiator $\frac{A}{C}$-N-N-A-T-G-$\frac{G}{A}$ (23). Because all ATG codons even farther upstream are closely followed by terminators, it appears likely that the *rel* protein initiates at this point and contains 12 helper-related amino acids at its NH_2 terminus. That v-*rel* sequences close to its 5′ end are indeed translated is indicated by our preliminary results with

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

Abbreviations: REV, reticuloendotheliosis virus; kb, kilobase(s); LTR, long terminal repeat; M-MuLV, Moloney murine leukemia virus.

† To whom reprint requests should be addressed.

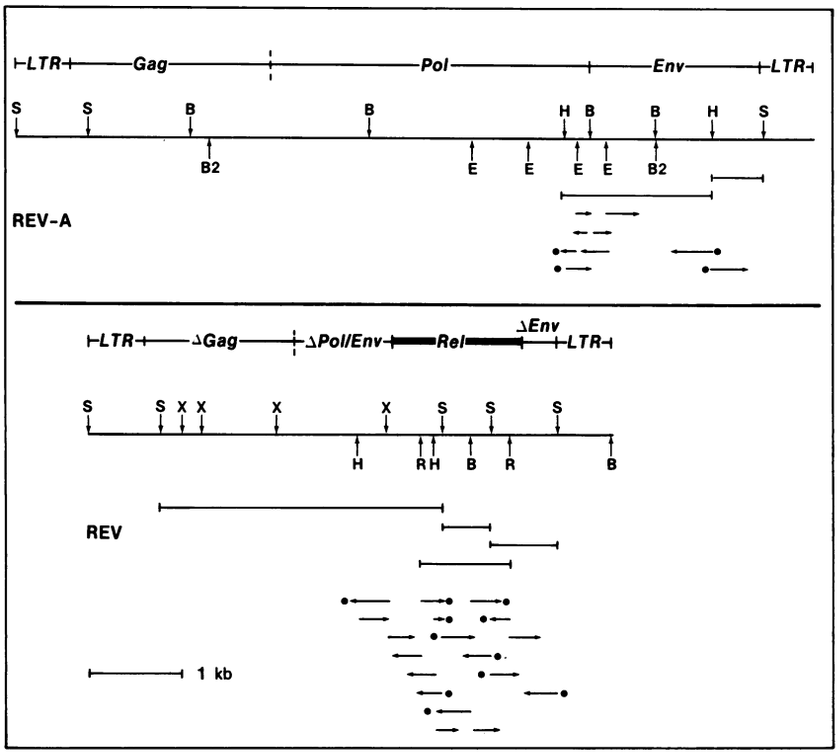


FIG. 1. Examples of sequence analysis strategy for REV-A and REV DNAs. The REV-A sequence was obtained from two pBR322 subclones, one containing the 1.6-kb *Hin*dIII fragment and the other the 0.55-kb *Hin*dIII–*Sac* I fragment immediately upstream from the 3′ LTR. The REV sequence was obtained from the three indicated *Sac* I clones and the *Eco*RI clone. DNA fragments were end-labeled at sites denoted by the beginning of each arrow and sequence was read over a distance denoted by the length of the arrow. Closed circles mean that end-labeling occurred at a site in vector DNA or that sequence was read into vector DNA. B, *Bam*HI; B2, *Bgl* II; E, *Bst*EII; H, *Hin*dIII; R, *Eco*RI; S, *Sac* I; and X, *Xba* I.

```
(-130) AGCAGTCGGG GCCACGGTGG GACGGACCCT ACACTGTAGT CCTCAGTGCC CCCACCGCGG TAAAGGTCGC TGGGAAGACC CCGTGGGTCC ACCACTCTCG ACTCTAGAAA GCTCCTGACA ACCAAGAAGA    -1

                                    env ←|→ v-rel
ATG GAC TTT CTC ACC AAC CTC CGA TTC ACT GAG GGT ATC TCA GAG CCC TAC ATT GAA ATA TTT GAA CAA CCC AGG CAA AGG GGT ACG CGT TTC AGA TAT AAA TGT GAA GGA AGA TCA GCT  120
Met Asp Phe Leu Thr Asn Leu Arg Phe Thr Glu Gly Ile Ser Glu Pro Tyr Ile Glu Ile Phe Glu Gln Pro Arg Gln Arg Gly Thr Arg Phe Arg Tyr Lys Cys Glu Gly Arg Ser Ala
  1                                 10                                  20                                  30                                  40

GGT AGC ATT CCA GGA GAA CAC AGT ACT GAC AAC AAC AAG ACA TTC CCA TCC ATA CAG ATC CTA AAC TAT TTT GGA AAA GTC AAA ATA AGA ACT ACA TTG GTA ACA AAG AAC GAA CCC TAC  240
Gly Ser Ile Pro Gly Glu His Ser Thr Asp Asn Asn Lys Thr Phe Pro Ser Ile Gln Ile Leu Asn Tyr Phe Gly Lys Val Lys Ile Arg Thr Thr Leu Val Thr Lys Asn Glu Pro Tyr
                                50                                  60                                  70                                  80

