Nucleotide Sequence of the *env*-Specific Segment of NFS-Th-1 Xenotropic Murine Leukemia Virus

ROY REPASKE,* RAYMOND R. O'NEILL, ARIFA S. KHAN, AND MALCOLM A. MARTIN

Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20205

Received 22 October 1982/Accepted 6 January 1983

The sequence of 863 contiguous nucleotides encompassing portions of the *pol* and *env* genes of NFS-Th-1 xenotropic proviral DNA was determined. This region of the xenotropic murine leukemia virus genome contains an *env*-specific segment that hybridizes exclusively to xenotropic and mink cell focus-forming but not to ecotropic proviral DNAs (C. E. Buckler et al., J. Virol. **41**:228–236, 1982). The unique xenotropic *env* segment contained several characteristic deletions and insertions relative to the analogous region in AKR and Moloney ecotropic murine leukemia viruses. Portions of an endogenous *env* segment cloned from a BALB/c mouse embryo gene library that had a restriction map and hybridization properties typical of xenotropic viruses (A. S. Khan et al., J. Virol. **44**:625–636, 1982) were also sequenced. The sequence of the endogenous *env* gene was very similar to the comparable region of the NFS-Th-1 xenotropic virus containing the characteristic deletions and insertions previously observed and could represent a segment of an endogenous xenotropic virus.

The chromosomal DNA of inbred mice contains multiple copies of murine leukemia virus (MuLV)-reactive sequences (8, 27, 28). Some of these type C proviral DNAs encode infectious ecotropic or xenotropic MuLVs (24), some are expressed only in the form of viral antigens (10, 20, 29), some may contribute portions of their envelope (env) genes during the formation of mink cell focus-forming (MCF) MuLVs (6, 16), and others contain large deletions and may not be expressed at all (16, 22). Evaluation of the molecular organization of these different endogenous MuLV proviruses and their integration sites in cellular DNA by nucleic acid hybridization techniques has been hampered by the extensive cross-reactivity of different MuLV types. To overcome some of these problems, we constructed a recombinant plasmid (pEcenv), previously designated pEc-B4, that contains sequences specific for the env gene of ecotropic MuLVs (5). This cloned 545-base-pair (bp) BglII-BamHI env segment maps 221 bp from the 5' terminus of the envelope gene of the Akv MuLV and partially overlaps a 400 bp SmaI fragment that also specifically anneals to ecotropic proviral DNAs (7). Radiolabeled pEcenv DNA has been used to quantitate the number of ecotropic proviruses in different inbred mouse strains and to evaluate the stability of ecotropic proviral DNA in AKR sublines (4, 5). More recently, we described the molecular cloning of an env-specific DNA fragment (pX_{env}) derived from NFS-Th-1 xenotropic MuLV that hybridizes to xenotropic and MCF proviruses but not to ecotropic proviral DNA (3, 16). Restriction enzyme mapping studies and hybridization experiments indicate that the *env*-specific segments present in pEc_{env} and pX_{env} DNAs are located in analogous regions of the *env* genes from which they were derived.

Since the structure of the envelope glycoprotein plays a major role in determining the tissue tropism and host range of MuLVs (1, 9, 18), we determined the nucleotide and deduced amino acid sequences of the xenotropic env-specific segment and compared them with the published sequences for Moloney (Mo) and Akv MuLVs. Our analysis indicated that although portions of the xenotropic *env*-specific segment were similar to the analogous region of ecotropic proviral DNA, this part of the xenotropic env gene contained multiple deletions and insertions relative to the ecotropic env DNA that explain its unique hybridization properties and could account for the biological characteristics of xenotropic MuLVs. The nucleotide sequence of a portion of the env region of an endogenous BALB/c MuLV provirus was also determined. It contained all of the deletions and insertions characteristic of the NFS-Th-1 xenotropic env gene and therefore could represent a portion of a potentially infectious xenotropic provirus in BALB/c chromosomal DNA.

(This work was performed in partial fulfilment



FIG. 1. Diagrammatic representation of NFS-Th-1 xenotropic and AKR ecotropic MuLV proviruses, showing the location of xenotropic and ecotropic *env*-specific segments. The construction and characterization of recombinant plasmids containing unique portions of the ecotropic and xenotropic *env* regions have been previously reported (3, 5). The strategy used for sequencing parts of the *pol* and *env* genes of NFS-Th-1 xenotropic proviral DNA is shown within dashed lines at the bottom. Abbreviations: A, *AvaII*; Acc, *AccI*; B, *BamHI*; Bg, *BgIII*; H, *HpaI*; R, *EcoRI*; S, *SmaI*.

of the requirements for the Ph.D. degree in genetics at George Washington University, Washington, D.C., by R.R.O.'N.)

MATERIALS AND METHODS

Cloned retroviral DNAs used for nucleotide sequencing studies. Two recombinant plasmids that contained segments of NFS-Th-1 xenotropic proviral DNA were used for sequencing the xenotropic env-specific region; their construction has been previously described (3). One of the plasmids (pXen_{5'-6.7}) contained an 8.9kilobase (kb) DNA insert that encompassed 6.7 kb of the 5' end of the xenotropic provirus and 2.2 kb of flanking cellular DNA. The second plasmid (pX_{env}), derived from the first, contained the 455-bp Bg/II-EcoRI env-specific segment of NFS-Th-1 xenotropic proviral DNA. The sequence of a portion of an endogenous MuLV env gene was determined from a 0.6-kb DNA segment contained within clone B-77 isolated from a BALB/c embryo gene library (16). Like other xenotropic MuLVs, the B-77 envelope segment hybridized to the xenotropic env-specific probe and not to the unique ecotropic env fragment.

DNA sequencing. DNA was sequenced by the partial degradation method of Maxam and Gilbert (19). Nucleotide sequence analyses were made with the computer program of Queen and Korn (21).

Reagents. Restriction enzymes were purchased from New England Biolabs, Beverly, Mass., Boehringer Mannheim, Indianapolis, Ind., or Bethesda Research Laboratories, Bethesda, Md., and used as described by the suppliers. T4 polynucleotide kinase and calf intestine alkaline phosphatase were obtained from PL Biochemicals, Inc., Milwaukee, Wis. $[\gamma^{-32}P]ATP$ (3,000 mCi/mmol) was purchased from Amersham Corp. Arlington Heights, Ill.

RESULTS AND DISCUSSION

Alignment of AKR ecotropic and NFS xenotropic proviral DNAs. Because of the extensive polynucleotide sequence homology involving large portions of the genomes of different types of MuLV, we constructed recombinant plasmids for use in hybridization experiments that contained segments specific for ecotropic and xenotropic env regions (3, 5). The ecotropic env segment consisted of the 545-bp BelII-BamHI fragment, which maps 6.4 to 7.0 kb from the 5' terminus of AKR ecotropic proviral DNA (Fig. 1). The 455-bp BgIII-EcoRI xenotropic env-specific segment maps to an analogous region (6.3 to 6.7 kb) of the NFS-Th-1 xenotropic provirus (3). Both of these segments have been used as hybridization probes to evaluate the molecular organization of endogenous MuLV proviral DNAs (4, 13). In view of the importance of the env gene in determining host range, the unique env segment of xenotropic proviral DNA was sequenced and compared with analogous regions of Mo and AKR proviruses.

