Nucleotide Sequence of the Akv *env* Gene

JACK LENZ, ROBERT CROWTHER, ANTHONY STRACESKI, AND WILLIAM HASELTINE*

Sidney Farber Cancer Institute, Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115

Received 4 December 1981/Accepted 21 December 1981

The sequence of 2,191 nucleotides encoding the *env* gene of murine retrovirus Akv was determined by using a molecular clone of the Akv provirus. Deduction of the encoded amino acid sequence showed that a single open reading frame encodes a 638-amino acid precursor to gp70 and p15E. In addition, there is a typical leader sequence preceding the amino terminus of gp70. The locations of potential glycosylation sites and other structural features indicate that the entire gp70 molecule and most of p15E are located on the outer side of the membrane. Internal cleavage of the *env* precursor to generate gp70 and p15E occurs immediately adjacent to several basic amino acids at the carboxyl terminus of gp70. This cleavage generates a region of 42 uncharged, relatively hydrophobic amino acids at the amino terminus of p15E, which is located in a position analogous to the hydrophobic membrane fusion sequence of influenza virus hemagglutinin. The mature polypeptides are predicted to associate with the membrane via a region of 30 uncharged, mostly hydrophobic amino acids located near the carboxyl terminus of p15E. Distal to this membrane association region is a sequence of 35 amino acids at the carboxyl terminus of the *env* precursor, which is predicted to be located on the inner side of the membrane. By analogy to Moloney murine leukemia virus, a proteolytic cleavage in this region removes the terminal 19 amino acids, thus generating the carboxyl terminus of p15E. This leaves 15 amino acids at the carboxyl terminus of p15E on the inner side of the membrane in a position to interact with virion cores during budding. The precise location and order of the large RNase T₁-resistant oligonucleotides in the *env* region were determined and compared with those from several leukemogenic viruses of AKR origin. This permitted a determination of how the differences in the leukemogenic viruses affect the primary structure of the *env* gene products.

Murine retroviruses have two polypeptides, gp70 and p15E, which comprise the protein portion of the viral envelope (12, 31). These two peptides are synthesized as a common precursor polypeptide in the following order: NH₂-gp70-p15E-COOH (22). Maturation then proceeds by several steps. Carbohydrate is attached to several sites on gp70 (45), but no carbohydrate is added to p15E (39). Proteolytic cleavage separates the two mature peptides (12, 22, 31). At least a fraction of the mature gp70 and p15E molecules are attached to each other by disulfide linkages (39). Both gp70 and p15E are accessible to serological interactions, although more antigenic sites are defined on the larger gp70 molecule (34, 35). p15E is more difficult to dissociate from virions than gp70 (39), suggesting that it is more tightly bound to the membrane.

Several important functions are associated with viral surface proteins. These include association with the viral membrane, tissue tropism and host range (1, 10, 18, 24, 44), the capacity for binding to surface structures on cells (29),

fusion of the lipid portions of the viral and host cell membranes (16), and interaction with the viral core particles during virion assembly (3). To determine how the structures of the Akv virus *env* proteins enable them to perform these functions, we determined the nucleotide sequence of the Akv virus *env* gene. Determination of the amino acid sequences of the encoded *env* polypeptides permitted deduction of several important structural features.

Previous studies have correlated the leukemogenic potentials of viruses from AKR mice with differences in the sequences in the 3' regions of the virus genomes (4-6, 10, 18, 26, 33, 36a, 37, 44, 47). The leukemogenic potentials of several viruses of AKR origin are associated with a limited number of differences compared with non-leukemogenic Akv virus. These differences are located in the carboxyl portion of the *env* coding sequences and the adjacent U₃ region of the long terminal repeat (LTR) (4, 36a, 37). Changes in the corresponding region are also associated with the leukemogenic potential of

avian leukosis virus (42). The leukemogenicity of another class of viruses, the MCF viruses (6, 10, 18, 26, 44, 47), has been correlated with more extensive changes in the 3' regions of the genomes. A comparison of the Akv *env* sequence with the genomes of leukemogenic viruses of AKR origin allowed a determination of how these changes affect the primary structures of the *env* gene products.

MATERIALS AND METHODS

Materials. Restriction enzymes were purchased from New England Biolabs, Beverly, Mass., or Bethesda Research Laboratories, Bethesda, Md. T4 DNA ligase, *Escherichia coli* polymerase I, and the Klenow fragment of *E. coli* polymerase I were obtained from New England Biolabs. DNase I was from Worthington Diagnostics, Freehold, N.J. Salmon sperm DNA was from Sigma Chemical Co., St. Louis, Mo., and ³²P-labeled nucleotides were from New England Nuclear Corp., Boston, Mass.

Cells. λAKR-623 was grown in *E. coli* LE392. pBR322 and pAKR59 were grown in *E. coli* HB101. NIH 3T3 cells were grown in Dulbecco modified Eagle medium containing 10% calf serum.

Generation of recombinant plasmid pAKR59. DNA was prepared from a lysate of λAKR-623 grown on *E. coli* LE392. A 40-μg amount of phage DNA was digested to completion with *Pst*I and *Eco*RI. The largest resulting fragment (see Fig. 1) contained the Akv sequences. Digestion with *Eco*RI is necessary to cleave a *Pst*I fragment containing phage and mouse cellular sequences which is approximately the same size as the 8.2-kilobase pair Akv fragment. The digested fragments were resolved on a 0.8% agarose gel containing 0.1 μg of ethidium bromide per ml. The band containing the Akv provirus was excised from the gel, and the DNA was electroeluted. This DNA was recovered and ligated to *Pst*I-linearized pBR322. The ligated DNA was used to transform *E. coli* HB101 to Tet^r. Resistant colonies were screened for hybridization to Akv containing the *Pst*I fragment which had been labeled with ³²P by nick translation.

Transfection of NIH 3T3 fibroblasts. DNA was isolated from one hybridizing, Tet^r clone (pAKR59), and this DNA was used to transfect NIH 3T3 fibroblasts by the calcium phosphate-glycerol method (8). Plasmid DNA was mixed 1:20 with salmon sperm DNA and diluted to a final concentration of 40 μg/ml. CaCl₂ was then added to a final concentration of 0.125 M, and 0.5 ml of this precipitate was added to each 60-cm² plate of NIH 3T3 fibroblasts. After 4 h the cells in each plate were treated for 4 min with 1 ml of 15% glycerol in HEPES-buffered saline (pH 7.05) (8). The cultures were passaged every third day and scored for the release of reverse transcriptase-containing viral particles.

