Nucleotide sequence of the 3' half of AKV

Winship Herr, Victoria Corbin and Walter Gilbert

Department of Biochemistry and Molecular Biology, Biological Laboratories, Harvard University, Cambridge, MA 02138, USA

Received 6 May 1982; Revised 29 August 1982

ABSTRACT

We report the nucleotide sequence of the 3' half of the ecotropic murine leukemia virus AKV genome. To obtain a preliminary sequence, we developed a sequencing strategy whereby a nested set of restriction fragments is chemically modified prior to gel purification and strand scission. The sequence defines the genetic map of the 3' half of AKV and locates recombinant regions previously identified in structural analyses of MCF viruses.

INTRODUCTION

The AKR mouse strain is characterized by a complex pattern of murine leukemia virus (MuLV) expression associated with a high incidence of thymic leukemias developed at 6-9 months of age. Two major types of MuLV occur in the germ line: ecotropic MuLV which replicate preferentially in murine cells, and xenotropic MuLV which grow in cells from a variety of heterologous species (1). As is characteristic of enveloped animal viruses, the host range is defined by the membrane-bound envelope gene products, gp70 and p15E (2,3), which are cleaved from a precursor molecule, pr80^{env} (4-8).

Neonatal AKR mice express high titers of the AKR ecotropic MuLV, AKV, (9) but this virus alone does not induce leukemia. Recombinant MuLV with both ecotropic and xenotropic host ranges (mink cell focus inducing, MCF viruses) appear in the preleukemic thymus (10) and may accelerate leukemogenesis if injected into young AKR mice (11,12). However, MCF viruses isolated from other strains, notably partially congenic Akv NIH Swiss mice, in general do not induce leukemia (11).

Tryptic peptide analysis of the MCF gp70's identified characteristic xenotropic and ecotropic peptides showing that these viruses contain recombinant env products (13). RNase T₁ oligonucleotide and restriction site mapping of the MCF viral genomes have identified complex recombinant structures principally within the 3' half of the viral genome (14-16). We report here the nucleotide sequence of the 3' half of AKV which identifies

coding sequences affected by recombination in MCF viruses.

MATERIALS and METHODS

Plasmids. A molecular clone of HindIII cut, closed circular AKV DNA, in Charon 21a lambda phage vector was generously provided by E. Gilboa, S. Goff, and D. Steffen. Two subclones (pAKV1 and pAKV2) of the AKV genome (in both orientations) were constructed as follows: The 8.8 kb, HindIII AKV insert was purified from the Charon 21a phage arms, circularized and ligated at 6 ug/ml, digested with PstI, and cloned into the PstI site of pBR322. A nested set of deleted subclones of both pAKV1 and pAKV2 were constructed by the method of Frischauf et al. (17) with minor modifications. DNase I cut plasmid DNA was ligated to EcoRI linkers, digested with EcoRI, and size fractionated in low temperature agarose. The DNA from separate gel fractions was circularized by ligation and plasmids conferring tetracycline resistance selected. Each subclone contains the tet^r genes of pBR322 ligated to a size-selected portion of the AKV genome at a randomly introduced EcoRI site. Thus, we constructed deletions at intervals of 400 nucleotides. Because the sets differed in insert orientation, we could easily sequence both DNA strands.

Sequencing reactions. Guanosine, cytidine and pyrimidine specific reactions were as described by Maxam and Gilbert (18,19). We used diethylpyrocarbonate to modify purines specifically. 3' end-labeled DNA was incubated in 200 ul 50 mM sodium cacodylate pH 8.0, 1 mM EDTA, with 5 ul diethylpyrocarbonate (Eastman, stored at -20°C) for 5' at 37°C, after mixing well. After the addition of 50 ul 1.5 M NaOAc, 200 ug/ml tRNA, the DNA was precipitated twice with ethanol and treated with piperidine as described (19).

Sequencing of pooled fragments. Plasmid DNA from each deleted subclone was prepared separately from 50 mls of chloramphenicol-amplified cultures by the "boiling" method (20). Twelve pooled subclones were further purified by either ethidium bromide-caesium chloride density gradient centrifugation or Sephadex G-150 gel filtration. Pooled subclones (0.15 picomoles each) were digested with EcoRI, 3' end labeled with [alpha-³²P]dATP + [alpha-³²P]TTP (3000 Ci/mole, New England Nuclear) and the large fragment of Polymerase I (19), and digested with HindIII and PstI (pAKV1 deletions) or HindIII and BamHI (pAKV2 deletions). The pooled DNA was chemically modified as described above. The four sets of premodified fragments were resuspended in 12 ul formamide loading buffer (19), heat denatured and separated by polyacrylamide gel electrophoresis. The top 3/4 of the gels (0.8 mm thick with lanes 1.2 cms wide) were 3.5 or 4% polyacrylamide (3.3% bisacrylamide: 96.7% acrylamide) and

the bottom 1/4 of 6% polyacrylamide (to retard the smallest fragment) in 7 M Urea, 50 mM Tris borate pH 8.3, 1 mM EDTA. The capacity of these gels is about 0.5 ug/10 mm²/fragment. To separate the larger fragments (about 4kb long) the gels were electrophoresed slowly (10 v/cms).

After autoradiography, the fragments were carefully cut out of the gel and placed in 1.5 ml Eppendorf tubes. To remove urea and buffer, the gel slices were soaked in 1 ml H₂O for 60 minutes. After this, the gel slices were immersed in 300 ul 1 M piperidine and incubated at room temperature for 1-2 hours followed by 30 mins at 90°C. The piperidine was lyophilized and chemically cleaved fragments (<500-700 nucleotides) were eluted by rehydrating the gel slices for 12 - 24 hrs in 0.5 ml 0.3 M NH₄OAc. The eluted fragments were transferred to a new tube, precipitated with ethanol, resuspended with 300 ul 0.5 M NH₄OH and lyophilized twice. Samples were resuspended in 5 ul formamide gel electrophoresis loading buffer (19). The sequencing reactions were electrophoresed through thin (21), 0.4 mm, polyacrylamide (5% bisacrylamide: 95% acrylamide) gels in 7M urea, 100 mM Tris borate pH 8.3, 2 mM EDTA.