Since we had previously ascertained that the xenotropic *env*-specific fragment did not hybridize to ecotropic proviral DNA under stringent hybridization conditions (3), it become important to establish precisely the region of heterolo-



FIG. 2. Comparison of *pol* gene sequences in NFS-Th-1 xenotropic (top line) and Mo MuLV proviral (bottom line) DNAs. The sequence of 351 nucleotides extending in the 3' direction from the *HpaI* site located approximately 5.9 kb from the 5' terminus of NFS-Th-1 xenotropic proviral DNA (see Fig. 1) was determined and aligned with the analogous region of Mo MuLV. Homologous deduced amino acids are shaded, and the *env* splice acceptor sequence is indicated. The asterisk marks the position of the initiator codon of the *env* precursor polypeptide, which is translated in a different reading frame (see Fig. 3). The numbers above the Mo MuLV sequence are identical to those used to Shinnick et al. (26).

gy and to determine whether the two *env* genes had any polynucleotide sequence homology. The strategy for sequencing the cloned xenotropic proviral DNA is shown in Fig. 1. Nucleotide sequencing was initiated at the *HpaI* site (5.9 kb) located in the *pol* region of the NSF-Th-1 xenotropic MuLV provirus, which is positioned approximately 360 bp from the 5' amino terminus of the *env* gene. As anticipated, this 3' portion of the xenotropic MuLV *pol* gene was 80% homologous with the analogous segment of Mo MuLV (Fig. 2). Nonhomologous nucleotides frequently occurred as third-base changes, which resulted in 84% homology for the deduced amino acids. The most striking example of conservation involved the 3' terminus of the xenotropic and ecotropic *pol* genes, in which 16 of the 17 final amino acids and the last three codons, including the TAA terminator, were identical. In this segment of Mo MuLV (and, by analogy, NFS-Th-1 xenotropic MuLV), the *pol* and *env* genes overlap one another (26). The nearly identical structure of the two proviral DNAs within the reading frame encoding the *pol* gene product aligned the two proviruses.

Sequencing of the *env*-specific segment of NFS-Th-1 xenotropic proviral DNA. The sequence of the first 573 nucleotides of the NFS-Th-1 xenotropic *env* gene could be aligned with the initial 666 nucleotides of the analogous segments of Mo

FIG. 3. Sequence of the 5' portion of the xenotropic NFS-Th-1 env region and its alignment with the env genes of AKR and Mo ecotropic MuLVs. The sequence of 573 nucleotides extending from the 5' terminus of the env gene of NFS-Th-1 xenotropic MuLV to the EcoRI site located at 6.7 kb was determined as described in the legend to Fig. 1 and compared with the comparable regions of two ecotropic MuLVs. The ecotropic env-specific segments of Akv extends from the Bg/II site at nucleotide 6,004 to the BamHI site at 6,542; the xenotropic env-specific fragment is located between the Bg/II site at 5,901 and the EcoRI site at 6,442. Homologous deduced amino acids are shaded, and the 3' terminus of the overlapping pol gene is indicated (+). Letters A through I indicate deletions or insertions in the xenotropic env sequence relative to the ecotropic env sequences. The amino termini of AKR and Mo MuLV gp70s and potential glycosylation sites (outlined tripeptides) in all three proviral DNAs are shown. Numbers refer to nucleotide number in the Mo MuLV genome (26).