DNA sequencing. The sequence of the *env* gene was determined by using standard techniques. Briefly, a series of overlapping restriction fragments was generated by digesting plasmid DNA with restriction enzymes which generate protruding 5' ends. The ends were labeled by filling out the 3' ends, using α-³²P-labeled deoxyribonucleoside triphosphates and the Klenow fragment of Pol I (14). Labeling reactions

were performed in 20-μl volumes containing 50 mM NaCl, 6 mM Tris-hydrochloride (pH 7.4), 6 mM MgCl₂, 1 mM dithiothreitol, 20 μg of restriction enzyme-digested pAKR59 DNA, 20 μCi of each deoxyribonucleoside triphosphate, and 1 to 2 U of the Klenow fragment for 1 h at 13°C. Unlabeled deoxyribonucleoside triphosphates at concentrations of 1 mM were added for the last 5 min of the reaction. Labeled fragments were isolated by electrophoresis on 5% polyacrylamide gels. Our general strategy was to first isolate large, labeled fragments by using *Bam*HI, *Bgl*II, and *Pst*I and then cut these fragments into a series of smaller fragments with *Hin*FI, *Sau*3AI, *Sau*96I, *Ava*I, or *Ava*II. These fragments were then labeled and resolved on gels. To generate fragments containing only a single labeled end, the double-end-labeled fragments were either cut internally with another restriction enzyme or denatured. The single-end-labeled fragments were resolved on polyacrylamide gels, and their sequences were determined by the method of Maxam and Gilbert (28).

RESULTS AND DISCUSSION

Generation of recombinant plasmid pAKR59. DNA sequencing was performed by using an infectious, molecular clone of the Akv provirus in pBR322. A molecular clone of the Akv provirus in phage lambda (λAKR-623) was kindly provided by Doug Lowy (25). DNA from this clone is active in a transfection assay on NIH 3T3 cells (25). To separate the Akv sequences from the flanking cellular and phage sequences of the clone, the proviral DNA was inserted into pBR322. DNA isolated from λAKR-623 was digested with *Pst*I, which cleaves the provirus once in each LTR. The resulting large *Pst*I fragment is comprised of one complete copy of the Akv provirus containing a single LTR. However, the LTR sequence is present in two discontinuous segments, one at either end of the fragment. This *Pst*I fragment was inserted into the *Pst*I site of pBR322 and used to transform *E. coli* HB101. A restriction map of the plasmid DNA from a transformed clone, pAKR59, is shown in Fig 1. This DNA is infectious in a transfection assay on NIH 3T3 cells, although portions of both the left LTR and the right LTR are deleted (Fig. 1). The RNase T₁ fingerprint of the virus produced on transfection is identical to that of Akv (38) and includes all of the oligonucleotides characteristic of the LTR region. Apparently, functional RNA can be produced from proviruses containing such deletions.

DNA sequence of the Akv *env* gene. Using DNA from pAKR59, we determined the sequence of 2,191 nucleotides comprising the *env* gene by the techniques of Maxam and Gilbert (28). Previous studies have shown that the *env* gene of murine retroviruses is translated as a precursor polyprotein which is subsequently

cleaved to generate the final peptide products, gp70 and p15E (12, 22, 31). Analysis of the sequence presented in Fig. 2 showed that there is a single open reading frame which encodes the precursor. The termination codon for this open reading frame is located 34 to 36 nucleotides upstream from the beginning of the LTR sequence.

Localization of gp70 within the encoded sequence. To determine the position of gp70 within the precursor protein, the deduced amino acid sequence of Akv gp70 was compared with the partial amino acid sequence reported by Oroszlan and Gilden (36) for Rauscher murine leukemia virus (R-MuLV) gp70. Figure 3A shows a comparison of the homologous regions of the two polypeptides. Notably, one of the regions of homology includes the amino terminus of the R-MuLV protein. Therefore, we assigned the amino terminus of Akv gp70 to the valine residue at position 1 in Fig. 2. Confirmation of the precise amino terminus of Akv gp70 awaits direct determination by protein sequencing.

The carboxyl terminus of gp70 has not been clearly defined yet for any murine retrovirus. Since the amino terminus of Moloney murine leukemia virus (Mo-MuLV) p15E corresponds to the glutamate residue at amino acid 440 (see below), the carboxyl terminus of Akv gp70 is probably located just before this position. Termination at the arginine residue that immediately precedes the amino terminus of p15E would result in a gp70 molecule with a predicted length of 439 amino acids. It is possible that a short block of amino acids located between gp70 and p15E is removed during processing to yield the final products.

Location of p15E within the encoded sequence. Sutcliffe et al. (51) have reported the nucleotide sequence and deduced the amino acid sequence

for Mo-MuLV p15E. Figure 3B shows a comparison of the Akv and Mo-MuLV p15E sequences extending from the amino termini to the ends of the open reading frames for the precursor polypeptides. The sequences of the two p15E proteins are very similar. The Akv amino terminus corresponds to the glutamate residue at amino acid 440 in Fig. 2. The carboxyl terminus of the open reading frame extends for an additional three amino acids in Akv compared with Mo-MuLV. The carboxyl terminus of p15E in Mo-MuLV corresponds to the leucine residue 17 amino acids before the carboxyl terminus of the precursor polypeptide (A. Schultz, T. Copeland, L. Henderson, and S. Oroszlan, personal communication). Akv contains the corresponding leucine residue at amino acid 619 in Fig. 2, suggesting that 19 amino acids are proteolytically removed to generate the final p15E carboxyl terminus. Thus, the mature p15E product is predicted to be 179 amino acids long. This is considerably larger than the size predicted for it by its mobility on sodium dodecyl sulfate-polyacrylamide gels (12, 22, 31). This discrepancy may be due to large regions of completely uncharged, mostly hydrophobic amino acids (see below).

Initiator methionine and leader sequence. Membrane proteins are usually synthesized as a precursor that contains a hydrophobic leader sequence which is proteolytically removed during maturation (2, 11, 53). Figure 2 shows that the valine at position 1 of gp70 is preceded by 20 uncharged, mostly hydrophobic amino acids. Two basic residues precede the hydrophobic stretch. Such a structure is typical in leader sequences (11). Two methionine codons located 31 and 49 amino acids upstream from valine at position 1 are in phase with the *env* coding sequence. Either of these might function as the

FIG. 1. Partial restriction map of pAKR59. The single lines indicate the portions of the plasmid derived from pBR322, and the double line indicates the portion derived from Akv. The provirus is oriented such that the 5' end of the plus sense strand is at the left and the 3' end is at the right. The solid black areas represent the LTR sequences. Abbreviations for restriction enzymes: Ba, *Bam*HI; Bg, *Bgl*II; Bs, *Bst*EII; Kp, *Kpn*I; Ps, *Pst*I; Pv, *Pvu*II; RI, *Eco*RI; Sa, *Sal*I; Sm, *Sma*I; Xb, *Xba*I; Xh, *Xho*I. The precise relative order of the *Bam*HI, *Sal*I, *Pvu*II, *Xho*I, and *Bgl*II sites located about 4 kilobase pairs from the left end of the provirus is not known. The nucleotide sequence was determined for the region from approximately 6 kilobase pairs from the left end of the inserted provirus to the *Pst*I site at its right end.