RESULTS

Figure 1 shows the two plasmid constructions (pAKV1 and pAKV2) and the fragments used to determine the nucleotide sequence of the 3' half of the AKV genome (shown in black in Fig. 1). A nested set of twelve deletions (Fig. 1a and b) was constructed from each of pAKV1 and pAKV2 by the method of Frischauf et al. (17). Rather than do individual chemical modification reactions on each deletion, we pooled each set of twelve deletions for the labeling and chemical modification reactions as described in the Materials and Methods prior to gel purification. The acidic depurination (A+G) reaction (19) causes considerable breakdown of the DNA which prevents purification of large acid treated fragments, but DNA modified by diethylpyrocarbonate is stable and generates an A>G pattern (Fig. 2A), as in RNA (22). Figure 2B shows the electrophoretic separation of pAKV1 end-labeled fragments modified by diethylpyrocarbonate. Sequencing patterns generated by this strategy of DNA sequencing are shown in Figure 2C. After deriving a preliminary sequence (Fig. 1a and b), the sequence was completed by sequencing the fragments shown in Figure 1c individually. We found that some of the deleted subclones contained unrelated sequences (dotted in Fig. 1) beginning at the inserted linker. These probably reflect artifacts which occurred during the linker ligation.

Figure 1. Structure of AKV subclones used to construct deleted subclones and map of sequenced fragments. Top: Restriction site maps of the subcloned AKV genome (bar) in pBR322 (line). The sequenced 3' region is shown in black. Middle: Restriction site map of the AKV provirus from (49). Bottom: The arrows identify the direction and extent of sequence analysis. (a) pAKV1 and (b) pAKV2 subclones sequenced as described in the text. (c) Fragments used to complete the sequence. Numbered arrows represent deleted subclones. (d) Sequence determined from an infectious clone of AKV (32). (e) AKV cDNA sequences determined previously (33). The dotted lines represent artifactual sequences resulting from the deletion constructions. The restriction sites are: B, BamHI; Bg, BglIII; H, HindIII; P, PstI; S, SacI; X, XbaI.

Figure 3 shows the nucleotide sequence of the 3' region of the AKV genome starting at the unique SacI site within *pol*. The sequence begins at a position corresponding to nucleotide 3294 of the Moloney MuLV (Mo-MuLV) genome (23), 1056 nucleotides downstream from the beginning of the Mo-MuLV *pol* reading frame. The *pol* open reading frame extends to nucleotide 2532 in Figure 3. The very large *pol* open reading frame noted in Mo-MuLV (23) appears to be conserved within AKV.

The amino acid sequence of the AKV gp70 and p15E peptides has not been analyzed. However the probable envelope gene structure can be deduced by comparison with partial amino acid sequences of the related Rauscher (R-MuLV) and Friend (F-MuLV) MuLV gp70's (24,25) and Mo-MuLV p15E molecule (26,27), and with the nucleotide sequence of Mo-MuLV (23). The N-terminal regions of R-MuLV and F-MuLV gp70's contain a common glycosylation site sequence (Asn-Ile-Thr) beginning at amino acid twelve (25). An analogous potential glycosylation site is present in the AKV sequence (Asn-Leu-Thr, nucleotides

Figure 2. Chemical modification patterns. (A) Maxam-Gilbert (19) chemical degradation pattern with an A>G pattern generated by modification with diethylpyrocarbonate. The sequence presented is from pAKV2 subclone 13. (B) Separation of pooled diethylpyrocarbonate modified pAKV1 fragments on a one meter long 3.5% polyacrylamide gel (see Materials and Methods). The shortest fragment (270 bases) is not shown. (C) Chemical cleavage patterns generated by the scheme described in the text. The sequences shown are nucleotides 200 to 300 of pAKV1 subclones 6 through 9 electrophoresed through a one meter long 5% polyacrylamide gel. The reactions are as in (A).