AAG CCA CAC CCT CAC GAT CTA GTT GGA AAA GGC TGC AGA GAT GGC TAC TAT GAA GCA GAG TTT GGG CCC GAA CGG CAA GTC TTG TCT TTT CAG AAT TTG GGA ATT CAA TGT GTG AAG AAA  360
Lys Pro His Pro His Asp Leu Val Gly Lys Gly Cys Arg Asp Gly Tyr Tyr Glu Ala Glu Phe Gly Pro Glu Arg Gln Val Leu Ser Phe Gln Asn Leu Gly Ile Gln Cys Val Lys Lys
                                90                                  100                                 110                                 120

AAA GAC CTG AAA GAA TCA ATT TCT TTG CGA ATC TCA AAG AAA ATC AAT CCC TTT AAT GTG CCT GAG GAA CAG TTG CAT AAC ATC GAT GAG TAC GAT CTC AAC GTT GTC CGC CTC TGT TTC  480
Lys Asp Leu Lys Glu Ser Ile Ser Leu Arg Ile Ser Lys Lys Ile Asn Pro Phe Asn Val Pro Glu Glu Gln Leu His Asn Ile Asp Glu Tyr Asp Leu Asn Val Val Arg Leu Cys Phe
                                130                                 140                                 150                                 160

CAA GCT TTC CTC CCT GAT GAA CAT GGC AAC TAC ACA TTG GCT CTT CCT CCT TTG ATT TCC AAC CCA ATC TAT GAC AAC AGA GCT CCC AAC ACG GCA GAA CTG AGG ATT TGT CGT GTG AAT  600
Gln Ala Phe Leu Pro Asp Glu His Gly Asn Tyr Thr Leu Ala Leu Pro Pro Leu Ile Ser Asn Pro Ile Tyr Asp Asn Arg Ala Pro Asn Thr Ala Glu Leu Arg Ile Cys Arg Val Asn
                                170                                 180                                 190                                 200

AAG AAC TGT GGA AGT GTA AAG GGA GGA GAT GAA ATT TTT CTT CTG TGT GAC AAA GTT CAA AAA GAT GAC ATA GAG GTC AGA TTT GTC TTG GGC AAC TGG GAG GCA AAG GGC TCC TTC TCC  720
Lys Asn Cys Gly Ser Val Lys Gly Gly Asp Glu Ile Phe Leu Leu Cys Asp Lys Val Gln Lys Asp Asp Ile Glu Val Arg Phe Val Leu Gly Asn Trp Glu Ala Lys Gly Ser Phe Ser
                                210                                 220                                 230                                 240

CAA GCT GAT GTT CAT CGC CAG GTC GCA ATT GTA TTT AGA ACA CCG CCG TTC CTC GGA GAC ATC ACA GAA CCC ATC ACG GTG AAG ATG CAG TTA CGA AGG CCT TCA GAC CAG GCA GTC AGT  840
Gln Ala Asp Val His Arg Gln Val Ala Ile Val Phe Arg Thr Pro Pro Phe Leu Gly Asp Ile Thr Glu Pro Ile Thr Val Lys Met Gln Leu Arg Arg Pro Ser Asp Gln Ala Val Ser
                                250                                 260                                 270                                 280

GAA CCA GTG GAT TTC AGA TAT TTA CCA GAT GAA GAG GAT CCG TCT GGC AAC AAA GCA AAA AGG CAA AGA TCA ACA CTG GCT TGG CAA AAA CCC ATA CAG GAC TGC GGA TCA GCT GTG ACA  960
Glu Pro Val Asp Phe Arg Tyr Leu Pro Asp Glu Glu Asp Pro Ser Gly Asn Lys Ala Lys Arg Gln Arg Ser Thr Leu Ala Trp Gln Lys Pro Ile Gln Asp Cys Gly Ser Ala Val Thr