CCACACCGTCCTACTGACCACCCCGCTCTCAAGGTAGACGGCATCTCTGC
 -49 Met Asp Thr Arg Arg Pro Arg Gln Gly Ser Asp His Ala Pro
 ATG GAT ACA CGG CGC CCA CGT CAA GGC AGC GAC CAC GCC CCC
 (96)
 -35 -30 -20 -10 -1
 Asp Lys Thr Ile Met Glu Ser Thr Thr Leu Ser Lys Pro Phe Lys Asn Gln Val Asn Pro Trp Gly Pro Leu Ile Val Leu Leu Ile Leu Gly Gly Val Asn Pro
 GAT AAA ACC ATC ATG GAG AGT ACA ACG CTC TCA AAA CCC TTT AAA AAT CAG GTT AAC CCG TGG GGC CCC CTA ATT GTC CTT CTG ATT CTC GGA GGG GGC AAC CCC
 (201)
 NH₂-gp70
 1 10 CBO 20 30 35
 Val Thr Leu Gly Asn Ser Pro His Gln Val Phe Asn Leu Thr Trp Glu Val Thr Asn Gly Asp Arg Glu Thr Val Trp Ala Ile Thr Gly Asn His Pro Leu Trp
 GTT ACG TTG GGA AAC AGC CCC CAC CAG GTT TTT AAC CTC ACC TGG GAA GTG ACT AAT GGA GAC CGA GAA ACG GTG TGG GCA ATA ACC GGC AAT CAC CCT CTG TGG
 (306)
 36 40 50 60 70
 Thr Trp Trp Pro Asp Leu Thr Pro Asp Leu Cys Met Leu Ala Leu His Gly Pro Ser Tyr Trp Gly Leu Glu Tyr Arg Ala Pro Phe Ser Pro Pro Pro Gly Pro
 ACT TGG TGG CCT GAC CTC ACA CCA GAT CTC TGT ATG TTG GCC CTC CAG GGG CCG TCC TAT TGG GGC CTA GAA TAT CGG GCT CCT TTT TCT CCT CCC CCG GGG CCC
 (411)
 71 80 90 100 105
 Pro Cys Cys Ser Gly Ser Asp Ser Thr Pro Gly Cys Ser Asp Cys Glu Glu Pro Leu Thr Ser Tyr Thr Pro Arg Cys Asn Thr Ala Trp Asn Arg Leu
 CCC TGC TGT TCA GGA AGC AGC GAC TCC ACG CCA GGC TGT TCC AGA GAT TGT GAG GAG CCC CTG ACT TCA TAT ACT CCC CGG TGC AAT ACG GCC TGG AAC AGA CTT
 (516)
 106 110 120 130 140
 Lys Leu Ser Lys Val Thr His Ala His Asn Gly Gly Phe Tyr Val Cys Pro Gly Pro His Arg Pro Arg Trp Ala Arg Ser Cys Gly Gly Pro Glu Ser Phe Tyr
 AAG TTA TCT AAA GTG ACA CAT GCA CAC AAT GGA GGA TTC TAT GTC TGC CCC GGG CCA CAT CGC CCC CGG TGG GCC CGG TCA TGT GGT GGT GAA TCC TTC TAT
 (621)
 141 150 160 170 175
 Cys Ala Ser Trp Gly Cys Glu Thr Thr Gly Arg Ala Ser Trp Lys Pro Ser Ser Ser Ser Asp Tyr Ile Thr Val Ser Asn Asn Leu Thr Ser Asp Gln Ala Thr
 TGT GCC TCT TGG GGC TGC GAA ACC ACA GGC CGA GCA TCC TGG AAA CCA TCC TCG TCC TGG GAC TAC ATC ACA GTA AGC AAC AAT CTA ACC TCA CAG CAG GCA ACC
 (726)
 176 180 190 200 210
 Pro Val Cys Lys Gly Asn Glu Trp Cys Asn Ser Leu Thr Ile Arg Phe Thr Ser Phe Gly Lys Gln Ala Thr Ser Trp Val Thr Gly His Trp Trp Gly Leu Arg
 CCA GTA GTC AAA GGT AAT GAG TGG TGC AAC TCC TTA ACT ATC CGG TTC ACG AGC TTT GGA AAA CAG GCC ACC TCC TGG GTC ACA GGC CAT TGG TGG GGA TTG CGC
 (831)
 211 220 230 240 245
 Leu Tyr Val Ser Gly His Asp Pro Gly Leu Ile Phe Gly Ile Arg Leu Lys Ile Thr Asp Ser Gly Pro Arg Val Pro Ile Gly Pro Asn Pro Val Leu Ser Asp
 CTA TAC GTC TCT GGA CAT GAC CCA GGG CTC ATC TTT GGG ATC CGA CTT AAA ATT ACA GAC TCG GGG CCC CGG GTC CCA ATA GGG CCA AAC CCC CTC TTG TCA GAC
 (936)
 246 250 260 CBO 270 280
 Arg Arg Pro Pro Ser Arg Pro Arg Pro Thr Arg Ser Pro Pro Pro Ser Asn Ser Thr Pro Thr Glu Thr Pro Leu Thr Leu Pro Glu Pro Pro Ala Gly Val
 CGA CGA CCA CCT TCC CGG CCT AGA CCC ACC AGA TCT CCC CGG CCT TCA AAC TCC ACC CCA ACC GAG ACA CCC CTC ACC CTC CCC GAA CCC CCG CCA GCG GGA GTC
 (1041)
 281 290 300 310 315
 Glu Asn Arg Leu Leu Asn Leu Val Lys Gly Ala Tyr Gln Ala Leu Asn Leu Thr Ser Pro Asp Lys Thr Gln Glu Cys Trp Leu Cys Leu Val Ser Gly Pro Thr
 GAA AAC CCA TTG TTA AAT CTA GTA AAA GGA GCC TAC CAA GCC CTC AAC CTC ACC AGT CCT GAT AAA ACC CAA GAG TGC TGG TTA TGC CTA GTA TCG GGA CCC CCA
 (1146)
 316 320 330 CBO 340 350
 Tyr Tyr Glu Gly Val Ala Val Leu Gly Thr Tyr Ser Asn His Thr Ser Ser Ala Pro Ala Asn Cys Ser Val Ala Ser Gln His Lys Leu Thr Leu Ser Ser Glu Val Thr
 TAC TAC GAG GGG GTT GCC GTC CTA GGT ACC TAC CAC CAT ACT TCT GCC CCA GCT AAC TGC TCT GTG GCC TCT CAA CAC AAA TTG ACC TTT TCC GAA GTG ACC
 (1251)
 351 360 370 CBO 380 385
 Gly Gln Gly Leu Cys Ile Gly Ala Val Pro Lys Thr His Gln Val Leu Cys Asn Thr Thr Gln Lys Thr Ser Asp Gly Ser Tyr Tyr Leu Ala Ala Pro Thr Gly
 GGA CAG GGA CTC TGC ATA GGA GCG GTC CCT AAA ACC CAT CAA GTC TTG TGT AAT ACC ACC CAA AGC ACA AGC GAT GGG TCC TAT TTG GCC GCT CCC ACA GGA
 (1356)
 386 390 400 410 420
 Thr Thr Trp Ala Cys Ser Thr Gly Leu Thr Pro Cys Ile Ser Thr Thr Ile Leu Asp Leu Thr Thr Asp Tyr Cys Val Leu Val Glu Leu Trp Pro Arg Val Thr
 ACT ACC TGG GCT TGT AGT ACT GGA CTT ACT CCC TGT ATC TCA ACC ACC ATA CTT GAC CTC ACC ACC GAT TAC TGT GTC CTG GTC GAG CTT TGG CCA AGG GTG ACC
 (1461)
 421 430 440 NH₂-p16E 450 455
 Tyr His Ser Pro Ser Tyr Val Tyr His Gln Phe Glu Arg Arg Ala Lys Tyr Lys Arg Glu Pro Val Ser Leu Thr Leu Ala Leu Leu Gly Gly Leu Thr Met
 TAC CAT TCC CCT AGT TAT GTT TAC CAC CAA TTT GAA AGA CGA GCC AAA TAT AAA AGA GAA CCC GTC TCA CTA ACT CTG GCC CTA CTA TTA GGA GGA CTC ACT ATG
 (1566)
 456 460 470 480 490
 Gly Gly Ile Ala Ala Gly Val Gly Thr Gly Thr Thr Ala Leu Val Ala Thr Gln Gln Phe Gln Gln Leu Gln Ala Ala Met His Asp Asp Leu Lys Glu Val Glu
 GGC GGA ATT GCC GCT GGA GTG GGA ACA GGG ACT ACC GCC CTA GTG GCC ACT CAG CAG TTC CAA CAA CTC CAG GCT GCT ATG CAC GAT GAC CTT AAA GAA GTT GAA
 (1671)
 491 500 510 520 525
 Lys Ser Ile Thr Asn Leu Glu Lys Ser Leu Thr Ser Leu Ser Glu Val Val Leu Gln Asn Arg Arg Gly Leu Asp Leu Leu Phe Leu Lys Glu Gly Leu Cys
 AAG TCC ATC ACT AAT CTA GAA AAA TCT TTG ACC TCC TTG TCC GAA GTA GTG TTA CAG AAT CGT AGA GGC CTA GAT CTA CTA TTC CTA AAA GAG GGA GGT TTG TGT
 (1776)
 526 530 540 550 560
 Ala Ala Leu Lys Glu Glu Cys Cys Phe Tyr Ala Asp His Thr Gly Leu Val Arg Asp Ser Met Ala Lys Leu Arg Glu Arg Leu Ser Gln Arg Gln Lys Leu Phe
 GCT GCC TTA AAA GAA GAA TGC TGT TTC TAT GCC GAC CAC ACA GGA TTG GTA CGG GAT AGC ATG GCC AAA CTT AGA GAA AGA TTG AGT CAG AGA CAA AAG CTC TTT
 (1881)
 561 570 580 590 595
 Glu Ser Gln Gln Gly Trp Phe Glu Gly Leu Phe Asn Lys Ser Pro Trp Phe Thr Thr Leu Ile Ser Thr Ile Met Gly Pro Leu Ile Ile Leu Leu Ile Leu
 GAA TCC CAA CAA GGG TGG TTT GAA GGG CTG TTT AAT AAG TCC CCT TGG TTC ACC ACC CTG ATA TCC ACC ATC ATG GGT CCC CTG ATA ATC CTC TTG TTA ATT TTA
 (1986)
 596 600 610 620 630
 Leu Phe Gly Pro Cys Ile Leu Asn Arg Leu Val Gln Phe Ile Lys Asp Arg Ile Ser Val Val Gln Ala Leu Val Leu Thr Gln Gln Tyr His Gln Leu Lys Thr
 CTC TTT GGG CCT TGT ATT CTC AAT CAC CTG GTC CAG TTT ATC AAA CAG AGG ATT TCG GTA GTG CAG GCC CTG GTT CTG ACT CAA CAA TAT CAT CAA CTT AAG ACA
 (2091)
 631 638
 Ile Glu Asp Cys Lys Ser Arg Glu END
 ATA GAA GAT TGT AAA TCA CGT GAA TAA AAGATTATTACAGTTTACAGAAAGAGGGGGAATGAAGACCCCTTCATAAGGCTTAGCCAGCTAACTGCAG
 (2191)