NFS-Th-1(xeno) ATG GAA	GGT TCA GC	G TTC TCA A	AA CCC C	TT AAA GA	T AAG ATT P THE ILE	AAC CCG T	9 90 9 90 99	+ CTA ATA	G T T	ATG GGG ATC TIG MET GLY
AKR(A-eco) ATG GAG	AGT ACA AC SER THR TH	G CTC TCA A	AA CCC TI	TT AAA AA HE 1000 AS	T CAG GTT N GLN VAL	AAC CCG T))) 3 00 0 0	CTA ATT :	5 T C	CTT CTG ATT CTC
HoMuLV(M-	eco)	CGT TCA AC ARG TH	58 G CTC TCA A R LEU	00 AA CCC C1	TT AAA AA	T AAG GTT N Mark Val	AAC CCG CC	G GGC CCC		CCC TTA ATT FRO LEU DLE	5360 CIT CTG ATG CIC LEU LEU MET
xeno	GTG AGG GCA GGA Val arg ala gly	GCC TCG GT Ala Ser Va	A CAA CGT G L GLN ARG A	AC AGC CO	B CT CAC CA	91 II G ATC TTC ILE	AAT GTT A	T TGG AGA	GIT ACC	AC A	CTA ATG ACA GGA LEU MET COM GLY
A-eco	GGA GGG GTC AAC GLY GLY VAL ASN	CCC GTT AC	70 G TTG GGA A R LEU GLY A	AC AGC CO	CC CAC CA	G GTT TTT	AAC CTC AC	C TGG GAA	GTG ACT	AT GGA GAC	CGA GAA ACG GTG
M-eco	AGA GGG GTC AGT ARG GLY VAL SER	ACT GCT TC THR ALA SE	70 G CCC GGC T R PRD GLY S	CC AGT CC	CT CAT CA	GTC TAT	AAT ATC AC	C TGG GAG GLU	GTA ACC	GLY ASP	5950 CGG GAG ACG GTA ARG GLU
xeno	CAA ACA GCT AAC GLN THR ALA ASN	GCC ACC TC ALA THR SE	C CTC CTG G R LEU gang g	GG ACG AT	TG ACA GA ET THR AS	ACC TTC THR PHE	CCT AAA CI	A TAT TTT TYR PHE	GAC CTG	IGT GAT TTA	C GTA VAL
A-eco	TGG GCA ATA ACC TRP ALA ILE THR	GGC AAT CA Gly ASN HI	C CCT CTG T S PRO	GG ACT TG	GG TGG RP TRP		CCT GAC CT	C ACA CCA THR PRO	91 II GAT CTC 1	GT ATG TTG MET	GCC CTC CAC GOG Ala leu his gly
M-eco	TGG GCA ACT TCT TRP ALA THR SER	GGC AAC CA Gly ASN HI	C CCT CTG T S PRO T	GG ACC TG RP TR	GG TGG RP TRP	В	CCT GAC CT	T ACC CCA Thr Pro	GAT TTA 1	IGT ATG TTA MET	6031 GCC CAC CAT GGA Ala HIS HIS GLY
xeno	GGA GAC TAC TGG GLY ASP			D			GAT GAC CC	A GAA CCC Glu pro	GAT ATT C ASP ILE	GG GAT GGT ASP GLY	Sma I TGC CGC ACT CCC CYS ARG
A-eco	CCG TCC TAT TGG PRO SER	GGC CTA GA GLY LEU GLI	A TAT CGG G U TYR ARG A	CT CCT TT La pro Ph	T TCT CC	Sma I CCC CCG PRO PRO	GGG CCC CC GLY PRO	C TGC TGT CYS CYS	TCA G	GA AGC AGC Ser Ser	GAC TCC ACG CCA
M-eco	CCA TCT TAT TGG Pro ser	GGG CTA GAN Gly Leu Gli	A TAT CAA T J TYR GLN S	CC CCT TT Er pro ph	T TCT TCT IE SER SEI	Sma I CCC CCG PRO PRO	GGG CCC CC Gly pro	I IGT IGC Cys Cys	TCA GGG G SER GLY	GC AGC AGC Ser Ser	6112 CCA
xeno	GGG GGA AGA AGA	AGG ACA AG Arg thr ari	A CTG TAT G G LEU TYR A	AC SP				Е			
A-800	GGC TGT TCC AGA	GAT TGT GAD ASP CYS GLU	G GAG CCC C U GLU PRO L	TG ACT TC EU THR SE	R TAT ACT	CCC CGG	TGC AAT AC CYS ASN TH	G GCC TGG R ALA TRP	AAC AGA C ASN ARG L	TT ANG TTA EU LYS LEU	TCT AAA GTG ACA SER LYS VAL THR
M-eco	GGC TGT TCC AGA GGC TGT TCC AGA GGC TGT TCC AGA CYS SER	GAT TGT GAG ASP CYS GLU GAC TGC GAJ ASP CYS GLU	G GAG CCC C U GLU PRO L A GAA CCT T U GLU PRO L	TG ACT TC EU THR SE TA ACC TC EU THR SE	C CTC ACC	CCC CGG PRO ARG CCT CGG PRO ARG	TGC AAT AC CYS ASN TH TGC AAC AC CYS ASN TH	G GCC TGG R ALA TRP T GCC TGG R ALA TRP	AAC AGA C ASN ARG L AAC AGA C ASN ARG L	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU	TCT AAA GTG ACA SER LYS VAL THR 6202 GAC CAG ACA ACT ASP GLN THR THR
M-eco xeno	GGC TGT TCC AGA GGC TGT TCC AGA CYS SER C	GAT TGT GAG ASP CYS GLU GAC TGC GAJ ASP CYS GLU	G GAG CCC C U GLU PRO L A GAA CCT T U GLU PRO L C TAT GTT T	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG	A TAT ACT R TYR THI C CTC ACC R LEU THI T CAT ACT HIS THR	CCC CGG PRO ARG CCT CGG PRO ARG	TGC AAT AC CYS ASN TH TGC AAC AC CYS ASN TH ATA GGG ILE GLY	G GCC TGG R ALA TRP T GCC TGG R ALA TRP	AAC AGA C ASN ARG L AAC AGA C ASN ARG L Igt gga g	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCG GGA	TCT AAA GTG ACA SER LYS VAL THR 6202 GAC CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY
M-eco xeno A-eco	GGC TGT TCC AGA CYS SER GGC TGT TCC AGA CYS SER E CAT GCA CAC AAT HIS ALA HIS ASN	GAT TGT GAG ASP CYS GLI GAC TGC GAA ASP CYS GLI TTC GGGA GGA TTC GLY GLY	G GAG CCC C U GLU PRO L A GAA CCT T U GLU PRO L C TAT GTT TI	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG Smo I GC CCC GG	T CAT ACT R TYR THE R LEU THE HIS THE G CCA CAT	CCC CGG PRO ARG CCT CGG PRO ARG GTA CCA VAL	TGC AAT AC CYS ASN TH TGC AAC AC CYS ASN TH ATA GGG ILE GLY CGG TGG GC ARG TRP AL	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TCA A ARG SER	AAC AGA C ASN ARG L AAC AGA C ASN ARG L IGT GGA G	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCG GGA GG CCG GGA GLY V= II GT CCA GAA	TCT AAA GTG ACA SER LVS VAL THR GAC CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TTC TAT TGT SCP PHE
M-eco xeno A-eco M-eco	GGC TGT TCC AGA CYS SER GGC TGT TCC AGA CYS SER E CAT GCA CAC AAT HIS ALA HIS ASN CAT AAA TCA AAT HIS LYS SER ASN	GAT TGT GAL ASP CYS GLI GAC TGC GAL ASP CYS GLI TTG GGA GGA TTG GLY GLY GAG GGA TTT GLU GLY	G GAG CCC C U GLU PRO L A GAA CCT T U GLU PRO L C TAT GTT TI C TAT GTC TI	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG GC CCC GG GC CCC GG Sma I GC CCC GG	A TAT ACT R TYR THI C CTC ACC R LEU THI HIS THA G CCA CAT PRO HIS G CCC CAC	CCC CGG PRO ARG CCT CGG PRO ARG GTA CCA VAL CGC CCC ARG	TGC AAT AC CYS ASN TH TGC AAC AC CYS ASN TH ATA GGG ILE GLY CGG TGG GC ARG TRP AL CGA GAA TC ARG GLU SE	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TCA A ARG SER C AAG TCA R LYS SER	AAC AGA C ASN ARG L AAC AGA C ASN ARG L IGI GGA G IGI GGA G IGI GGG G	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCG GGA GL CCA GAC GL CCA GAC ASP	TCT AAA GTG ACA SER LYS VAL TAR GAC CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TTC TAT TGT SEP PHE SER PHE
M-eco Xeno A-eco M-eco	GGC TGT TCC AGA CYS SER GGC TGT TCC AGA CYS SER E CAT GCA CAC AAT HIS ALA HIS ASH CAT AAA TCA AAT HIS LYS SER ASH GLY LYS	GAT TGT GAG ASP CYS GLI GAC TGC GAA ASP CYS GLI TTG GGA GGA TTG GLY GLY GAG GGA TTT GLU GLY TGT GAG ACC	G GAG CCC C J GLU PRO L A GAA CCT T J GLU PRO L C TAT GTT TI C TAT GTC TI L TAT GTT TI C TAT GTT TI C ACT GGA C	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG Sma I GC CCC GG Sma I GC CCC GG	A TAT ACT R TYR THI C CTC ACC R LEU THF IT CAT ACT HIS THF G CCA CAT PRO HIS G CCC CAC C TGG AAG	CCC CGG PRO ARG CCT CGG PRO ARG CTA CCA VAL CGC CCC CGC CCC ARG CGC CCC CCA TCA	TGC AAT ACC TGC AAC ACC TGC AAC ACC CYS ASN TH ATA GGG ILE GLY CGG TGG GC ARG GTRP AL CGA GAA TC ARG GLU SE TCA TGA TG	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TGA A ARG SER C AAG TGA C AA	AAC AGA C ASN ARG L AAC AGA C ASN ARG L TGT GGA G TGT GGG G TGT GGG G ATT TCC C	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCC GGA GLY S II GT CCA GAA GLU S II GT CCA GAC ASP TT AAG CCA	TCT AAA GTG ACA SER LYS VAL THR GAC CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TTC TAC TGT SCP PHE CCC TTC TAC TGT SER PHE GGA AAC ACT CCT GLY ASN ACT CCT