GluLeuAspCysGlnGlnGlyThrArgAlaLeuLeuLeuThrLeuGlyAsnLeuGlyTyrArgAlaSerAlaLysLysAlaGlnLeuCys GAGCTCGACTGCCAACAGGTTACTCGGGCCCTGTTACTAACCTAGGAAACCTCGGGGTATCGGGCTCGGCCAAGAAAGCCCACTTTGC pol →	90
GlnLysGlnValLysTyrLeuGlyTyrLeuLeuGlyGlnArgTrpLeuThrGluAlaArgLysGluThrValMetGlyGlnPro CAGAAACAGGTCAAGTATCTGGGGTATCTCTAAAGAGGGTCAGAGATGGCTGACTGAGGCCAGAAAGAGACTGTGTATGGGCAGCCT	180
ThrProLysThrProArgGlnLeuArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGluMetAlaAla ACTCCGAAGACCCCTCGAACAACTAAGGAGTTCCTAGGACCGGAGGCTTCTGTCGCTCTGGATCCCTGGGTTTGGCGAATATGGCAGCC	270
ProLeuTyrProLeuThrLysThrGlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTyrGlnGluLysGlnAlaLeuLeu CCCTTGATCTCTTACCAAAACGGGACTCTGTTCAATTGGGGCCAGACCAGCAAAAGGCTATCAAGAAATCAACAGGCCTCTCTA	360
ThrAlaProAlaLeuGlyLeuProAspLeuThrLysProPheGluLeuPheValAspGluLysGlnGlyTyrAlaLysGlyValLeuThr ATCCCCCGCCCTGGCATTCGCAGATTGTACTAAGCCCTTTGAATCTTTGTGACAGAGAAGCAGGGCTACGCCAAGGCGCTCTAAAG	450
GlnLysLeuGlyProTrpArgArgTrpLeuValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTrpProProCysLeuArgMet CAAAACATGGGACCTTGGCGTTCGGCCCTGGCCTACCTGTGCCAAAAGTAGATCCAGTGGCAGCTGGGTGGCCCCCTTGTCTACCGATG	540
ValAlaAlaAlaValLeuThrLysAspAlaGlyLysLeuThrMetGlyGlnProLeuValLysAlaAlaProHisAlaValGluAla GTAGAGCCCATGGCGTTCTGCACAAAGATGCAGCAAGCTAACTATGGGACAGCGCTAGTCACTCGGCCCCCATCGAGTACGAGAGCA	630
LeuValLysLeuProAspArgTrpLeuSerAsnAlaArgMetThrHisTrpGlnAlaMetLeuLeuAspThrAspArgValGlnPhe CTGGTCAAGCAACCCCTGACCCCTGGCTATCCACCGCCGATGACCCACTACCAAGCAATGCTCTAGACACTGACCGAGTTCAGTTTC	720
GlyProValValAlaLeuAsnProAlaThrLeuLeuProLeuProGluGluGlyAlaProHisAspCysLeuGluLysLeuAlaGluThr GACCACTGGTGGCTCAATCTTCGCCCTTACTCCCTTCCCGGAAGAAGGAGCCCTGATGATGGCTCGAGATCTTGGCTTGGCTAAAG	810
HisGlyThrArgProAspLeuThrAspGlnProIleProAspAlaAspHisThrTrpTyrThrAspGlySerSerPheLeuGlnGluGly CATGGAACAGACCGGATCTCACCGACGAGCCATCCAGACGCGACCCACCTGGTATACCGATGGGAGGAGCTTTTGTCAAGAAAGGA	900
GlnArgLysAlaGlyAlaAlaValThrThrGluThrGluValLysTrpAlaArgAlaLeuProAlaGlyThrSerAlaGlnArgAlaGlu ACCGAAAGGCTGGGCGAGCGGTCACCAAGGAGGTAATCTGGGACAGGGCATGCGCGCTAGGACATCTGCGCAAGCGGCGCA	990
LeuIleAlaLeuThrGlnAlaLeuLysMetAlaGluGlyLysArgLeuAsnValTyrThrAspSerArgTyrAlaPheAlaThrAlaHis CTGATGACCTCACTCAAGCCTTGAAGATGGCAGAGGTAAGAGGCTAAACGTTTACACTGACAGCCGATATGTCTTTCGCCACGGCCCAT	1080
IleHisGlyGluIleTyrArgArgArgGlyLeuLeuThrSerGluGlyArgGluLysAsnLysSerGluLysLeuAlaLys ATCCATGGAGATCTATAGAGGCGAGGGTGTCTAACCTCAGAGGTAGAGAAATCAAAACAGAGCCAGATCTGCGCTTACTGTAAA	1170
AlaLeuPheLeuProLysArgLeuSerIleIleHisCysLeuGlyHisGlnLysGlyAspSerAlaGluAlaArgGlyAsnArgLeuAla GCTCTTTTTCTGCCAAGAGACTCAGTATAATCACTGCCTGGGGCATCAAAAGGAGATAGTCCGAGGCTAGGGCAACCGCCTAGCA	1260
AspGlnAlaArgAlaGluAlaAlaIleLysThrProProAspThrSerThrLeuLeuLysGluAspSerThrProTyrThrProAlaTyr GACCAAGCGGCCCGGAGGACGCTATAAAGACGCTCCAGATACATCCACTCTCTTATAGAGGATCAACCCATATACGCTTGCCTAT	1350
PheHisTyrThrGluThrAspLeuLysLysLeuArgGluLeuGlyAlaThrTyrAsnGlnSerLysGlyTyrTrpValPheGlnGlyLys TTCCATTATACGTAACAGATCTAAAAAATCAGAGAGCTGGGGCCACCTATAACAGAGCAAGGATACTGGGTCTTCCAAGGCAAG	1440
ProValMetProAspGlnPheValPheGluLeuAspSerLeuHisArgLeuThrHisLeuGlyTyrGlnLysMetLysAlaLeuLeu CCGGTGATGCCGACCAATTGTATTGTAACCTGTAGACTACTCCACCGGCTACCCACCTCGGCTACCAGAAATGAAAGCACTCCTT	1530
AspArgGlyGluSerProTyrTyrMetLeuAsnArgAspLysThrLeuGlnTyrValAlaAspSerCysThrValCysAlaGlnValAsn GACAGAGGAGAAAGCCCTACTACTGCTAAACCGGGCAAAACCTCCAATATGTGGCAGATTCTGCGACGGCTCTGCGCCCAAGTAAAT	1620
AlaSerLysAlaLysIleGlyAlaGlyValArgValArgGlyHisArgProGlySerHisTrpGluIleAspPheThrGluValLysPro GCCAGCAAGCTAAATATCGGGCAGGAGTGAGAGTACGAGGACATCGGUCAGGCTCCATTTGGGAGATCGATTTTACAGAAAGTCAAGCCA	1710
GlyLeuTyrGlyTyrLysTyrLeuLeuValPheValAspThrPheSerGlyTrpValGluAlaPheProThrLysArgGluThrAlaArg GGGCTGTATGGGTACAAGTACCTCTGGTATTCTGGGACACCTTCTCTGGTTGGGTGAAGCCCTTTTCAACCAAGAGAGAAACAGCAAGA	1800
ValValSerLysLysLeuLeuGluGluIlePheProArgPheGlyMetProGlnValLeuGlySerAspAsnGlyProAlaPheThrSer GTCGTGTCGAAGAAATGCTGGAAGAAATATTCGAGATTTCGGAATGCCACAGGTATTTGGGATCTGATAACGGGCGCTGCCTTACCTCC	1890
GlnValSerGlnSerValAlaAspLeuLeuGlyIleAsp***LysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArg CAGGTAAGTCAGTCGGTGGCGGATTACTGGGATCGATTGAAATTACATGTGCTTATAGACCCAGAGCGCCCTGTGTAACTTCACTGAGAA	1980
MetAsnArgThrIleLysGluThrLeuThrLysLeuThrLeuAlaAlaGlyThrArgAspTrpValLeuLeuLeuProAlaLeuTyr ATGAATAGAACCATCAAGGAGACTCTAACTAAATTAACGCTTGCAGCTGGCACTAGAGACTGGGATCTCTACTCCCTTAGCTCTTTAC	2070
ArgAlaArgAsnThrProGlyProHisGlyLeuThrProTyrGluIleLeuTyrGlyAlaProProLeuValAsnPheHisAspPro CGAGCCGGAACATCCGGGCCCATTCGATTCGATTCGATTAAGAACTTGTACGGGGCGCCCGCCCTGTGTAACTTCACTGAGCAACCC	2160
AspMetSerGluLeuThrAsnSerProSerLeuGlnAlaHisLeuGlnAlaLeuGlnThrValGlnArgGluIleTrpLysProLeuAla GACATGTCAGAAATAACTAATAGCCCATCTTCAAGCTCACTTACAGGCCCTCCAAACGCTGACGAGAAATTTGGAACCACTGGCC	2250
GluAlaTyrArgAspGlnLeuAspGlnProValPheProHisProPheArgIleGlyAspSerValTrpValArgArgHisGlnThrLys GAGGCCATCCGGGACCAACTAGACCAACCACTGATACCACTCCCTTCCGATTTGGAGACTCCGCTGTGGTGGCGCCGACCAAGCAAA	2340
AsnLeuGluProArgTrpLysGlyProTyrThrValLeuLeuThrThrProThrAlaLeuLysValAspGlyIleSerAlaTrpIleHis AACTTAGAACCTCGCTGCAAGGAGCCTACACGCTCTACTGACCACCCCAACCGCTCTCAAGTAGAGCGCATCTCTGATGGATACAC	2430