FIG. 3. (A) Comparison of R-MuLV gp70 and Akv gp70. Two portions of the R-MuLV gp70 sequence have been determined (36). This figure shows the Akv sequences in the two corresponding regions, one of which includes the amino terminus of gp70. Where the Akv and R-MuLV sequences differ, the difference in R-MuLV is indicated above the corresponding Akv residue. (B) Comparison of Mo-MuLV p15E and Akv p15E. The Akv sequence is shown, and the differences in the Mo-MuLV sequence are indicated above the corresponding residue in Akv. The sequence begins at the amino terminus of p15E and ends at the carboxyl terminus of the *env* precursor polypeptide. The presumed carboxyl terminus of p15E is indicated. Long regions of at least 20 amino acids containing no charged amino acids are underlined.

translational initiator. The *env* gene products of murine retroviruses are translated from spliced mRNAs (13, 46). No matches for the consensus 3' splice sequence (50) are located within the sequence shown in Fig. 2 preceding the amino terminus of gp70.

Mechanism of association of the *env* proteins with the membrane. Figure 4B presents a model in which the *env* proteins are anchored to the

membrane via a hydrophobic sequence near the carboxyl terminus of p15E. In this structure, most of p15E and all of gp70 are located outside the lipid portion of the membrane. A short hydrophilic region at the immediate carboxyl terminus of p15E is located on the inner side of the membrane. The association of gp70 with the membrane occurs via interactions with p15E. These include disulfide linkages (39), as well as other interactions, since only a fraction of the gp70 molecules may be linked to p15E via disulfide bridges (39).