M-eco xeno A-eco M-eco Xeno A-eco	GGC TGT TCC AGA CYS SER GGC TGT TCC AGA CYS SER E CAT GCA CAC AAT HIS ALA HIS ASN CAT AAA TCA AAT HIS LYS SER ASN GGC AAAA TGG GGA GLY LYS GCC TCT TGG GGC	GAT TGT GAG ASP CYS GLI GAC TGC GAA ASP CYS GLI TTG GLY GLY GAG GGA TTG GLY GLY TGT GAG ACC TGC GAA ACC	G GAG CCC C G GU PRO L A GAA CCT T G GU PRO L C TAT GTT TI C TAT GTT TI C TAT GTC TI C TAT GTT TI C TAT GTT TI C ACT GGA CI C ACA GGC CI A	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG GC CCC GG GC CCC GG GC S== I GC CCC GG GC GC GG AG GCA TA N GC GCA TC SA GCA TC	A TAT ACC R TYR THI C CTC ACC R LEU THI T CAT ACC HIS THE G CCA CAN PRO HIS G CCC CACC PRO HIS C TGG AAAA R	CCC CGG PRO ARG CCT CGG PRO ARG ARG CGC CCC ARG CGC CCC CCA TCA CCA TCC	TGC AAT AC CYS ASN TH TGC AAC AC CYS ASN TH ATA GGG ILE GLY CGG TGC GC ARG TRP AL CGA GAA TC ARG GLU SE ICA TCA TG ICG TCC TG	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TCA A ARG SER C AAG TCA R LYS SER AV-0 II G GAC CTA LEU G GAC TAC	AAC AGA CAGA CAGA CAGA CAGA CAGA CAGA C	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCG GGA GLY VI II GT CCA GAC GLU VI II GT CCA GAC ASP TT AAG CCA EU LYS ARG TA AGC AACC	TCT AAA GTG ACA SER LYS VAL THR GAC CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TTC TAT TGT SCP PHE CCC TTC TAT TGT SER PHE GGA AAC ACT CTT GLY ASN PEO AAT CTA ACC TCA SSN LEU SER
M-eco Xeno A-eco M-eco A-eco M-eco	GGC TGT TCC AGA CYS SER GGC TGT TCC AGA CYS SER E CAT GCA CAC AAT HIS ALA HIS ASH CAT AAA TCA AAT HIS LYS SER ASH GGC AAA TGG GGA GLY LYS GGC ALA SER GCC TCT TGG GGC ALA SER	GAT TGT GAG ASP CYS GLI GAC TGC GAJ ASP CYS GLI TTT GGA GGA TTT GLI GLV TGT GAG ACC TGT GAG ACA	G GAG CCC C J GLU PRO L A GAA CCT T J GLU PRO L C TAT GTT TI C TAT GTT TI C TAT GTT TI C TAT GTC TI C TAT GTC TI C TAT GTC TI C ACT GGA CA GI C ACA GGC CA A	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG GC CCC GG CCC GC CC GG GC CCC GC CCC GG GC CCC GC CCC CCC GC CCC CCC GC CCC CCC GC CCC	C TGG AAG	CCC CGG PRO ARG CCT CGG PRO ARG CATA CCA VAL CGC CCC CCA TCA CCA TCC CCC TCC	TEC AAT AC CYS ASN TH TEC AAC AC CYS ASN TH ATA GGG ILE GLY CGG TGG GC TGG TGG GC TGG ACA TC TGA TCA TG ICG TCC TG ICG TCC TG	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TCA A ARG SEC LVS SER AVB II G GAC TCA LVS SER AVB II G GAC TAC TYR S GAT TTC	AAC AGA C ASN ARG C AASN ARG C TGI GGA G TGI GGA G TGI GGA G TGI GGG G ATT TCC C ATT TCC C ATT ACA G TH TCC ACA G TH TCC ACA G	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCC GGA GLY VI II GT CCA GAC GA GT CCA GAC ASP TT AAG CCA AL SER ASN TA AAC AAC AL ASN ASN	TCT AAA GTG ACA SER LYS VAL THR 632 GAC CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TTC TAT TGT SFR PHE GAG ACA ACT GAG GGC TAC TGT GLU GLY GGA AAC ACT GGA AAC ACT GGA AAC ACT AAT CTA ACC AAT CTA ACC GAT CTC ACC AAT CTC ACC GAT CTC TC AAT CTC ACC GAT CTC TC AAT CTC ACC GAT CTC TC AAT CTC ACC GAT CTC TCT AST CTC ACC GAT CTC TCT
M-eco Xeno A-eco M-eco A-eco M-eco	GGC TGT TCC AGA CYS SER GGC TGT TCC AGA CYS SER E CAT GCA CAC AAT HIS ALA HIS ASH CAT AAA TCA AAT HIS LYS SER ASH GGC AAA TGG GGA GLY LYS GCC TCT TGG GGC ALA SER GCC TAT TGG GGC ALA TYR AAG GAT CAG GGC GLY	GAT TGT GAG ASP CYS GLI GAC TGC GAJ ASP CYS GLI TTT GGA GGA TTT GLI GLY TGT GAG ACC TGT GAG ACC TGT GAG ACC TGT GAG ACA	G GAG CCC C J GLU PRO L A GAA CCT T J GLU PRO L C TAT GTT TI C TAT GTT TI C TAT GTC TI L TAT GTC TI C ACT GGA C. ACT GGA CCC A L ACC GGT AC AF	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG Sma I GC CCC GG Sma I GC CCC GG GC CCC GG AG GCA TA SE GC CCC GG SC SC CCC GG SC CCC GG SC CCC GG SC SC S	C TGG AAG C TGG AAG	CCC CGG PRO ARG CCT CGG PRO ARG GTA CCA VAL CGC CCC ARG CGC CCC CCA TCA CCCA TCA CCCA TCC CCC TCC C GTC CAG	TGC AAT AC CYS ASN TH TGC AAC AC CYS ASN TH ATA GGG ILE GLY CGG TGG GC ARG TRF AL CGA GAA TC TCA TGA TG ICA TCA TG ICA TCA TG ICA TCA TG ICA TCA TG	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TCA A ARG SER C AAG TCA R LVS SER AVA II G GAC TAC LEU G GAC TAC TVR S GAT TYC R	AAC AGA C ASN ARG C AASN ARG C TGT GGA G TGT GGA G TGT GGA G TGT GGG G ATT TGC C ATT TGC C ATT AGA G ATT AGA G THR V	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCC GGA GLU YS II GT CCA GAA GLU YS II GT CCA GAA ASP TT AAG CCA AL SER ASN TA AAC AAC AL ASN ASN GUY GIN	TCT AAA GTG ACA SER LYS VAL THR GAC CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TTC TAT TGT SER PHE GGA AAC ACT GTC TAT ACC GGA AAC ACT GGA AAC ACT GTC ACT GAT CTA ACC AAT CTA ACC GAT CTC ACC TCT ASP LEU SER AAT TGT ACC ACC AAT TGT ACC ACC TCT
M-eco xeno A-eco H-eco A-eco H-eco A-eco	GEC TGT TCC AGA CYS SER GEC TGT TCC AGA CYS SER E CAT GCA CAC AAT HIS ALA HIS ASH CAT AAA TCA AAT HIS LYS SER ASH GLY LYS GCC AAA TGG GGA GLY LYS GCC TCT TGG GGC ALA SER ACC GAT CAG GGC LYS GAC CAG GAC CAG	GAT TGT GAG ASP CYS GLI GAC TGC GAA ASP CYS GLI TTT GGA GGA TTT GLI GLY TGT GAG ACC TGT GAG ACC TGT GAG ACC TGT GAG ACC	G GAG CCC C J GLU PRO L A GAA CCT T J GLU PRO L C TAT GTT TI C TAT GTT TI C TAT GTC TI L TAT GTC TI C ACT GGA C. A A GGC CT A A GAS SER S	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG GC CCC GG GC CCC GG GC S== I GC CCC GG GC GCC GG GC GCC GG GC TTA SG GCT TA SER VAL SI	A TAT ACC R TYR THI C CTC ACC R LEU THI T CAT ACT HIS THE G CCA CAT PRO HIS G CCC CACC PRO HIS C TGG AAG C TGG AAG C TGG AAG C TGG AAG C TGG AAG	CCC CGG PRO ARG CCT CGG PRO ARG CAT CCA ARG CGC CCC CCA TCA CCA TCC CCA TCC CCG TCC CGTC CAG	TEC AAT AC AC CYS ASN TH TEC AAC AC AC CYS ASN TH ATA GGG ILE GLY CGG TEG GC ARG TRP AL CGA GAA TC TCA TCA TG ICA TCA TG ICA TCA TG ICA TCA TG ICA TCA TG ICA TCA TG ICA A	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TCA A ARG SER C AAG TCA G GAC TAC G GAC TAC TVR G GAC TAC TVR C GAC TAC TVR C GAC TAC TVR C CCA GAC	AAC AGA C ASN ARG L AAC AGA C ASN ARG L IGI GGA G IGI GGA G IGI GGA G ATT TCC C AIT TCC C AIT TCC C AIC ACA G AIT TCC AAA AIT TCC AAA AIC ACA G IN V	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCC GGA GLY SGT CCA GAC GLY SGT CCA GAC ASP TT AAG CCA AL SER ASN GCG CCT GLY GLY SGT AAT GAC	TCT AAA GTG ACA SER LYS VAL THR GAO CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TTC TAC TGT SCP PHE GGA AAC ACT CCT GLY ASN LEU AAT CTA ACC TCA ASN LEU GGA AAC ACT CCT GLY ASS AAT CTC ACC TCA ASS CTC ACC TCA ACT CTA ACC TCA CST CTA CTC CTC CST CTA CTC CTC CST CTA CTC CTC CST CTC CTC CTC CTC CTC CTC CTC CTC CTC CT