AlaAlaHisValLysAlaAlaThrThrProProIleLysProSerTrpArgValGlnArgSerGlnAsnProLeuLysIleArgLeuThr
MetGluSerThrThrLeuSerLysProPheLysAsnGlnValAsnPro
GCGCCCCACGTCAAGGCAGCGACCCGCCGCGGATAAAACCATCATGGAGAGTACAACGCTCTCAAACCCCTTTAAAAATCAGGTTAACC 2520

28
ArgGlyAlaPro***
TrpGlyProLeuIleValLeuLeulleLeuGlyGlyValAsnProValThrLeuArgAsnSerProHisGlnValPheAsnLeuThrTrp
CGTGGGGCCCCCTAATTGCTCTCTGATCTCTCGAGGGGTCACCCCGTTACGTTGAGAAACAGCCCCCAGGCTTTTAACTCTCACCT 2610

16
GluValThrAsnGlyAspArgGluThrValTrpAlaIleThrGlyAsnHisProLeuTrpThrTrpProAspLeuThrProAspLeu
GGGAAGTGACTAATGGAGCAGAGAACCGGTGGGCAATAACCGCAATCACCTCTGTGGACTGTGGGCTGACCTACACACGATC 2700

46
CysMetLeuAlaLeuHisGlyProSerTrpTrpGlyLeuGluThrArgAlaProPheSerProProGlyProProCysCysSerGly
TCTGTATGTGGCCCTCCAGGGCGGCTCTATTGGGGCTAGAATATCGGGTCTCTTTCTCTCCCGGGGCCCCCTGCTGTTTCAG 2790

76
SerSerAspSerThrProGlyCysSerArgAspCysGluGluProLeuThrSerTrpThrProArgCysAsnThrAlaTrpAsnArgLeu
GAAGCAGCGCTCCACCGCAGGCTGTCAGAGATTGTGAGGAGCCCTGACTTCATATCTCCCGGTGCAATACGGGCTGCAACAGAC 2880

106
LysLeuSerLysValThrHisAlaHisAsnGlyGlyPheTyrValCysProGlyProHisArgProArgTrpAlaArgSerCysGlyGly
TTAAGTATCTAAAGTGACACATGCACACAATGGAAGATTCTATGTCTGCCCCGGGCCATCGCCCCGGTGGGCCGCTCATGCGGTG 2970

136
ProGluSerPheTyrCysAlaSerTrpGlyCysGluThrThrGlyArgAlaSerTrpLysProSerSerTrpAspTyrIleThrVal
GTCCAGAAATCTTCTATTGTCCTTGGCTTCGAGGCTGCGAAACACAGCGCGAGCATCTCTGGAACCATCTCTGCTGGGACTACATCAGAC 3060

166
SerAsnAsnLeuThrSerSerGlnAlaThrProValCysLysGlyAsnGluTrpCysAsnSerLeuThrIleArgPheThrSerPheGly
TAAGCAAAATCTAACCTCAGACAGGCAACCCAGTATGCAAGGTAAAGAGTGGTCAACTCCTTAACATTCGCGGTTCAGAGCTTTG 3150

196
LysGlnAlaThrSerTrpValThrHisTrpTrpGlyLeuArgLeuTyrValSerGlyHisAspProGlyLeuIlePheGlyIleArg
GAAACAGCGCCACCTCTCGGTCACAGGCTTGGTGGGATGTGCGCTATACGCTCTGGACATACCCAGGGGTGCATCTTGGGATCC 3240