The structure proposed in Fig. 4B is based on several general observations regarding viral surface proteins, as well as specific data regarding the murine leukemia virus *env* proteins. One common structural feature of membrane-associated proteins is that they contain at least one completely uncharged, mostly hydrophobic region that anchors the protein to the lipid portion of the membrane (11). This region must be at least approximately 20 amino acids long to traverse the membrane in an α -helical configuration (11). A structure of this type is present near the carboxyl terminus of p15E (Fig. 2, positions 573 through 603). Immediately after this uncharged region is a basic arginine residue. The model presented in Fig. 4B predicts that this arginine residue is located on the inner side of the lipid portion of the membrane. This type of structure consisting of a hydrophobic membrane-traversing region followed by a basic amino acid on the inner side of the membrane is typical of both cellular and viral membrane-associated proteins (7, 15, 37, 40, 43, 52).

Several other features of gp70 and p15E are consistent with the structure shown in Fig. 4B. The functions of gp70 in binding to target cells and determining viral host range (1, 10, 18, 24, 29, 44) support the location of gp70 outside the lipid portion of the membrane. Most of the amino acids in p15E are predicted to lie outside the membrane. This prediction is supported by the observation that several antisera to p15E react with intact virions and virus-producing cells (34, 39). In addition, three cysteine residues in p15E are located on the outer side of the membrane. This permits at least one of them to form disulfide links with gp70. Another interesting feature of this model is that the only amino acids located on the inner side of the membrane

FIG. 2. Nucleotide sequence and deduced amino acid sequence determined for the Akv *env* gene. The nucleotide sequence is shown from an *Ava*I site on the 5' side of the amino terminus of the *env* gene to the *Pst*I site 39 nucleotides after the beginning of the LTR sequence. The numbers above the sequence refer to the numbers of amino acids from the predicted amino terminus of gp70, whereas the numbers on the right of each line refer to the number of nucleotides from the beginning of the entire sequence presented. The positions of the large, RNase T₁-resistant oligonucleotides (T1) from this region and several organizational features of the sequence are indicated.

FIG. 4. Localization of structural features of the Akv *env* proteins. (A) Linear representation of the *env* precursor from the amino terminus of gp70 to the carboxyl terminus of the precursor. The amino terminus of p15E is indicated by an arrow, as is the position of the boundary of the LTR. Potential glycosylation sites (Asn-X-Thr or Ser) are indicated by CHO. Regions of uncharged, hydrophobic amino acids are indicated by wavy lines. Three of these regions are present, one near the carboxyl terminus of gp70, one at the amino terminus of p15E, and one near the carboxyl terminus of p15E. (B) Figurative representation of the potential structure of gp70 and p15E relative to the lipid portion of the membrane. Four potential membrane traversing regions exist. These are the leader signal sequence at the amino terminus and the three regions described above. The latter three are depicted by short loops. According to this structure, only the hydrophobic region near the carboxyl terminus of p15E is located in the lipid bilayer in mature peptides. The relatively hydrophilic regions which form most of the peptides are indicated by the curved lines. The arrows indicate the amino terminus of gp70 after removal of the leader sequence and the amino terminus of p15E. No implication of secondary or higher order structure is intended, except that the three potential membrane traversal regions are sufficiently long to be capable of crossing the membrane in a helical configuration.

are at the extreme carboxyl terminus of p15E. Therefore, any interactions with the virion core particles (3) most likely occur through these residues.

Two other extensive regions of uncharged, mostly hydrophobic amino acids exist in the *env* gene precursor. These could also be located in the lipid portion of the membrane. One is located near the carboxyl terminus of gp70, and the other is located at the amino terminus of p15E (Fig. 2, positions 376 through 403 and 441 through 482). However, several observations argue that the amino acids in these regions lie on the outer side of the membrane.

The amino acids in positions 399 through 407 (Fig. 2) have been demonstrated to be predictive for the phenotype of surface antigenic determinant G_{IX} (9). G_{IX}⁺ viruses, such as Akv, contain

an aspartate residue (Fig. 2, position 404) which is altered to asparagine in G_{IX}⁻ viruses. This alteration creates the sequence Asn-Leu-Thr, which is a potential glycosylation site (45). The presence of this sequence is associated with the presence of an additional carbohydrate chain in the gp70 of G_{IX}⁻ viruses (45). Glycosylation at this site might mask the G_{IX} determinant (45). If the amino acids in this region were located on the inner side of the membrane, they would not be accessible to glycosylation enzymes or be in a position to affect the G_{IX} surface determinants. Therefore, it is likely that the amino acids in this region are located on the outer side of the membrane.

The amino acid sequence in the vicinity of the cleavage site at the gp70-p15E junction also favors a location on the outer side of the mem-

brane for this region of the *env* precursor. Five of the seven amino acids immediately preceding the cleavage site are basic (Fig. 2, positions 433 through 439), a feature characteristic of sites cleaved by tryptic-like enzymes. A similar site has been described for many viral and cellular membrane proteins (7, 9, 11, 23, 30, 43, 52, 53). In each case, the cleavage occurs outside the lipid bilayer. Frequently, such incisions are followed by carboxypeptidase B activity, which excises at least a portion of the amino acids at the cleavage site (17, 21). The Akv *env* gene precursor is most likely processed by the same type of mechanism. Therefore, the amino acids in this region should be located on the outer side of the membrane.

A location outside the lipid portion of the membrane for the amino acids in positions 376 through 403 and 441 through 482 is also supported by plotting the hydrophobicity indices of these regions on the matrix described by Segrest and Feldman (49). The calculated indices for these regions are both approximately 1.6 and plot in a region of the matrix inconsistent with their being localized within the lipid bilayer. In contrast, the index for the hydrophobic region near the carboxyl terminus of p15E (Fig. 2, positions 573 through 603) is approximately 2.9, which is consistent with its location within the lipid bilayer.

Taken together, these observations strongly argue that the amino acids in positions 376 through 482 are located on the outer side of the membrane (Fig. 4).

Location of potential glycosylation sites. Carbohydrate can be attached to protein via the side chains of asparagine residues at the sites of sequences Asn-X-Thr and Asn-X-Ser (27). Previous work with G_{IX}^+ murine retroviruses has indicated that carbohydrate is attached to the viral gp70 molecules at six different sites (45). The amino acid sequence of gp70 contains seven potential sites (Fig. 2 and 4A). All of these sites are located on the outer side of the membrane in the structure shown in Fig. 4B. Most of these sites are located in the carboxyl portion of gp70 (Fig. 4A), although one is located very near the amino terminus. In addition, one potential glycosylation site is present in p15E. However, no carbohydrate is attached to p15E. (45).