H-eco xeno A-eco H-eco H-eco A-eco H-eco H-eco	GGC TGT TCC AGA CYS SER GGC TGT TCC AGA CYS SER E CAT GCA CAC AAT HIS ALA HIS ASN CAT AAA TGG CA AAT HIS LYS SER ASN GAC TAT TGG GGC ALA SER ALA TYR GCC TAT TGG GGC ALA TYR AG GAT CAG GGC LYS GAT CAG GGC GU GG GAC CAG GAC CAG	GAT TGT GAG ASP CYS GLU GAC TGC GAA ASP CYS GLU TTG GLY GLY GAG GGA TTT GLU GLY TGT GAG ACC TGT GAG ACC TGT GAG ACCA CCC TGT TA PRO CYS TY	G GAG CCC C J GLU PRO L A GAA CCT T J GLU PRO L TAT GTT TI TAT GTT TI TAT GTT TI TAT GTT TI ACT GGA C: ACT C: ACT C: ACT C: ACT C: ACT C: ACT C: ACT C: ACT C: ACT C: AC	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG GC TC SE SA GCT TA SE SA GCT TA SE SA GCT TA SE SA GCT TA	C TGG AAAG C TGG AAG C C AGC AGC C C AGC CAG C C AGC CAG	CCC CGG PRO ARG CCT CGG PRO ARG GTA CCA VAL CGC CCC ARG CGC CCC CCC TCC CCC TCC VAL GLN	TEC AAT AC AC CYS ASN TH TEC AAC AC AC CYS ASN TH ATA GGG ILE GLY CGG TCG GC AA ARG GLU SE ILCA TCA TG ICA TCA TG ICA TCA TG ICA ACA TG ICA ACA TG I	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TGA A ARG SER C AAG TGA C AAG TGA C AAG TGA C CAA C CCA C CCA C CAA GT/ AL GLN VAL	AAC AGA CAGA CAGA CAGA CAGA CAGA CAGA C	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCG GGA GLY VI II GT CCA GAC GLU LYS ARG TT AAG CCA EU LYS ARG TA AGC AAC AL ASN ASN GCG GCT GLY GLY GGT AAT GAG ASP ASN LYS	TCT AAA GIG ACA SER LYS VAL THR GAG CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TIC TAT TGT SCP PHE ACT CTT SCP PHE ACT CCT GGA AAC ACT CCT GLY ASN PRO AAT CTC ACC TCT ASN LEU SER GGA AAC ACT CCT GLY ASN PRO GGA CTC ACC TCT ASN LEU SER GGA CAC CCC GIG TGC AAC CCC GIG TGC AAC CCC GIG TGC AAC CCC SER GGG TGC AAC CCC SER
M-eco xeno A-eco H-eco A-eco M-eco A-eco H-eco H-eco	GEC TGT TCC AGA GEC TGT TCC AGA GEC TGT TCC AGA CYS SER CAT GCA CAC AAT HIS ALA HIS ASH CAT AAA TCA AAT HIS LYS SER ASH GLY LYS GCC TAT TGG GGC ALA SER GCC TAT TGG GGC ALA SER GAC CAG GAC CAG CIG GTC TTA GAA CAT GAC LEU GLU	GAT TGT GAG ASP CYS GLI GAC TGC GAA ASP CYS GLI TTT GGLY GLY GAG GGA TTT GLU GLY TGT GAG ACC TGT GAG ACC TGT GAG ACC TGT GAG ACC CCC TGT TA PRO CYS TY RI TTT	G GAG CCC C J GLU PRO L A GAA CCT T J GLU PRO L C TAT GTT TI C TAT GTC TI C TAT GTC TI C TAT GTC TI C TAT GTC TI C ACT GGA C. A A GGC CCT A A GGC CCT A A GAT TCC 1 R ASP SER S	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG GC CCC GG GC CCC GG GC S== I GC CCC GG GC S== I GC CCC GG GC S== I GC CCC GG GC TC TA SA GC CCT TA SA GC CCT TA SA GC TT TA SA SA GCA TC SE SA GCA TA SA SA GCA TA SA SA GCA TA SA SCA SCA TA SA SCA SCA SCA SCA SCA SCA SCA SCA SCA S	A TAT AC R TAT AC R TYR THI C CTC ACC R LEU THI T CAT ACT HIS THE G CCA CAT PRO HIS G CCC CACC PRO HIS C TGG AAG C TGG AAG C TGG AAG C TGG AAG	CCC CGG PRO ARG CCT CGG PRO ARG VAL CGC CCC ARG CGC CCC CCA TCA CCA TCC CCC TCC CCG TCC CGTC CAG	TEC AAT AC AC CYS ASN TH TEC AAC AC AC CYS ASN TH ATA GGG ILE GLY CGG TEG GC ARG TRP AL CGA GAA TC ARG GLU SE ICA TCA TG ICA TCA TCA TG ICA TCA TCA TG ICA TCA TCA TG ICA TCA TCA TCA TG ICA TCA TCA TCA TCA TCA TCA TCA TCA TCA T	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TCA A ARG SER C AAG TCA G GAC TAC G GAC TAC C CAG TAC C CAG GT/A C CCG C CCA GT/A C CGG T/A C CGG GT/A	AAC AGA C ASN ARG L ANC AGA C ASN ARG L TGI GGA G TGI GGA G TGI GGA G TGI GGA G AIT TCC C TGC GGA AIT TCC C TGC AAA AIC ACA G THR V I I I I I I I I I I I I I I I I I I I	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCG GGA GLY SI II GT CCA GAC GLY GLY TT AAG CCA GLU LYS ARG TA AGC AAC AL SER ASN GLY GLY GGT AAT GAG GLY ASN GLU GAT AAT AAG	TCT AAA GIG ACA SER LYS VAL THR 6202 GAC CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TTC TAC TGT SCP PHE CCC TTC TAC TGT SER PHE GGA AAC ACT CCT GLY ASN LEU CGA TCT ACC TCA AAT CTA ACC TCA SER AAT CTC ACC TCA AAT CTC ACC TCA SER STGG TGC AAC CCC TRP TC ACC TCC STGG TGC AAC CCC
M-eco xeno A-eco H-eco Xeno A-eco H-eco H-eco Xeno A-eco H-eco	GEC TGT TCC AGA GEC TGT TCC AGA GEC TGT TCC AGA CYS SER CAT GCA CAC AAT HIS ALA HIS ASH CAT AAA TCA AAT HIS LYS SER ASH GLY LYS GCC TAT TGG GGC ALA SER GCC TAT TGG GGC ALA SER GCC CAG GAC CAG GAC CAG GAC CAG CTG GTC TTA CAA CAT ACT ATC CGG	GAT TGT GAG ASP CYS GLI GAC TGC GAA ASP CYS GLI TTG GGLY GLY GAG GGA TTG GGLY GLY TGT GAG ACC TGT GAG ACC TGT GAG ACC TGT GAG ACC CCC TGT TA PRO CYS TY RI TTC	G GAG CCC C J GLU PRO L A GAA CCT T J GLU PRO L C TAT GTT TI C TAT GTT TI C TAT GTC TI C TAT GTC TI C TAT GTC TI C ACT GGA CA A A GGC GGT A A GASP SER S	TG ACT TC EU THR SE TA ACC TC EU THR SE GC CCC GG GC TC GC CCC GG GC TC TC GC CCC TA GC CCC GG GC CCC GG GC CC TA GC CCC GG GC CC CC GG GC CCC GG GC CC TA GC CCC GC CC GC CC GC CC GC CC GC CC CC	A TAT AC R TAT AC R TYR THI C CTC ACC R LEU THI T CAT ACT HIS THE G CCA CAT PRO HIS G CCC CAC PRO HIS C TGG AAA C TGG AAA C TGG AAA C TGG AAA C TGG AAA C TGG AAA	CCC CGG PRO ARG CCT CGG PRO ARG VAL CGC CCC ARG CGC CCC CCA TCA CCA TCC CCC TCC CGTC CAG CGC CCC	TEC AAT AC AC CYS ASN TH TEC AAC AC AC CYS ASN TH ATA GGG ILE GLY CGG TEG GC ARG TRP AL CGA GAA TC ARG GLU SE ICA TCA TG ICG TCC TG ICA TCA TCA TG ICA TCA TCA TCA TG ICA TCA TCA TCA TG ICA TCA TCA TCA TCA TCA TCA TCA TCA TCA T	G GCC TGG R ALA TRP T GCC TGG R ALA TRP F C CGG TCA A ARG SER C AAG TCA C AAG TCA C AAG TCA C AAG TCA C AAG TAC C CCA C CCA C CCA C CCA C CAA G TAC	AAC AGA CAGA CAGA CAGA CAGA CAGA CAGA C	TT AAG TTA EU LYS LEU TC AAG CTA EU LYS LEU GG CCG GGA GLY V= II GT CCA GAC GLU LYS AG EU LYS AG EU LYS AG TT AAG CCA EU LYS AG TA AGC AAC AL ASC AAC AL ASC AAC GCY GLY GGT GAT GAC GAT AAT AAG ASP ASN LYS	TCT AAA GIG ACA SER LYS VAL THR GAO CAG ACA ACT ASP GLN THR THR GAG GGC TAC TGT GLU GLY TCC TTC TAT TGT SCP PHE CCC TTC TAT TGT SCP PHE GGA AAC ACT CCT GLY ASN CT CCT CCT CT CCT GLY ASN CT CCT CCT CT CCT GLY ASN CT CCT CCT CT CCT CT CT CT CT CT CT CT