                                290                                 300                                 310                                 320

GAG AGG CCA AAA GCG GCT CCT ATC CCC ACT GTC AAC CCT GAA GGA AAG CTG AAG AAA GAA CCA AAT ATG TTT TCA CCT ACG CTG ATG CTG CCT GGG CTA GGA ACA CTG AGC TCC AGT CAG 1080
Glu Arg Pro Lys Ala Ala Pro Ile Pro Thr Val Asn Pro Glu Gly Lys Leu Lys Lys Glu Pro Asn Met Phe Ser Pro Thr Leu Met Leu Pro Gly Leu Gly Thr Leu Ser Ser Ser Gln
                                330                                 340                                 350                                 360

ATG TAC CCT GCA TGC AGC CAG ATG CCC ACC CAG CCT GCG CAG CTT GGC CCT GGG AAG CAG GAC ACA CTC CAT TCC TGC TGG CAG CAG CTG TAC AGC CCC TCC CCT TCA GCC AGC AGC CTG 1200
Met Tyr Pro Ala Cys Ser Gln Met Pro Thr Gln Pro Ala Gln Leu Gly Pro Gly Lys Gln Asp Thr Leu His Ser Cys Trp Gln Gln Leu Tyr Ser Pro Ser Pro Ser Ala Ser Ser Leu
                                370                                 380                                 390                                 400

CTC AGC TTG CAC TCA CAC AGC AGC TTC ACA GCG GAA GTG CCT CAG CCT GGT GCT CAG GGC AGT AGC TCT CTC CCG GCC TAT AAC CCA CTG AAC TGG CCT GAT GAG AAG AAT TCC AGT TTT 1320
Leu Ser Leu His Ser His Ser Ser Phe Thr Ala Glu Val Pro Gln Pro Gly Ala Gln Gly Ser Ser Ser Leu Pro Ala Tyr Asn Pro Leu Asn Trp Pro Asp Glu Lys Asn Ser Ser Phe
                                410                                 420                                 430                                 440

TAC AGG AAT TTT GGC AAC ACA CAT GGG ATG GGA GCA GCG TTG GTG TCA GCT GCA GGC ATG CAG AGT GTT TCC AGT AGC AGC ATC GTC CAG GGC ACT CAT CAG GCC AGT GCC ACT ACT GCA 1440
Tyr Arg Asn Phe Gly Asn Thr His Gly Met Gly Ala Ala Leu Val Ser Ala Ala Gly Met Gln Ser Val Ser Ser Ser Ser Ile Val Gln Gly Gly Thr His Gln Ala Ser Ala Thr Thr Ala
                                450                                 460                                 470                                 480

      v-rel ←|→ env
AGC ATC ATG ACC ATG CCT CGC ACT CCA GGA GAA GTG CCG TTT TTA CGC CAA CAA GTC GGG TAT CGT TCG TGA CAAGATCCGG AAACTCCAAG GGGACCTTGT CGCGAGGAAA CGTGCACCCC          1562
Ser Ile Met Thr Met Pro Arg Thr Pro Gly Glu Val Pro Phe Leu Arg Gln Gln Val Gly Tyr Arg Ser END
                                490                                 500

CTGTGGAACG GCTTGAACGG CTTCCTTCCA TATTTGCTAC CCTTGTTGGG CCCCCTGTTT GGGCTCATAT TGTTCCTGAC CCTCGGCCCG                                                        1652
```