Similarities to the hemagglutinin protein of influenza virus. The *env* gene products have structural features that resemble the structural features of the hemagglutinin protein of influenza virus (3, 21, 22, 39, 40, 54). The proteins perform analogous functions in the different viruses, suggesting that the similar structures may perform similar functions. Both proteins are translated as precursors which are cleaved at an internal site to yield two disulfide-linked

peptides. In each case, the amino-terminal peptide (gp70 or HA1) is located on the outer side of the membrane, whereas the carboxyl-terminal peptide (p15E or HA2) anchors the protein to the membrane via a hydrophobic region in a similar position. The amino-terminal peptide determines the host range and surface binding properties. HA2 contains a hydrophobic, amino-terminal sequence that is involved in membrane fusion activity (16, 20, 21, 40, 54). A similar function might be associated with the hydrophobic region at the amino terminus of p15E. Most of the carbohydrate attached to the hemagglutinin is located in a region closer to the membrane than the portion of the molecule that forms the receptor for the target cell (54). The carbohydrate-containing portions of gp70 might be in an analogous region, proximal to the lipid bilayer.

Changes associated with leukemogenic potential in the 3' region of viruses from AKR mice. Knowledge of the DNA sequence of the 3' region of Akv permitted an analysis of the differences in genomic structure of a family of leukemogenic viruses derived from AKR mice. Akv virus is non-leukemogenic, although it can replicate to high titers in N-tropic mice (33, 48). However, other viruses that closely resemble Akv in genomic structure exhibit potent leukemogenic activity. The genomes of several of these viruses have been analyzed by RNase T_1 fingerprinting (4, 26, 36a, 37). This technique resolves about 15% of the genomic RNA into unique, RNase T_1 -resistant oligonucleotides (38). Figure 5 shows the exact order of the 3' oligonucleotides from Akv and their locations within the *env* coding sequence.

One leukemogenic virus, SL3-3, which was isolated from a thymoma cell line of AKR origin, differs from Akv by only a few single-base changes, yet exhibits potent leukemogenic activity. The locations of the differences in the genome of SL3-3 detectable by RNase T_1 fingerprinting are shown in Fig. 5. Regions having identical sequences are also indicated. The data show that the *env* genes of SL3-3 and Akv are nearly identical. Only one oligonucleotide difference is detectable within the *env* coding sequence. This difference encodes a change from Arg-Leu to His-Met in the first two amino acids on the inner side of the membrane traversal region near the carboxyl terminus of p15E (Fig. 2, positions 604 and 605). The other detectable change occurs within the LTR sequence. Both of the changes detectable in SL3-3 compared with Akv occur in several independently isolated leukemogenic viruses of AKR origin (4, 36a, 37).

Pedersen et al., submitted for publication).

A similar analysis is shown for a leukemogenic, tissue culture-passaged isolate of the

FIG. 5. Relationship between the *env* gene structure and the RNase T₁-resistant oligonucleotides of Akv and several leukemogenic viruses derived from AKR mice. The top line is a linearized representation of the structure of the Akv *env* gene products, similar to that shown in Fig. 4A. The line beneath it shows the RNase T₁-resistant oligonucleotides of Akv. The numbering system is the same as that used previously (38). Oligonucleotides are arranged such that they are oriented underneath the corresponding portion of the *env* gene products which they encode. Vertical lines are included for reference. Four leukemogenic viruses are compared with Akv. When a virus shares an oligonucleotide with Akv, the oligonucleotide is denoted by a solid circle. When an oligonucleotide from one of the leukemogenic viruses has demonstrable intersequence homology to the Akv sequence, its position is noted beneath the corresponding part of the Akv *env* proteins. Intersequence homologies were found by using a computer program (41). Parameters were adjusted such that no loopouts were permitted and a maximum of two base changes were permitted per oligonucleotide. When a continuous block of sequences is substituted in a virus relative to Akv, a horizontal line bounded by parentheses was inserted.

Gross passage A virus (Gross A/NIH). This virus was originally derived from an AKR tumor and has been maintained by extensive animal and tissue culture passage (4, 5). Its genome contains the same changes that distinguish SL3-3 from Akv. In addition, the *env* region of Gross A/NIH differs from that of Akv by several other changes (4). These differences are related to Akv by a number of single-base changes (Fig. 5). They are localized within two regions of the *env* protein, the carboxyl terminus of gp70 and the carboxyl terminus of p15E (Fig. 6). These observations are consistent with the hypothesis that Gross A/NIH initially arose as a virus similar to SL3-3 but sustained additional point mutations upon long-term animal and tissue culture passage. It is interesting that the changes detectable in the Gross A/NIH genome are localized within two regions of the *env* gene precursor.

Several leukemogenic viruses which have been derived from AKR mice contain major substitutions in the gp70 region of the genome (18, 26, 36a, 37). One of these, SL3-2, which was isolated from a thymoma cell line of AKR origin, contains a substitution of a block of sequences which precisely spans the gp70 portion of the genome relative to Akv (Fig. 5). Nonetheless, this virus has an ecotropic host range (37). In

addition, SL3-2 contains the same single base changes that distinguish SL3-3 from Akv.

Another class of leukemogenic viruses is the MCF viruses (6, 18, 47). These differ from ecotropic viruses in that they have an extended host range in tissue culture, which correlates with sequence substitutions within the gp70 gene (10, 26, 44). A comparison of the oligonucleotides of one such isolate, MCF-247, with the Akv sequence (Fig. 5) showed that it contains two substitutions relative to Akv. One of these substitutions spans the amino two-thirds of gp70. The carboxyl one-third of gp70 is identical to that of Akv. The other substitution begins in the coding sequence near the carboxyl terminus of p15E and extends through the U₃ region of the LTR.

A comparison of the genomes of all of the leukemogenic viruses described above (Fig. 5) shows that the only region altered in all of them relative to Akv is the region extending from approximately the carboxyl terminus of p15E into the U₃ portion of the LTR sequence. A corresponding region has been implicated in the leukemogenic potential of avian leukosis virus (42). In addition to the coding function, this region plays a role in the origin of second-strand DNA synthesis during reverse transcrip-

FIG. 6. Localization of changes in the ecotropic, leukemogenic viruses relative to Akv. Intersequence homologies were determined as described in the legend to Fig. 5. (A) Changes at the carboxyl terminus of gp70 between Gross A/NIH and Akv. The Akv sequence presented shows the two long, uncharged regions near the carboxyl end of gp70 and the amino terminus of p15E. The amino-terminal glutamate residue of p15E is indicated. The positions and the amino acid changes of the Gross A/NIH oligonucleotides are also indicated. (B) Changes at the carboxyl terminus of p15E in Gross A/NIH, SL3-2, and SL3-3 relative to Akv. The Akv sequence presented shows the membrane traversal region near the carboxyl end of p15E. The positions and the amino acid changes of the Gross A/NIH oligonucleotides relative to Akv are shown. Oligonucleotide 208B is shared by Gross A/NIH, SL3-2, and SL3-3. This oligonucleotide is also found in viruses produced by a number of other thymoma cell lines established from spontaneous tumors of AKR mice (4, 36a, 37; Pedersen et al.).

tion and possibly in the control of RNA transcription (19, 32, 42). The role of these alterations in the leukemogenic potential of these viruses remains to be determined.