226
LeuLysIleThrAspSerGlyProArgValProIleGlyProAsnProValLeuSerAspArgArgProProSerArgProArgProThr
GACTTAAATTCACAGCTCGGGGCCCCGGTCCCAATAGGCGCAACCCGCTCTGTGACAGCCGAGCACCCTCTCCCGGCTAGACCCA 3330

256
ArgSerProProProSerAsnSerThrProThrGluThrProLeuThrLeuProGluProProProAlaGlyValGluAsnArgLeuLeu
CCAGATCTCCCGCGCTTCAAACTCCAGCCCAACGAGACACCCCTCTGCTCCCGCAACCCCGGACGGGGTGCATAACCCAGTTGT 3420

286
AsnLeuValLysGlyAlaTyrGlnAlaLeuAsnLeuThrSerProAspLysThrGlnGluCysTrpLeuCysLeuValSerGlyProPro
TAAATCTAGTAAAGAGCGCTACCAAGCCCTCAACCTCAGCTCTGATAAAACCAAGAGTGTGGTTATGCTAGTATCGGAGCCCC 3510

316
TyrTrpGluGlyValAlaValLeuGlyThrThrSerAsnHisThrSerAlaProAlaAsnCysThrValAlaSerGlnHisLysLeuThr
CATACTACGAGGGGTTGCCGCTCTAGGTACTACTCCAACTACTCTGCCCCAGCTAACTGCTCTGTGGCTCTCAACACAATTGA 3600

346
LeuSerGluValThrGlyGlnGlyLysIleGlyAlaValProLysThrHisGlnValLeuCysAsnThrThrGlnLysThrSerAsp
CCTGTGCGAAGTGACCGGACAGGACTCTGATAGGAGCGGTCCCTAAACCCATCAAGCTCTGTGTAATACCCACCAAGACAGCG 3690

376
GlySerTyrTyrLeuAlaAlaProThrGlyThrThrTrpAlaCysSerThrGlyLeuThrProCysIleSerThrIleLeuAspLeu
ATGGGTCTACTATTGGCGGCTCCACAGGAACCTACTGGGCTGTAGTACTGGACTACTCCCTGTATCTCAACCCACTACTTGACC 3780

406
ThrThrAspTyrCysValLeuValGluLeuTrpProArgValThrTyrHisSerProSerTyrValTyrHisGlnPheLysArgArgAla
TCACCACGTTACTGTCTGCTGCGAGGCTTGGCAAGGTTGACCTACCATCTCCCTAGTATATTTTACCACCAATTTAAAGACGAG 3870

436
LysTyrLysArgGluProValSerLeuThrLeuAlaLeuLeuGlyGlyLeuThrMetGlyGlyIleAlaAlaGlyValGlyThrGly
CCAAATATAAAGAGAACCGCTCTCACTAACTTGCCCTACTATTAGGAGGACTCACTATGGCGGAATTGCGCTGGAGTGGGAACAG 3960

476
ThrThrAlaValAlaThrGlnGlnPheGlnGlnLeuAlaAlaMetHisAspAspLeuLysGluValGluLysSerIleThrAsn
GGACTACGCCCTTAGCCTAGCAGCTCCAGCACTCCAGCACTGCCATGCACGATGACCTTAAAGAAAGTTGAAAGTCCATCAGTA 4050

506
LeuGluLysSerLeuThrSerLeuSerGluValValLeuGlnAsnArgArgGlyLeuAspLeuLeuPheLeuLysGlyGlyLeuCys
ATCTAGAAAAATCTTGACCTCTGTCCGAAGTAGTGTACAGAATCGTAGAGGCTAGATCTACTATTCTCTAAAGAGGGAGGTTGT 4140

536
AlaAlaLeuLysGluGluCysCysPheTyrAlaAspHisThrGlyLeuValArgAspSerMetAlaLysLeuArgGluArgLeuSerGln
GTGCTCCCTTAAAGAAAGATGCTGTTCTATGCGGACACACAGGATTGGTACGGGATAGCATGGCCAAAGTTAGAGAAAGATTGAGTC 4230

566
ArgGlnLysLeuPheLysSerGlnGlnGlyTrpPheGluGlyLeuPheAsnLysSerProTrpPheThrThrLeulleSerThrIleMet
AGACACAAAGCTCTTTAAATCCCAACAGGGTGGTTGAAGGGCTGTTTAAATAGTCCCTTGGTTCACACCCCTGATATCACCATCA 4320

596
GlyProLeullelleLeuLeuLeulleLeuPheGlyProCysIleLeuAsnArgLeuValGlnPheIleLysSerVal
TGGGTCCCTGATAATCTCTTGTAAATTTACTCTTGGGCTGTGATTCTCAATCGCTGGTCCAGTTTATCAAGACAGGATTTCGG 4410

626
ValGlnAlaValLeuThrGlnGlnTyrHisGlnLeuLysThrIleGluAspCysLysSerArgGlu***
TAGTGACGCGCTGTTCTGACTCAACAATATCATCACTTAAGACAATAGAGATTGTAATACAGTGAATAAAGATTATTATTCAGTT 4500

656
TACAGAAAGAGGGGGGAAAGAAAGCCCTTCATAGCGCTTAGCCAGCTAACTGCAG 4558

Figure 3. DNA sequence of the AKV envelope gene and surrounding regions. The DNA strand corresponding to the viral RNA is shown in the 5' to 3' direction from left to right. The numbering begins at the Sac I site within the *pol* open reading frame. The amino acid sequence of the *pol* and *env* reading frames is shown above the nucleotide sequence. The deduced N-terminus of each *env* peptide and the 5' end of the LTR (48) is identified. Large AKV RNase T₁ oligonucleotides identified by Rommelaere et al. (29) and sequenced by Pedersen and Haseltine (30) are underlined and numbered as in (29). Potential glycosylation site sequences are overlined and labeled CHO.