and Akv ecotropic MuLVs (17, 26) (Fig. 3). The difference in the size (93 nucleotides) of comparable regions between the xenotropic and ecotropic env genes was due to several small and two larger deletions of 36 and 78 nucleotides (Fig. 3, D and E) in the xenotropic proviral DNA. Conversely, the xenotropic *env*-specific region contained two small insertions and a larger 42-bp insertion (Fig. 3, H) relative to the two ecotropic env genes. The xenotropic and ecotropic env-specific segments shared two discrete regions of polynucleotide sequence homology. The first involved sequences encoding amino acids near the amino terminus of gp70 (corresponding to Mo MuLV nucleotides 5,891 to 5,932 [26]), in which 29 of 42 nucleotides and 11 of 14 amino acids were identical. The second occurred between Mo MuLV nucleotides 6,299 and 6,352, encompassing a stretch of 54 nucleotides, of which 48 were identical to those present in either Akv or Mo MuLV.

The xenotropic env gene sequence shown in Fig. 3 had a single open reading frame. The amino terminus of Mo MuLV gp70 was previously positioned at nucleotide 5,876, 100 bp downstream from the beginning of the env gene (26). By analogy, the amino terminus of Akv gp70 was located at the same position, although the amino acid at position 1 (valine) (17) is different from the initial amino acid (alanine) of the Mo MuLV gp70 (Fig. 3). Since the amino acid sequence of the NFS-Th-1 xenotropic MuLV gp70 has not been determined, its location within the deduced env precursor polypeptide can only be speculated on. If the amino terminus of the xenotropic MuLV gp70 is located at precisely the same position within the env gene as it is in the Akv and Mo MuLV gp70s, then the first amino acid would be serine (Fig. 3). However, this region of the xenotropic env gene shared little polynucleotide sequence homology with either of the two ecotropic env segments. A shift in the alignment of the xenotropic env sequence in this region by one codon in the 3' or 5' direction would result in a gp70 with an amino-terminal amino acid identical to Mo MuLV (alanine) or Akv MuLV (valine).