ACKNOWLEDGMENTS

We thank Winship Herr and Victoria Corbin for helpful discussions and for communicating unpublished data on sequences. We thank Doug Lowy for the λ AKR-623 clone and C. Van Beveren, D. Lowy, and I. Verma for communicating the LTR sequence of Akv virus before publication. We thank A. Schultz, T. Copeland, L. Hendersen, and S. Oroszlan for communicating their protein sequencing data.

J.L. is a postdoctoral fellow at the National Institutes of Health. W.H. is a recipient of an American Cancer Society Faculty Research Award. This work was supported by Public Health Service grants CA 19341 and CP87-1057 from the National Cancer Institute.

LITERATURE CITED

- Besmer, P., and D. Baltimore. 1977. Mechanism of restriction of ecotropic and xenotropic murine leukemia viruses and formation of pseudotypes between the two viruses. *J. Virol.* 21:965-973.
- Blobel, G., and B. Dobberstein. 1975. Transfer of proteins across membranes. *J. Cell Biol.* 67:835-851.
- Bolognesi, D. P., and R. C. Montelaro. 1978. Assembly of type C oncornavirus: a model. *Science* 199:183-186.
- Buchhagen, D. L., F. S. Pedersen, R. L. Crowther, and W. A. Haseltine. 1980. Most sequence differences between the genomes of Akv virus and a leukemogenic Gross A virus passaged *in vitro* are located near the 3' terminus. *Proc. Natl. Acad. Sci. U.S.A.* 77:4359-4363.
- Buchhagen, D. L., T. Pincus, O. Stutman, and E. Fleissner. 1976. Leukemogenic activity of murine type C viruses after long-term passage *in vitro*. *Int. J. Cancer* 18:835-842.
- Cloyd, M. W., J. W. Hartley, and W. P. Rowe. 1980. Lymphomagenicity of recombinant mink cell focus-inducing murine leukemia viruses. *J. Exp. Med.* 151:542-552.
- Colligan, J. E., T. J. Kindt, H. Uehara, J. Martinko, and S. G. Nathenson. 1981. Primary structure of a murine transplant antigen. *Nature (London)* 291:35-39.
- Copeland, N. G., and G. M. Cooper. 1979. Transfection by exogenous and endogenous murine retrovirus DNAs. *Cell* 16:347-356.
- Donis-Keller, H., J. Rommelaere, R. W. Ellis, and N. H. Hopkins. 1980. Nucleotide sequences associated with differences in electrophoretic mobility of envelope glycoprotein gp70 and with G_{1X} antigen phenotype of certain murine leukemia viruses. *Proc. Natl. Acad. Sci. U.S.A.* 77:1642-1645.
- Elder, J. H., J. W. Gautsch, F. C. Jensen, R. A. Lerner, J. W. Hartley, and W. P. Rowe. 1977. Biochemical evidence that MCF murine leukemia viruses are envelope (*env*) gene recombinants. *Proc. Natl. Acad. Sci. U.S.A.* 74:4676-4680.
- Engelman, D. M., and T. A. Steitz. 1981. The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. *Cell* 23:411-422.
- Famulari, N. G., D. L. Buchhagen, H. D. Klenk, and E. Fleissner. 1977. Presence of murine leukemia virus envelope proteins gp70 and p15(E) in a common polyprotein of infected cells. *J. Virol.* 20:501-508.
- Fan, H., and I. M. Verma. 1978. Size analysis and relationship of murine leukemia virus-specific mRNA's: evidence for transposition of sequences during synthesis and processing of subgenomic mRNA. *J. Virol.* 26:468-478.
- Friedman, T., P. La Porte, and A. Estey. 1978. Nucleotide sequence studies of polyoma DNA. *J. Biol. Chem.* 253:6561-6567.
- Garoff, H., A. M. Frischau, K. Simons, H. Lehrach, and H. Delius. 1980. Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. *Nature (London)* 288:236-241.
- Gething, M. J., J. M. While, and M. D. Waterfield. 1978. Purification of the fusion protein of Sendai virus: analysis of the NH₂-terminal sequence generated during precursor activation. *Proc. Natl. Acad. Sci. U.S.A.* 75:2737-2740.
- Habener, J. F., H. T. Chang, and J. T. Potts. 1977. Enzymic processing of parathyroid hormone by cell-free extracts of parathyroid glands. *Biochemistry* 16:3910-3917.
- Hartley, J. W., N. K. Wolford, L. J. Old, and W. P. Rowe. 1977. A new class of murine leukemia viruses associated with development of spontaneous lymphomas. *Proc. Natl. Acad. Sci. U.S.A.* 74:789-792.
- 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)* 290:475-480.
- Hsu, M. C., A. Scheid, and P. W. Choppin. 1981. Activation of the Sendai virus fusion protein (F) involves a conformational change with exposure of a new hydrophobic region. *J. Biol. Chem.* 256:3557-3563.
- Jou, W. M., M. Verhoeven, R. Devos, E. Saman, R. Fang, D. Huylebroeck, W. Flers, G. Threlfall, C. Barber, N. Carey, and S. Entage. 1980. Complete structure of the hemagglutinin gene from the human influenza A/Victoria/3/75 (H3N2) strain as determined from cloned DNA. *Cell* 19:683-696.
- Karshin, W. L., L. J. Arcement, R. C. Naso, and R. B. Arlinghaus. 1977. Common precursor for Rauscher leukemia virus gp69/71, p15(E), and p12(E). *J. Virol.* 23:787-798.
- Kemmler, W., D. F. Steiner, and J. Borg. 1973. Studies on conversion of proinsulin to insulin. *J. Biol. Chem.* 248:4544-4551.
- Levy, J. A. 1973. Xenotropic viruses: murine leukemia viruses associated with NIH Swiss, NZB, and other mouse strains. *Science* 182:1151-1153.
- Lowy, D. R., E. Rands, S. K. Chattopadhyay, C. Garon, and G. L. Hager. 1980. Molecular cloning of infectious, integrated murine leukemia virus DNA from infected mouse cells. *Proc. Natl. Acad. Sci. U.S.A.* 77:614-618.
- Lung, M. L., C. Hering, J. W. Hartley, W. P. Rowe, and N. Hopkins. 1980. Analysis of the genomes of mink cell focus-inducing murine type C viruses: a progress report. *Cold Spring Harbor Symp. Quant. Biol.* 44:1269-1274.
- Marshall, R. D. 1974. The nature and metabolism of the carbohydrate-protein linkages of glycoproteins. *Biochem. Soc. Symp.* 40:17-26.
- Maxam, A., and W. Gilbert. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
- McGrath, M. S., and I. R. Weissman. 1979. AKR leukemogenesis: identification and biological significance of thymic lymphoma receptors for AKR retroviruses. *Cell* 17:65-75.
- Nakanishi, S., A. Inoue, T. Kita, M. Nakamura, A. C. Y. Chang, S. N. Cohen, and S. Numa. 1979. Nucleotide sequence of cloned cDNA for bovine corticotropin- β -lipotropin precursor. *Nature (London)* 278:423-427.
- Naso, R. C., L. J. Arcement, W. L. Karshin, G. A. Janjoom, and R. B. Arlinghaus. 1976. A fucose-deficient glycoprotein precursor to Rauscher leukemia virus gp69/71. *Proc. Natl. Acad. Sci. U.S.A.* 73:2326-2330.
- Neel, B. G., W. S. Hayward, H. L. Robinson, J. Fang, and S. M. Astrin. 1981. Avian leukosis virus-induced tumors have common proviral integration sites and synthesize discrete new RNAs: oncogenesis by promoter insertion. *Cell* 23:323-334.
- Nowinski, R. C., and E. F. Hays. 1978. Oncogenicity of AKR endogenous leukemia viruses. *J. Virol.* 27:13-18.
- Nowinski, R. C., M. E. Lostrum, M. R. Tam, M. R. Stone, and W. M. Burnette. 1979. The isolation of hybrid cell lines producing monoclonal antibodies against the p15(E) protein of ecotropic murine leukemia viruses. *Virology*