2601-2609; potential glycosylation sites are overlined in Fig. 3), which suggests the amino terminus of the AKV gp70 molecule is valine. Comparison with the Mo-MuLV sequence indicates the initiator methionine codon lies at nucleotides 2475-2477 (Fig. 3). This predicts that the precursor env protein has a 31 amino acid long 'signal' sequence which contains a hydrophobic region and partially overlaps the pol reading frame as in Mo-MuLV (23). Yet, unlike Mo-MuLV, the AKV sequence shows a large open reading frame extending 5' of the methionine codon (starting at nucleotide 2148). The C-terminal sequence of gp70 is not known.

The Mo-MuLV p15E peptide is 180 amino acids long. The N-terminus results from proteolytic cleavage after a Lys-Arg sequence between gp70 and p15E (23,26); the C-terminus results from proteolytic cleavage after an Ala-Leu sequence which clips off the small (16 amino acids) 'R' peptide (27,28). The sequence of the AKV p15E peptide is nearly identical to that of Mo-MuLV with only 12 amino acid differences and may therefore be processed similarly (as indicated in Fig. 3).

The location of large RNase T₁ oligonucleotides mapped by Rommelaere et al. (29) are underlined and identified by their numerical designation. To locate the oligonucleotides we used the sequences determined by Pedersen and Haseltine (30).

MuLV acquire mutations frequently during infection and replication (31). Although the AKV clone we sequenced shows no gross rearrangements, we have been unable to obtain infectious virus upon transfection of murine cells. Consistent with this, the pol open reading frame (nucleotides 1-2532) contains a UGA termination codon (nucleotides 1930-1932) which is not present in the genome of Mo-MuLV (23). To estimate the number of mutations which may be present in our sequence, we compared it with one we derived of the p15E coding region from an env subclone (generously provided by C. Holland) of the infectious AKV clone AKR-623 (32) and with the sequence of AKV Hae III cDNA restriction fragments (33, see Fig. 1e). Of 1100 nucleotides used in this comparison, we identified two differences (nucleotides 3861 and 4248) shown below the sequence in Figure 3. The nucleotide sequence of the entire AKV env gene from the infectious AKR-623 clone reported by Lenz et al. (34) begins at nucleotide 2366 (Fig. 3) and contains, in addition to the two differences noted above, two base changes at nucleotides 2432 and 2577, and the deletion of a T at nucleotide 2368 and a G between nucleotides 4511 and 4517. The deletion of nucleotide 2368 may be important because it causes the pol reading frame to terminate thirteen nucleotides downstream.

DISCUSSION

AKV envelope peptide structure. The AKV envelope gp70 and p15E peptides share many structural homologies with the influenza virus envelope hemagglutinin peptides HA₁ and HA₂ respectively (see refs. 35 and 36 for reviews of hemagglutinin structure). Analogous to HA₁, the gp70 molecule is highly glycosylated (the sequence reveals seven potential glycosylation sites [asn-x-ser or thr, see Fig. 3] most of which, if not all, are probably glycosylated [37]), has the viral receptor binding site (38-40), and is structurally and antigenically variable among different strains of MuLV (13,40,41). HA₁ is anchored to the virus surface by HA₂ which contains a hydrophobic transmembrane sequence. To identify hydrophobic regions within gp70 and p15E we analysed the amino acid sequences by computer, using the hydrophobicity indices of Rose and Roy (42). Consistent with the analogy to hemagglutinin, we do not identify any potential hydrophobic transmembrane sequence within gp70, whereas the p15E peptide has candidate transmembrane (residues 135 to 164) and hydrophilic cytoplasmic domains (residues 165-180). The N-terminus of p15E is a glutamic acid residue after which lies an apolar domain (residues 2-43). An analogous apolar N-terminal region of HA₂ has been implicated in cell-virus membrane fusion during infection (43-45).

AKV and Mo-MuLV sequence comparison. In Figure 4, we compare the nucleotide and amino acid sequences of AKV with the Mo-MuLV sequence (23). Nucleotide differences which cause amino acid replacements are shown above the map of pol and env coding regions, whereas silent changes are shown below. Three regions (a-c) contain one or more nucleotide deletions which were not included in the comparison. Between AKV and Mo-MuLV, 96% of the first 600 pol nucleotides are identical. This region may be related through recombination. Elsewhere, silent nucleotide changes appear to be scattered along the sequence, suggesting uniform drift. The replacement pattern identifies conserved sequences (stippled regions in Fig. 4): two regions within pol (nucleotides 1-1200 and 1681-2160), the C-terminal region of gp70 (3421-3840), and the p15E (3901-4440) coding sequence. These are regions of 80 or more amino acids in which there is 90% or greater homology. The conservation in p15E is consistent with the wide antigenic cross reactivity of p15E peptides from different retroviruses (40,46). Chien et al. (47) identified three regions of non-homology between the 3' half of the AKV and Mo-MuLV genomes by heteroduplex mapping. One of these probably represents the divergent region within pol (nucleotides 1201-1680), however, we have not been able to identify the other two regions. They may lie within the C-terminal pol and N-terminal

Figure 4. Nucleotide differences between AKV and Mo-MuLV (23). Nucleotide changes (for every 60 nucleotides) which cause amino acid replacements are shown above the genetic map of the *pol* and *env* regions sequenced. Below the map are the silent nucleotide changes. Regions with greater than 90% amino acid sequence homology are stippled in the genetic map.

gp70 sequences (nucleotides 2161-3421).

MCF structure. Preleukemic and leukemic thymuses from AKR mice express MCF MuLV which are related to AKV through recombination. Rommelaere et al. (14) identified large AKV RNase T₁ oligonucleotides which were missing from the genomes of four MCF viruses, two isolated from AKR mice (MCF 247 and MCF 13) and two from NSF mice partially congenic for the *Akv-1* (MCF V1-36) or *Akv-2* (MCF V2-34) loci of AKR mice (10,13). Both MCF 247 and MCF 13 accelerate the development of thymic leukemias if injected into young AKR mice, whereas MCF V2-34, like virtually all non-AKR MCF isolates, does not (11,12). MCF V1-36 is the only MCF virus isolated from non-AKR mice which has been shown to induce leukemia, albeit with lower frequencies than MCF 247 and MCF 13 (11).