As pointed out previously, the 5'-terminal portion of the xenotropic *env* gene was highly conserved relative to the analogous segments of Mo and Akv MuLVs, particularly in the region of overlap with the *pol* gene. The direction of the shift from the *pol* to the *env* reading frame was identical. This part of the MuLV *env* gene encodes a hydrophobic leader sequence (17, 26) that is removed during the maturation of gp70 (11, 15). This portion of the NFS-Th-1 xenotropic *env* region also contains uncharged, mostly hydrophobic amino acids that are probably components of an envelope precursor. MuLV envelope glycoproteins are translated from spliced mRNAs (12, 23). A consensus 3' acceptor sequence (25) was identified 275 nucleotides upstream from the 5' terminus of the xenotropic *env* gene, at a location and with a nucleotide sequence identical to that present in Mo MuLV (Fig. 2).

Within the sequence shown (Fig. 3), the *env* regions of both Mo and Akv MuLVs encoded two potential glycosylation sites. One of these (between Mo MuLV nucletides 5,909 and 5,917) was also present in the xenotropic *env* gene. A second potential glycosylation site in the xenotropic *env* region with no counterpart in the ecotropic gp70 was located between nucleotides 5,960 and 5,968 (Fig. 3).

As pointed out above, the xenotropic env gene contains several insertions and deletions relative to the analogous segment of the ecotropic genome, which may be responsible for the unique biological properties of each MuLV type. Segment D (Fig. 3), which is missing from xenotropic MuLVs, is particularly rich in proline residues. Its absence would certainly affect the secondary structure of this portion of the xenotropic gp70 molecule. The base substitutions found in the xenotropic env gene that generate regions of polynucleotide heterology could have a less obvious effect on the structure of gp70 than would deletions or insertions. For example, both ecotropic and xenotropic env gene products have a peak of hydrophilicity (average indices of 1.02 and 1.34, respectively, calculated as described by Hopp and Woods [14]) in the region just 5' to E (Fig. 3). Unexpectedly, an adjacent region of the xenotropic gp70 was also extremely hydrophilic (an average index of 0.94), in the area (between nucleotides 6.032 and 6.094) in which the ecotropic env protein had an index of -0.07. Assuming that this portion of the xenotropic envelope glycoprotein, like other strongly hydrophilic regions, is located on the surface of the gp70 molecule, this polypeptide segment could determine some of the characteristic properties of xenotropic MuLVs.

The recently published sequence of the *env* gene of BALB/c Mo MCF MuLV (2) is remarkably similar to the *env* segment of NFS-Th-1 xenotropic proviral DNA shown in Fig. 3. A comparison of both the nucleotide and derived amino acid sequences shows that homology was greater than 85%. Without exception, the MCF sequence also had the same insertions and deletions identified in the xenotropic *env* segment relative to the ecotropic MuLV *env* region (Fig. 3, A through I). In addition, the two putative glycosylation sites in the NFS-Th-1 xenotropic *env* gene were located in the same positions in the MCF sequence; the second site (nucleotides 5,960)





--- --- --- --- --

FIG. 4. Sequence of a portion of an endogenous MuLV env gene cloned from BALB/c mouse embryo DNA and its relationship to the env region of NFS-Th-1 xenotropic MuLV. (Panel 1) Diagrammatic representation of the xenotropic env-specific segment of NFS-Th-1 xenotropic proviral DNA and its alignment with the analogous region of clone B-77, isolated from a BALB/c embryo gene library and found to have a restriction map and hybridization properties unique to xenotropic proviruses (16). The sequencing strategy is indicated by the arrows; the regions of clone B-77 DNA not sequenced are shown by the dashed lines. (Panel 2) The sequence of 270 nucleotides of clone B-77 DNA (bottom line), extending from the pol-env gene junction in the 3' direction, was aligned with the analogous segment of NFS-Th-1 xenotropic proviral DNA (top line). Identical deduced amino acids are shaded. The letters designate insertions and deletions unique to the xenotropic MuLV env gene relative to the ecotropic env segment and follow the convention described in the legend to Fig. 3. The boxed tripeptides denote potential glycosylation sites. (Panel 3) Further comparison of the nucleotide sequence of clone B-77 env region (bottom line) with the 3' portion of the NFS-Th-1 xenotropic env-specific segment (top line) as shown in panel 1. Identical deduced amino acids are shaded. The letters to the comparable regions in the ecotropic MuLV env gene relative to the cortopic env specific segment (top line) as shown in panel 1. Identical deduced amino acids are shaded. The letters indicate the insertion (H) and deletion (I) in the xenotropic MuLV env gene relative to the cortopic MuLV env gene relative to the cortopic MuLV env gene relative to the comparable regions in the ecotropic MuLV genu gene relative to the comparable regions in the ecotropic cenv-specific segment (top line) as shown in panel 1. Identical deduced amino acids are shaded. The letters indicate the insertion (H) and deletion (I) in the xenotropic MuLV env gene relative to the comparable reg

5,968, Fig. 3), as mentioned previously, was not coincident with any known ecotropic MuLV envelope glycosylation site.

Identification of a xenotropic env gene present in an endogenous MuLV provirus cloned from BALB/c mouse DNA. Khan et al. (16) cloned several endogenous MuLV proviruses from AKR and BALB/c mouse DNAs. Five of the twelve clones that contained long terminal repeat segments also had env regions that hybridized to the xenotropic env-specific probe. Restriction enzyme mapping of the five endogenous env segments indicated that some contained sites characteristic of MCF MuLVs (such as a BamHI site at 6.2 kb associated with an EcoRI site at 6.7 kb). More revealing, however, were the unique hybridization properties of four of the five xenotropic *env*-reactive clones. Unlike typical xenotropic MuLV proviral DNA, which anneals exclusively to the pX_{env} probe and not to the ecotropic env-specific DNA, these four endogenous env segments reacted strongly with labeled pX_{env} DNA and weakly but reproducibly with the pEc_{env} probe (16). This pattern of hybridization (dual reactivity) was shown to be characteristic of the MCF env gene, and the results obtained with the four endogenous env segments suggested that they might be progenitors of infectious MCF MuLVs (16).