- 93:111-126.
35. O'Donnell, P. V., E. Stockert, Y. Obata, A. B. De Leo, and L. J. Old. 1980. Murine-leukemia virus-related cell-surface antigens as serological markers of AKR ecotropic, xenotropic, and dualtropic viruses. Cold Spring Harbor Symp. Quant. Biol. 44:1255-1264.
 36. Oroszlan, S., and R. Gilden. 1980. Primary structure analysis of retrovirus proteins, p. 299-339. In J. Stephenson (ed.), Molecular biology of RNA tumor viruses. Academic Press, Inc., New York.
 - 36a. Pedersen, F. S., R. L. Crowther, E. F. Hays, R. C. Nowinski, and W. A. Haseltine. 1982. Structure of retroviral RNAs produced by cell lines derived from spontaneous lymphomas of AKR mice. J. Virol. 41:18-29.
 37. Pedersen, F. S., R. L. Crowther, D. Y. Tenney, A. M. Reimbold, and W. A. Haseltine. 1981. Novel leukemogenic retroviruses isolated from cell line derived from spontaneous AKR tumor. Nature (London) 292:167-170.
 38. Pedersen, F. S., and W. A. Haseltine. 1980. Analysis of the genome of an endogenous ecotropic retrovirus of the AKR strain of mice: micromethod for detailed characterization of high-molecular-weight RNA. J. Virol. 33:349-365.
 39. Pinter, A., and E. Fleissner. 1977. The presence of disulfide-linked gp70-p15(E) complexes in AKR murine leukemia virus. Virology 83:417-422.
 40. Porter, A. G., C. Barber, N. H. Carey, R. A. Hallowell, G. Threlfall, and J. S. Emtage. 1979. Complete nucleotide sequence of an influenza virus hemagglutinin gene from cloned DNA. Nature (London) 282:471-477.
 41. Queen, C. L., and L. J. Korn. 1980. Computer analysis of nucleic acids and proteins. Methods Enzymol. 65:595-609.
 42. Robinson, H. L., M. N. Pearson, D. W. DeSimone, P. N. Tschlis, and J. M. Coffin. 1980. Subgroup-E avian-leukosis-virus-associated disease in chickens. Cold Spring Harbor Symp. Quant. Biol. 44:1133-1142.
 43. Rogers, J., P. Early, C. Carter, K. Calame, M. Bond, L. Hood, and R. Wall. 1980. Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chain. Cell 20:303-312.
 44. Rommelaere, J., D. V. Faller, and N. Hopkins. 1978. Characterization and mapping of RNase T₁-resistant oligonucleotides derived from the genomes of Akv and MCF murine leukemia viruses. Proc. Natl. Acad. Sci. U.S.A. 75:495-499.
 45. Rosner, M. R., J. S. Tung, N. Hopkins, and P. W. Robbins. 1980. Relationship of G_{1x} antigen expression to the glycosylation of murine leukemia virus glycoprotein. Proc. Natl. Acad. Sci. U.S.A. 77:6420-6424.
 46. Rothenberg, E., D. J. Donoghue, and D. Baltimore. 1978. Analysis of a 5' leader sequence on murine leukemia virus 21S RNA. Heteroduplex mapping with long reverse transcriptase products. Cell 13:435-451.
 47. Rowe, W. P., M. W. Cloyd, and J. W. Hartley. 1980. Status of the association of mink cell focus-forming viruses with leukemogenesis. Cold Spring Harbor Symp. Quant. Biol. 44:1265-1268.
 48. Rowe, W. P., and T. Pincus. 1972. Quantitative studies of naturally occurring murine leukemia virus infection of AKR mice. J. Exp. Med. 135:429-436.
 49. Segrest, J. P., and J. R. Feldman. 1974. Membrane proteins: amino acid sequence and membrane penetration. J. Mol. Biol. 87:853-858.
 50. Sharp, P. 1981. Speculations on RNA splicing. Cell 23:643-646.
 51. Sutcliffe, J. G., T. M. Shinnick, I. M. Verma, and R. A. Lerner. 1980. Nucleotide sequence of Moloney leukemia virus: 3' end reveals details of replication, analogy to bacterial transposons, and an unexpected gene. Proc. Natl. Acad. Sci. U.S.A. 77:3302-3306.
 52. Tomita, M., and V. T. Marchesi. 1975. Amino-acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. Proc. Natl. Acad. Sci. U.S.A. 72:2964-2968.
 53. Wickner, W. 1979. The assembly of proteins into biological membranes: the membrane trigger hypothesis. Annu. Rev. Biochem. 48:23-45.
 54. Wilson, I. A., J. J. Skehel, and D. C. Wiley. 1981. Structure of the hemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature (London) 289:366-373.