Figure 5 shows the location of the AKV RNase T₁ oligonucleotides (29) with respect to a restriction map and the genetic map of the 3' terminal half of AKV. The AKV LTR sequence is from Van Beveren et al. (48). We show the AKV RNase T₁ oligonucleotides which are present (+) or missing (0) in the MCF viruses as identified by Rommelaere et al. (14). A missing oligonucleotide may be due to a point mutation rather than recombination, for example, in our AKV clone oligonucleotide 25 is missing (see Fig. 3). Nevertheless, regions

Figure 5. AKV RNase T_1 oligonucleotide maps of four MCF viruses. The top portion of the figure shows a restriction map (49) of the 8.8 Kb AKV provirus. The expanded region below shows a map of AKV restriction sites from the DNA sequence (Fig. 3): B, BamHI; C, ClaI; K, KpnI; H, HpaI; Pv, PvuII; S, SalI; Sm, SmaI; X, XhoI; Xb, XbaI. Maps of AKV RNase T_1 oligonucleotides (29) missing (o) or present (+) in four MCF viral genomes (14) are shown above a map of the coding regions of *pol* and the *env* peptides. The position of the one LTR RNase T_1 oligonucleotide is from (48).

which are missing several oligonucleotides probably represent recombinant nonectropic sequences.

The maps in Figure 5 indicate, as suggested by Lung et al. (15), that the oncogenic MCF viruses 247 and 13 contain multiple recombinant regions. The sequence shows they are missing RNase T_1 oligonucleotides from the gp70 and LTR sequences, and retain ecotropic N-terminal p15E oligonucleotides. Furthermore, Lung et al. (15) have shown that the oncogenic MCF viruses acquire two characteristic non-AKV RNase T_1 oligonucleotides within the LTR. The nononcogenic MCF V2-34 virus neither retains ecotropic p15E oligonucleotides nor acquires new LTR sequences (14,15). The weakly oncogenic MCF VI-36 has an *env* structure similar to MCF 247 and MCF 13 but does not acquire the non-AKV LTR oligonucleotides characteristic of the oncogenic MCF viruses (15). Consistent with the RNase T_1 oligonucleotide maps, restriction site analysis of MCF viral genomes (16) has shown that they acquire recombinant restriction sites within the gp70 sequences, and that only the oncogenic viruses retain an ecotropic XbaI site which lies within the N-terminal region of p15E (Fig. 5). Some MCF viruses also contain recombinant

sequences outside of the env-LTR region. For example MCF 13 lacks the AKV SacI site within pol and has acquired a new KpnI site in this region (16). This shows that many regions of MCF viruses may recombine.

The sequence of the AKV env region identifies the coding sequences affected by recombination in oncogenic MCF viruses: they acquire certain nonectropic gp70 sequences but retain ectropic N-terminal p15E sequences. It may be that the nucleotide sequences confer oncogenicity and not alterations in protein structure, but the structural analogy between influenza hemagglutinin and the AKV env peptides suggests that the apolar N-terminal region of p15E may be involved in virus-cell membrane fusion. Oncogenic MCF viruses (eg. MCF 247 and MCF 13) are thymotropic when injected into mice whereas nononcogenic MCF viruses are not (11,12). Perhaps it is the juxtaposition of a nonectropic gp70 receptor binding activity with an ectropic N-terminal p15E membrane fusion activity that dictates a thymus specific host range.

ACKNOWLEDGEMENTS

We thank E. Gilboa, S. Goff, and D. Steffen for the unpublished molecular clone of AKV; C. Holland for the p15E subclone and an infectivity assay; J. Lenz, R. Crowther, A. Straceski, W. Haseltine, T. Schinnick, R. Lerner, and J. G. Sutcliffe for communicating nucleotide sequences prior to publication; R. Tizard and D. Wiley for computer programs; N. Hopkins, D. Peattie, G. Ruvkun, R. Tizard and D. Wiley for helpful discussions; and B. Bowen for preparation of the manuscript. This work was supported by National Institutes of Health Grant GM09541-21.

REFERENCES

1. Levy, J.A. (1973) *Science* 182, 1151-1153.
2. Ishimoto, A., Hartley, J.W. and Rowe, W.P. (1977) *Virology* 81, 263-269.
3. Besmer, P. and Baltimore, D. (1977) *J. Virol.* 21, 965-973.
4. Famulari, N.G., Buchhagen, D.L., Klenk, H-D. and Fleissner, E. (1976) *J. Virol.* 20, 501-508.
5. Naso, R.B., Arcement, L.J., Karshin, W.L., Jamjoom, G.A. and Arlinghaus, R.B. (1976) *Proc. Nat. Acad. Sci. USA* 73, 2326-2330.
6. Shapiro, S.Z., Strand, M. and August, J.T. (1976) *J. Mol. Biol.* 107, 459-477.
7. van Zaane, D., Dekker-Michielsen, M.J.A. and Bloemers, H.P.J. (1976) *Virology* 75, 113-129.
8. Witte, O.N., Tsukamoto-Adey, A. and Weissman, I.L. (1977) *Virology* 76, 539-553.
9. Rowe, W.P. and Pincus, T. (1972) *J. Exp. Med.* 135, 429-436.
10. Hartley, J.W., Wolford, N.K., Old, L.J. and Rowe, W.P. (1977) *Proc. Nat. Acad. Sci. USA* 74, 789-792.
11. Cloyd, M.W., Hartley, J.W. and Rowe, W.P. (1980) *J. Exp. Med.* 151, 542-552.
12. O'Donnell, P.V., Stockert, E., Obata, Y. and Old, L.J. (1981) *Virology* 112, 548-563.