The fifth cloned endogenous env segment that was isolated from a BALB/c mouse library (associated with clone B-77) behaved like a typical xenotropic proviral DNA and hybridized only to the pX_{env} DNA probe. Since the restriction map of the B-77 env region was also similar to the analogous segment of the NFS-Th-1 xenotropic provirus, portions of its env gene were sequenced and compared with the xenotropic envspecific sequence. The two segments of the B-77 envelope gene analyzed were identical to comparable portions of the xenotropic env gene in 380 out of 408 nucleotides (93% homology) and 123 out of 136 amino acids (90% homology) (Fig. 4). All of the characteristic insertions and deletions of the NFS-Th-1 xenotropic env gene shown in Fig. 3, including the 42-bp insertion H, were present in the B-77 endogenous env segment. In addition, the B-77 env region, like the xenotropic env gene, contained a second potential glycosylation site (located between A and B in Fig. 4, panel 2) that was absent from the comparable segment of the ecotropic proviral DNA (nucleotides 5,960 to 5,968, Fig. 3). All of the deduced amino acid differences between the B-77 and the NFS-Th-1 xenotropic MuLV envelopes were single base changes. These sequencing results showed several characteristic features of xenotropic env genes that distinguished them from analogous segments of ecotropic

MuLVs. The role of each in determining properties such as tissue tropism, host range, and binding to cell receptors awaits the construction of recombinants containing specific segments of ecotropic and xenotropic MuLV *env* genes.

ACKNOWLEDGMENTS

We thank Charles E. Buckler, who assisted in the alignment of the xenotropic MuLV *env* segment with ecotropic MuLV *env* genes. We also thank John Coligan and W. L. Maloy for helpful discussions of protein structure and function relationships.

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.
- Bosselman, R. A., F. van Straaten, C. Van Beveren, I. M. Verma, and M. Vogt. 1982. Analysis of the *env* gene of a molecularly cloned and biologically active Moloney mink cell focus-forming proviral DNA. J. Virol. 44:19–31.
- Buckler, C. E., M. D. Hoggan, H. W. Chan, J. F. Sears, A. S. Khan, J. L. Moore, J. W. Hartley, W. P. Rowe, and M. A. Martin. 1982. Cloning and characterization of an envelope-specific probe from xenotropic murine leukemia proviral DNA. J. Virol. 41:228-236.
- Buckler, C. E., S. P. Staal, W. P. Rowe, and M. A. Martin. 1982. Variation in the number of copies and in the genomic organization of ecotropic murine leukemia virus proviral sequences in sublines of AKR mice. J. Virol. 43:629-640.
- Chan, H. W., T. Bryan, J. L. Moore, S. P. Staal, W. P. Rowe, and M. A. Martin. 1980. Identification of ecotropic proviral sequences in inbred mouse strains with a cloned subgenomic DNA fragment. Proc. Natl. Acad. Sci. U.S.A. 77:5779-5783.
- Chattopadhyay, S. K., M. W. Cloyd, D. L. Linemeyer, M. R. Lander, E. Rands, and D. R. Lowy. 1982. Cellular origin and role of mink cell focus-forming viruses in murine thymic lymphomas. Nature (London) 295:25-31.
- Chattopadhyay, S. K., M. R. Lander, E. Rands, and D. R. Lowy. 1980. Structure of endogenous murine leukemia virus DNA in mouse genomes. Proc. Natl. Acad. Sci. U.S.A. 77:5774-5778.
- Chattopadhyay, S. K., D. R. Lowy, N. M. Teich, A. S. Levine, and W. P. Rowe. 1974. Qualitative and quantitative studies of AKR-type murine leukemia virus sequences in mouse DNA. Cold Spring Harbor Symp. Quant. Biol. 39:1085-1101.
- 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.
- Elder, J. J., F. C. Jensen, M. L. Bryant, and R. A. Lerner. 1977. Polymorphism of the major envelope glycoprotein (gp70) of murine C-type viruses: virion associated and differentiation antigens encoded by a multi-gene family. Nature (London) 267:23-28.
- Famulari, N. G., D. L. Buchhagen, H. D. Klenk, and E. Fleissner. 1976. 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 mRNAs: evidence for transposition of sequences during synthesis and processing of subgenomic mRNA. J. Virol. 26:468– 478.
- Hoggan, M. D., C. E. Buckler, J. F. Sears, H. W. Chan, W. P. Rowe, and M. A. Martin. 1982. Internal organization of endogenous proviral DNAs of xenotropic murine

Vol. 46, 1983

leukemia viruses. J. Virol. 43:8-17.

- Hopp, T. P., and K. R. Woods. 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. 78:3824-2828.
- 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.
- 16. Khan, A. S., W. P. Rowe, and M. A. Martin. 1982. Cloning of endogenous murine leukemia virus-related sequences from chromosomal DNA of BALB/c and AKR/J mice: identification of an *env* progenitor of AKR-247 mink cell focus-forming proviral DNA. J. Virol. 44:625-636.
- Lenz, J., R. Crowther, A. Straceski, and W. Haseltine. 1982. Nucleotide sequence of the Akv env gene. J. Virol. 42:519-529.
- Levy, J. A. 1973. Xenotropic viruses: murine leukemia viruses associated with NIH Swiss, NZB, and other mouse strains. Science 182:1151-1153.
- Maxam, A., and W. Gilbert. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. 65:499-560.
- Morse, H. C., III, T. M. Chused, M. Boehm-Truitt, B. J. Mathleson, S. L. Sharrow, and J. W. Hartley. 1979. XenCSA: cell surface antigens related to one major glycoprotein (gp70) of xenotropic murine leukemia viruses. J. Immunol. 122:443-454.

- Queen, C. L., and L. J. Korn. 1980. Computer analysis of nucleic acids and proteins. Methods Enzymol. 65:595-609.
- Roblin, R., J. M. Young, R. J. Mural, T. E. Bell, and J. N. Ihle. 1982. Molecular cloning and characterization of murine leukemia virus-related DNA sequences from C3H/HeN mouse DNA. J. Virol. 43:113–126.
- 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.
- Rowe, W. P. 1978. Leukemia virus genomes in the chromosomal DNA of the mouse. Harvey Lect. 71:173-192.
- 25. Sharp, P. 1981. Speculations on RNA splicing. Cell 23:643-646.
- Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) 293:543-548.
- Steffen, D., S. Bird, W. P. Rowe, and R. A. Weinberg. 1979. Identification of DNA fragments carrying ecotropic proviruses of AKR mice. Proc. Natl. Acad. Sci. U.S.A. 76:4554-4558.
- Steffen, D., and R. A. Weinberg. 1978. The integrated genome of murine leukemia virus. Cell 15:1003-1010.
- Stockert, E., L. J. Old, and E. A. Boyse. 1971. The G_{IX} system. A cell surface allo-antigen associated with murine leukemia virus; implications regarding chromosomal integration of the viral genome. J. Exp. Med. 133:1334–1355.