FIG. 2. DNA sequence of v-*rel* and amino acid sequence of the predicted *rel* protein. Nucleotides are numbered in the right-hand column, and every 10th amino acid is numbered. Numbering begins with the presumptive initiator methionine. Junctions of v-*rel* with REV-A-derived sequences are marked and were localized by comparing REV and REV-A proviral DNA sequences.

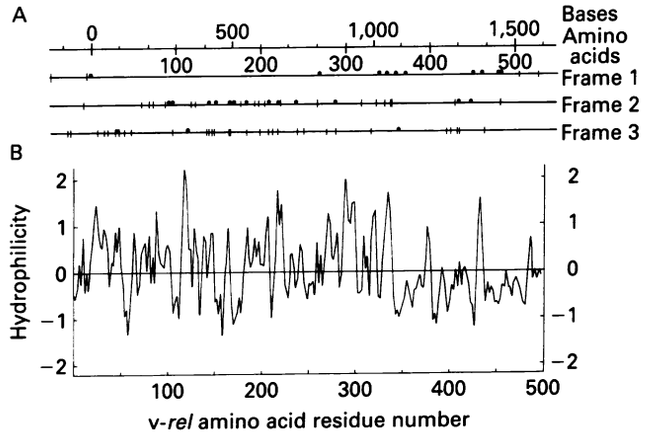


FIG. 3. Reading frames of v-*rel* and hydrophilicity of the predicted *rel* protein. (*A*) Numbering starts with the presumptive initiator ATG, which lies 37 bases upstream from the 5′ boundary of v-*rel*. The 3′ terminus of the v-*rel* insert is at base number 1,452. Closed circles denote methionines and bars denote termination codons. Thus, frame 1 is open for 1,509 bases (503 amino acids), whereas in frames 2 and 3 the maximal distance between a methionine codon and the first downstream termination codon is about 175 bases. (*B*) Hydrophilicity of the protein predicted from frame 1. According to the method of Hopp and Woods (13) each amino acid is assigned a hydrophilicity value ranging from 3.0 (very hydrophilic) to −3.4 (very hydrophobic). Average values are then computed along the length of the protein in overlapping segments of six amino acids. On this scale the average protein has a net hydrophilicity of 0.07; value for the *rel* protein is 0.1. The method of Kyte and Doolittle (14), which uses somewhat different hydrophilicity values for the amino acids, yields very similar results over most of the molecule, except that a few areas predicted above to be weakly hydrophobic are calculated to be weakly hydrophilic. These differences occur in the COOH-terminal half of the molecule and reflect the clustering of prolines and glutamines therein.

antisera raised against two synthetic peptides (amino acid residues 22–33 and 96–108 in Fig. 2). Both antisera precipitate a protein of apparent M_r of 64,000 from nonproducer transformed cells but not from uninfected chicken embryo fibroblasts (unpublished data).

The Predicted *rel* Protein. The DNA sequence predicts a protein of 503 amino acid residues, whose combined M_rs total 55,915. Notable features of the amino acid composition are elevated levels of serine, proline, and glutamine (>30% higher than that of the average protein) and depressed levels of alanine, cysteine, and tryptophan (at least 30% lower than average). Proline and serine levels stand out in particular. The total proline content is predicted to be 8.7%, more than 60% higher than average. Most of these prolines are found in the 179-residue segment between amino acids 255 and 433, where the proline content totals 14.5%. Total serine content is 9.3%, about 30% higher than average, and these residues are even more highly clustered than are the prolines. Approximately three-fourths of the 47 total serine residues are in the COOH-terminal half of the molecule, and a 97-residue segment between amino acids 385 and 481 contains 23 serines, for a content of 24%.

About 10% of the residues are acidic and 13% are basic, which are values close to average. The very low level of these charged residues in the COOH-terminal 140 amino acids results in the moderate hydrophobicity of the chain in this region, as shown in Fig. 3*B*. The hydrophobic amino acids themselves are rather evenly distributed throughout the molecule. Though there are several occurrences of seven consecutive hydrophobic residues, there are no longer stretches characteristic of membrane attachment sites. Overall the molecule is only slightly hydrophilic (average = 0.1 on a scale of 3.0 to −3.4), but it contains two extended regions (residues 119–152 and 272–303) where the level of charged residues exceeds 40%.

The presence of typical recognition sites for modifying enzymes indicates that the *rel* protein might be altered post-translationally. The cAMP-dependent protein kinase phosphorylates serine residues within sequences Lys-Arg-X-X-Ser-X or Arg-Arg-X-Ser-X (24), and two such serines are found at positions 275 and 304. There are three potential glycosylation sites at positions 52, 170, and 437 with the sequence Asn-X-Thr/Ser (25). However, glycosylated proteins tend to be exported, and the absence of a long hydrophobic region makes it doubtful that the *rel* protein is either secreted or membrane-bound. Indeed, our preliminary experiments have revealed neither glycosylation nor kinase activity of the *rel* protein, but further investigation relating to both of these possibilities is necessary.

v-*rel* Is Inserted into the *env* Gene. Inspection of the sequence surrounding the ATG codon at position 1–3 reveals marked similarity to that surrounding the initiator codon for the Moloney murine leukemia virus (M-MuLV) envelope protein (26). In M-MuLV the sequences specifying the COOH terminus of the *pol* gene product and the NH_2 terminus of the *env* gene product overlap but are read in different frames. The envelope initiator codon is found 58 nucleotides upstream from the *pol* termination codon (TAA). This same arrangement is evident in REV-A, with a spacing of 61 nucleotides. As shown in Fig. 4*A*, the translated sequence surrounding this ATG in REV-A is very similar to that at the extreme COOH terminus of the