13. Elder, J.H., Gautch, J.W., Jensen, F.C., Lerner, R.A., Hartley, J.W. and Rowe, W.P. (1977) *Proc. Nat. Acad. Sci. USA* 74, 4676-4680.
14. Rommelaere, J., Faller, D.V. and Hopkins, N. (1978) *Proc. Nat. Acad. Sci. USA* 75, 495-499.
15. Lung, M.L., Hering, C., Hartley, J.W., Rowe, W.P. and Hopkins, N. (1980) *Cold Spring Harbor Quant. Biol. Symp.* 44, 1269-1274.
16. Chattopadhyay, S.K., Lander, M.R., Gupta, S., Rands, E. and Lowy, D.R. (1981) *Virology* 113, 465-483.
17. Frischauf, A.M., Garoff, H. and Lerach, H. (1980) *Nucleic Acids Res.* 8, 5541-5549.
18. Maxam, A.M. and Gilbert, W. (1977) *Proc. Nat. Acad. Sci. USA* 74, 560-564.
19. Maxam, A.M. and Gilbert, W. (1980) *Meth. Enz.* 65, 499-560.
20. Holmes, D.S. and Quigley, M. (1981) *Anal. Biochem.* 114, 193-197.
21. Sanger, F. and Coulson, A.R. (1978) *FEBS Letts.* 87, 107-110.
22. Peattie, D.A. (1979) *Proc. Nat. Acad. Sci. USA* 76, 1760-1764.
23. Shinnick, T.M., Lerner, R.A. and Sutcliffe, J.G. (1981) *Nature* 293, 543-548.
24. Oroszlan, S. and Gilden, R.V. (1980) In *Molecular Biology of RNA Tumor Viruses*, Stephenson, J. R., Ed., pp. 299-344, Academic Press, New York.
25. Linder, D., Stirm, S., Schneider, J., Hunsmann, G., Smythers, G. and Oroszlan, S. (1982) *J. Virol.* 42, 352-355.
26. Oroszlan, S., Henderson, L.E., Copeland, T.D., Schultz, A.M. and Rabin, E.M. (1980) In *Biosynthesis, Modification, and Processing of Cellular and Viral Polypeptides*, Koch, G. and Richter, B. Eds., pp. 219-232, Academic Press, New York.
27. Henderson, L.E., Copeland, T.D., Sowder, R. and Oroszlan, S. (1981) *Fed. Proc.* 40, 1825.
28. Green, N., Shinnick, T.M., Witte, O., Ponticelli, A., Sutcliffe, J.G. and Lerner, R.A. (1981) *Proc. Nat. Acad. Sci. USA* 78, 6023-6027.
29. Rommelaere, J., Faller, D.V. and Hopkins, N. (1977) *J. Virol.* 24, 690-694.
30. Pedersen, F.S. and Haseltine, W.A. (1980) *J. Virol.* 33, 349-365.
31. Shields, A., Witte, O.N., Rothenberg, E. and Baltimore, D. (1978) *Cell* 14, 601-609.
32. Lowy, D.R., Rands, E., Chattopadhyay, S.K., Garon, C.F. and Hager, G.L. (1980) *Proc. Nat. Acad. Sci. USA* 77, 614-618.
33. Herr, W., Schwartz, D. and Gilbert, W., manuscript submitted.
34. Lenz, J., Crowther, R., Straceski, A. and Haseltine, W. (1982) *J. Virol.* 42, 519-529.
35. Wilson, I.A., Skehel, J.J. and Wiley, D.C. (1981) *Nature* 289, 366-373.
36. Wiley, D.C., Wilson, I.A. and Skehel, J.J. (1981) *Nature* 289, 373-378.
37. Rosner, M.R., Tung, J.S., Hopkins, N. and Robbins, P.W. (1980) *Proc. Nat. Acad. Sci. USA* 77, 6420-6424.
38. DeLarco, J. and Todaro, G.J. (1976) *Cell* 8, 365-371.
39. Fowler, A.K., Twardzik, D.R., Reed, C.D., Weislow, O.S. and Hellman, A. (1977) *J. Virol.* 24, 729-735.
40. Nowinski, R.C., Pickering, R., O'Donnel, P.V., Pinter, A. and Hammerling, U. (1981) *Virology* 111, 84-92.
41. Lostrum, M.E., Stone, M.R., Tam, M., Burnette, W.N., Pinter, A. and Nowinski, R.C. (1979) *Virology* 98, 336-350.
42. Rose, G.D. and Roy, S. (1980) *Proc. Nat. Acad. Sci. USA* 77, 4643-4647.
43. Skehel, J.J. and Waterfield, M.D. (1975) *Proc. Nat. Acad. Sci. USA* 72, 93-97.
44. Gething, M.J., White, J.M. and Waterfield, M.D. (1978) *Proc. Nat. Acad. Sci. USA* 75, 2737-2740.
45. Skehel, J.J., Bayley, P.M., Brown, E.B., Martin, S.R., Waterfield, M.D., White, J.M., Wilson, I.A. and Wiley, D.C. (1982)

- Proc. Nat. Acad. Sci. USA 79, 968-972.
46. Thiel, H.-J., Broughton, E.M., Matthews, T.J., Schafer, W. and Bolognesi, D.P. (1981) *Virology* 111, 270-274.
 47. Chien, Y.-H., Verma, I. M., Shih, T. Y., Scolnick, E. M. and Davidson, N. (1978) *J. Virol.* 28, 352-360.
 48. van Beveren, C., Rands, E., Chattopadhyay, S.K., Lowy, D.R. and Verma, I.M. (1982) *J. Virol.* 41, 542-556.
 49. Steffen, D.L., Bird, S. and Weinberg, R.A. (1980) *J. Virol.* 35, 824-835.