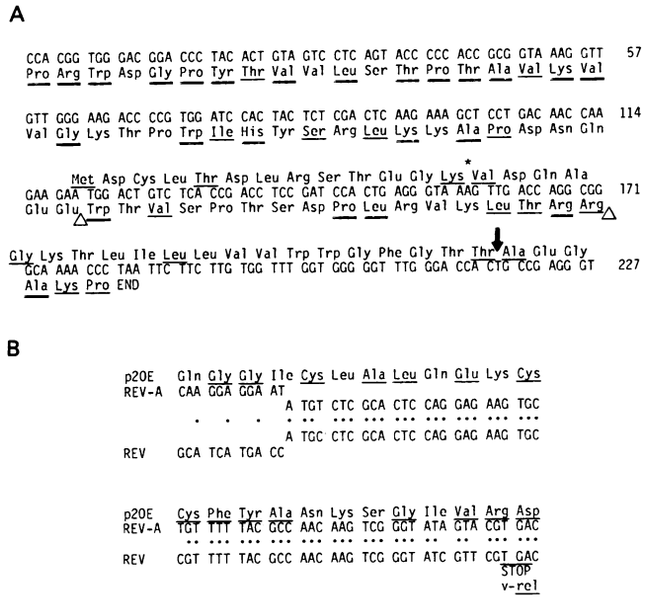
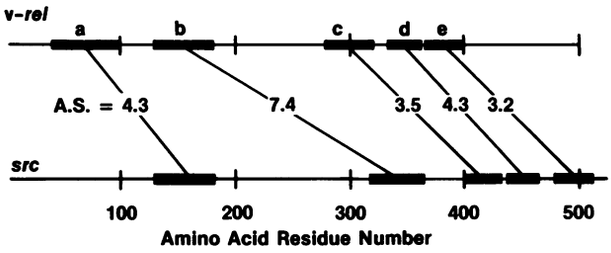


FIG. 4. v-*rel* is inserted into the *env* gene. (*A*) The translated nucleotide sequence at the REV-A *pol/env* junction. Translation in the *pol* reading frame is given below the DNA sequence and in the *env* reading frame above. Underlining indicates identity with either M-MuLV (26) or simian sarcoma virus (19); double underlining indicates identity with both. Triangles indicate a gap in REV-A relative to M-MuLV (position 120) or to simian sarcoma virus (position 171). In M-MuLV the envelope precursor is cleaved at a Thr-Ala linkage located 33 amino acids downstream from the initiator methionine (26, 27). REV-A also has a Thr-Ala doublet at this position (arrow). The asterisk marks the position of insertion of v-*rel* in REV. (*B*) The 3′ junction of v-*rel* with REV-A sequences. The sequence of REV from nucleotide 1,442 to 1,513 (Fig. 2 numbering system) is shown and is compared with that of REV-A. The REV-A sequence is translated in the p20E reading frame. Underlining indicates identity with M-MuLV p15E (26). Dots indicate identity between REV and REV-A.



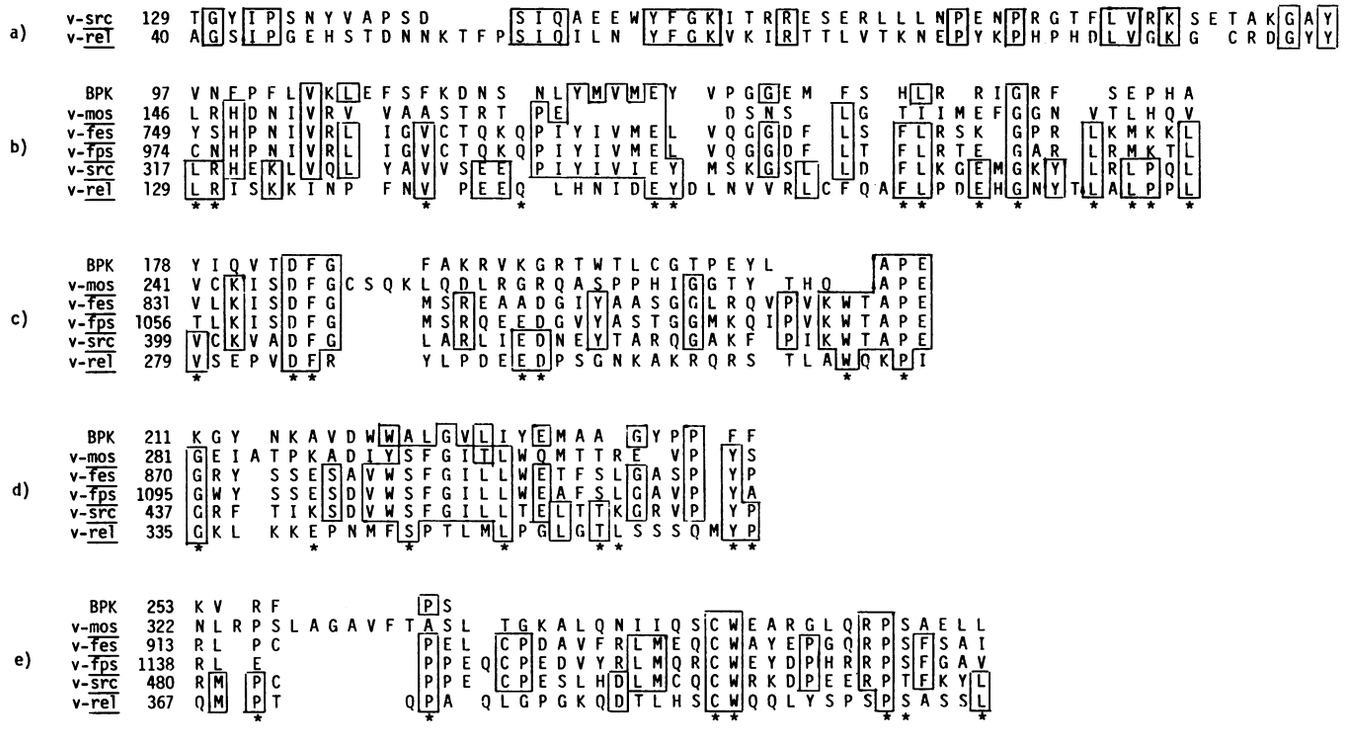


FIG. 5. Relatedness of v-*rel* and v-*src*. Segments of the predicted *rel* protein were compared with segments of v-*src* (30) by using the ALIGN program with the mutation data matrix, a matrix bias of 6, and a gap penalty of 6 (31). A.S. = alignment score. Replacing the 12 COOH-terminal amino acids of v-*src* with the 19 COOH-terminal residues of c-*src* (32) extends the length of segment e and raises its score to 4.0 SD. Alignment over the entire region containing c + d + e then yields a score of 3.6 SD (probability of chance occurrence $< 2 \times 10^{-4}$). The amino acid sequence in segments a–e is shown below, and corresponding sequences from v-*fes* (20), v-*fps* (21), v-*mos* (18) and the catalytic chain of the cAMP-dependent bovine protein kinase (33, 34) are included for comparison. Residues common to v-*rel* and v-*src* are boxed, as are residues common to v-*src*, v-*fes*, and v-*fps*. Asterisks indicate residues found in v-*rel* and at least two of the other proteins. The phosphorylated tyrosine in v-*src* is residue 416.

M-MuLV *pol* product. Farther upstream the homology between the two is even more striking (unpublished data), leaving little doubt that these are equivalent regions in the two genomes. That this ATG is REV-A's *env* initiator is revealed by inspection of the REV-A DNA sequence and comparison with that of M-MuLV. In M-MuLV a 33-amino acid signal peptide containing a highly hydrophobic region is cleaved from the envelope precursor at a Thr-Ala linkage (26, 27). In REV-A there is a Thr-Ala doublet at position 33–34, and the sequence between the methionine and this doublet contains 11 consecutive hydrophobic amino acids. Therefore, we conclude that v-*rel* has been inserted into REV-A just 37 nucleotides downstream from the start of the *env* gene and that the *rel* protein thus begins with 12 amino acids of the REV-A envelope gene product. This is exactly the same arrangement as seen with the insertion of v-*mos* into M-MuLV, except that in that case only five amino acids of the *env* gene product are present (18).

Initiation of translation of almost all of the v-*onc* gene products of the defective transforming viruses occurs at an ATG located in helper-derived sequences. Some of these products are *gag–onc* fusion proteins, whereas others such as v-*rel*, v-*mos* (18), and v-*sis* (19) contain several amino acids of *env*. A simple interpretation of these findings is that transduction of each of the c-*oncs* omitted the 5′-most region of the respective genes, including the normal initiator. It is not known whether this denotes a requirement of a helper-derived protein fragment for transforming activity or simply reflects the low frequency at which the entire c-*onc* could be inserted in a position that would allow self-initiation. Alternatively, perhaps it is only when its 5′-most region (exon?) is missing that the resultant v-*onc* has transforming ability.

env sequences are also found at the 3′ end of v-*rel*, as shown in Fig. 4*B*. In REV-A, sequences coding for p20E were easily identifiable based on amino acid sequence data for purified REV-A p20E (W. P. Tsai and S. Oroszlan, personal communication) and on the protein's similarity to M-MuLV p15E (26). In REV, p20E sequences abut the 3′ boundary of v-*rel*, but because the *rel* protein is read in the −1 frame relative to REV-A p20E, no p20E fragments will appear in it. The *rel* protein is predicted to end after incorporating 19 amino acids from the helper-derived region.

Because all frames upstream from the presumptive initiator ATG are blocked, it is likely that the *rel* protein is translated from a spliced message. In fact, a RNA of about 3 kb that con-

tains *rel* sequences has been observed in nonproducer transformed cells (unpublished data). It is also interesting to note that within v-*rel* itself there are a number of potential splice acceptor sites [e.g., $(Y)_{10}$-C-A-A-G$\downarrow$C at position 725, where Y = pyrimidine]. Thus, it is conceivable that a second protein is produced, translated from a mRNA of <3 kb and initiated at one of the several ATGs located within v-*rel*.

v-*rel* May Be Distantly Related to v-*src*. A preliminary search for similarities between the v-*rel* protein and the other reported *onc* gene proteins employed the computer program RELATE (28) and revealed a possible distant relatedness between parts of v-*rel* and v-*src*. These results were then evaluated by using the ALIGN program (28). Given a scoring matrix and a penalty for introduced gaps, ALIGN brings two sequences into optimal alignment and assigns a score to the comparison. The residues in each sequence are then repeatedly jumbled and recompared, and the so-called alignment score expresses how far (in standard deviation units) the test score is above the average score. An alignment score of 5 SD has been taken to imply evolutionary relatedness of two polypeptides and a score between 3 and 5 SD is suggestive of such relatedness (29).

The results of comparing v-*rel* and v-*src* (30) in this way are shown in Fig. 5. There are five segments of *rel*, together comprising 205 amino acids, which give alignment scores of >3 SD when compared with segments of v-*src*. The most similar coupling exhibits 18 identities out of 47 possible matches, yielding a score of 7.4 SD. The probability of these two segments achieving a score this high by chance is $<10^{-13}$. The four other pairs have scores between 3 and 5 SD (7/33, 8/33, 18/54, and 7/28 identities). Broadening the comparison, alignment over the 134-amino acid span encompassing segments c, d, and e (*rel* 279–413, *src* 399–526) gives a score of 3.0 SD (probability of chance occurrence = 10^{-3}). Thus, of the 472 (503 − 31) amino acids derived from v-*rel*, almost half are in segments that appear to be related to v-*src*. The fact that the linear arrangement of these five segments is the same in both *rel* and *src* also supports the possibility of their ancestral relatedness.

This apparent similarity between *rel* and *src* is much weaker than that between v-*src* and its other relatives. v-*src* and v-*yes* have 82% identities over a 436-amino acid span (15); v-*src* and v-*fes* or v-*fps* have about 40% identities over a 280-residue span (20, 21); v-*src* and v-*mos* show about 25% identities over the 374-amino acid length of v-*mos* (18); and v-*src* and the catalytic chain of the bovine protein kinase exhibit about 25% identities over a 166-amino acid span (33, 34). Although *src* and *rel* exhibit almost 30% identities in the segments shown in Fig. 5, homology over the region encompassing all of these segments is only about 20% of *rel*'s residues. The long areas of apparently unrelated sequence and the necessity of introducing gaps within the apparently related areas result in neither RELATE nor ALIGN assigning a significant score to this latter comparison. In contrast, the alignment score between v-*src* and the bovine protein kinase over a 166-amino acid length is about 9 (probability of chance occurrence = 10^{-19}) (34). Thus, we are unable to say whether the observed similarities between v-*rel* and members of the v-*src* family are merely fortuitous or whether they represent a distant relatedness reflecting descent from a common ancestor. Detection and isolation of the *rel* protein, rendered possible by knowledge of v-*rel*'s DNA sequence, should reveal whether there are functional similarities between the *rel* and *src* proteins.

We are very grateful to John Elser for assistance in sequence analysis, to Karen McNitt for performing the computer searches, to Jeannie Clarke for preparing the manuscript, and to Dr. I. Chen for the gift of a Charon 4A clone of REV-T. This research was sponsored by the National Cancer Institute, under Contract NO1-CO-23909 with Litton Bionetics, Inc., by National Institutes of Health Grant CA26169, and by Robert A. Welch Foundation Grant F849.

1. Hoelzer, J. D., Franklin, R. B. & Bose, H. R., Jr. (1979) *Virology* **93**, 20–30.
2. Breitman, M. L., Lai, M. M. C. & Vogt, P. K. (1980) *Virology* **100**, 450–461.
3. Gonda, M. A., Rice, N. R. & Gilden, R. V. (1980) *J. Virol.* **34**, 743–751.
4. Hu, S. S. F., Lai, M. M. C., Wong, T. C., Cohen, R. S. & Sevoian, M. (1981) *J. Virol.* **37**, 899–907.
5. Chen, I. S. Y., Mak, T. W., O'Rear, J. J. & Temin, H. M. (1981) *J. Virol.* **40**, 800–811.
6. Rice, N. R., Hiebsch, R. R., Gonda, M. A., Bose, H. R., Jr., & Gilden, R. V. (1982) *J. Virol.* **42**, 237–252.
7. Simek, S. & Rice, N. R. (1980) *J. Virol.* **33**, 320–329.
8. Wong, T. C. & Lai, M. M. C. (1981) *Virology* **111**, 289–293.
9. Chen, I. S. Y., Wilhelmsen, K. C. & Temin, H. M. (1983) *J. Virol.* **45**, 104–113.
10. Hoelzer, J. D., Lewis, R. B., Wasmuth, C. R. & Bose, H. R., Jr. (1980) *Virology* **100**, 462–474.
11. Lewis, R. B., McClure, J., Rup, B., Niesel, D. W., Garry, R. F., Hoelzer, J. D., Nazerian, K. & Bose, H. R., Jr. (1981) *Cell* **25**, 421–431.
12. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
13. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 3824–3828.
14. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
15. Kitamura, N., Kitamura, A., Toyoshima, K., Hirayama, Y. & Yoshida, M. (1982) *Nature (London)* **297**, 205–208.
16. Rushlow, K. E., Lautenberger, J. A., Papas, T. S., Baluda, M. A., Perbal, B., Chirikjian, J. G. & Reddy, E. P. (1982) *Science* **216**, 1421–1423.
17. Klempnauer, K.-H., Gonda, T. J. & Bishop, J. M. (1982) *Cell* **31**, 453–463.
18. Van Beveren, C., Galleshaw, J. A., Jonas, V., Berns, A. J. M., Doolittle, R. F., Donoghue, D. J. & Verma, I. M. (1981) *Nature (London)* **289**, 258–262.
19. Devare, S. G., Reddy, E. P., Law, J. D., Robbins, K. C. & Aaronson, S. A. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 731–735.
20. Hampe, A., Laprevotte, I., Galibert, F., Fedele, L. A. & Sherr, C. J. (1982) *Cell* **30**, 775–785.
21. Shibuya, M. & Hanafusa, H. (1982) *Cell* **30**, 787–795.
22. Alitalo, K., Bishop, J. M., Smith, D. H., Chen, E. Y., Colby, W. W. & Levinson, A. D. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 100–104.
23. Kozak, M. (1981) *Curr. Top. Microbiol. Immunol.* **93**, 81–123.
24. Krebs, E. G. & Beavo, J. A. (1979) *Annu. Rev. Biochem.* **48**, 923–959.
25. Marshall, R. D. (1974) *Biochem. Soc. Symp.* **40**, 17–26.
26. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293**, 543–548.
27. Oroszlan, S., Henderson, L. E., Copeland, T. D., Schultz, A. M. & Rabin, E. M. (1980) in *Biosynthesis, Modification, and Processing of Cellular and Viral Polyproteins*, eds. Koch, G. & Richter, D. (Academic, New York), pp. 219–232.
28. Dayhoff, M. O. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), Vol. 5, Suppl. 3, pp. 1–8.
29. Barker, W. C. & Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), Vol. 5, pp. 101–110.
30. Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* **32**, 853–869.
31. Schwartz, R. M. & Dayhoff, M. O. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), Vol. 5, Suppl. 3, pp. 353–358.
32. Takeya, T. & Hanafusa, H. (1983) *Cell* **32**, 881–890.
33. Shoji, S., Parmelee, D. C., Wade, R. D., Kumar, S., Ericsson, L. H., Walsh, K. A., Neurath, H., Long, G. L., Demaille, J. G., Fischer, E. H. & Titani, K. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 848–851.
34. Barker, W. C. & Dayhoff, M. O. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2836–2839.