# Nucleotide Sequence Analysis Establishes the Role of Endogenous Murine Leukemia Virus DNA Segments in Formation of Recombinant Mink Cell Focus-Forming Murine Leukemia Viruses

ARIFA S. KHAN

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

Received 6 December 1983/Accepted 13 March 1984

The sequence of 363 nucleotides near the 3' end of the *pol* gene and 564 nucleotides from the 5' terminus of the env gene in an endogenous murine leukemia viral (MuLV) DNA segment, cloned from AKR/J mouse DNA and designated as A-12, was obtained. For comparison, the nucleotide sequence in an analogous portion of AKR mink cell focus-forming (MCF) 247 MuLV provirus was also determined. Sequence features unique to MCF247 MuLV DNA in the 3' pol and 5' env regions were identified by comparison with nucleotide sequences in analogous regions of NFS-Th-1 xenotropic and AKR ecotropic MuLV proviruses. These included (i) an insertion of 12 base pairs encoding four amino acids located 60 base pairs from the 3' terminus of the *pol* gene and immediately preceding the *env* gene, (ii) the deletion of 12 base pairs (encoding four amino acids) and the insertion of 3 base pairs (encoding one amino acid) in the 5' portion of the env gene, and (iii) single base substitutions resulting in 2 MCF247-specific amino acids in the 3' pol and 23 in the 5' env regions. Nucleotide sequence comparison involving the 3' pol and 5' env regions of AKR MCF247, NFS xenotropic, and AKR ecotropic MuLV proviruses with the cloned endogenous MuLV DNA indicated that MCF247 proviral DNA sequences were conserved in the cloned endogenous MuLV proviral segment. In fact, total nucleotide sequence identity existed between the endogenous MuLV DNA and the MCF247 MuLV provirus in the 3' portion of the pol gene. In the 5' env region, only 4 of 564 nucleotides were different, resulting in three amino acid changes between AKR MCF247 MuLV DNA and the endogenous MuLV DNA present in clone A-12. In addition, nucleotide sequence comparison indicated that Moloneyand Friend-MCF MuLVs were also highly related in the 3' pol and 5' env regions to the cloned endogenous MuLV DNA. These results establish the role of endogenous MuLV DNA segments in generation of recombinant MCF viruses.

Mink cell focus-forming (MCF) murine leukemia viruses (MuLVs) have been isolated from preleukemic and leukemic mouse tissues (9, 10). Heteroduplex mapping (5, 7) oligonucleotide fingerprinting (23, 28), as well as restriction enzyme mapping and Southern blot analyses (4, 17) indicate that MCF viruses are recombinants between ecotropic MuLVs and nonecotropic sequences. Since ecotropic viruses lack characteristics of MCF MuLVs such as dualtropism, mink cell focus-forming activity, and ability to accelerate leukemias in newborn mice (6), it is most likely that the nonecotropic sequences acquired during recombination contribute to expression of some of these properties unique to MCF viruses. I have previously reported the characterization by restriction enzyme mapping and Southern blot analyses of several endogenous MuLV proviruses isolated from AKR/J and BALB/c mouse DNAs (17). The cloned endogenous MuLV DNA segments were distinct from known infectious MuLV proviruses due to the insertion of a 190-base-pair (bp) transposon-like segment in the long terminal repeats (LTR) (15) and by the presence of unique restriction sites in the gag and pol regions (17). However, the env segment associated with the endogenous MuLV DNAs was related to the env region in MCF MuLV proviruses based on similar-sized restriction fragments and dual-reactive hybridization properties. Furthermore, restriction enzyme and Southern blot analyses suggested that the env region in one cloned endogenous MuLV DNA segment, designated as A-12, might represent the nonecotropic progenitor of AKR MCF247 MuLV DNA. To establish this relationship between endogenous MuLV DNAs and MCF proviruses, the nucleotide

sequence in the 3' end of the *pol* gene and in the 5' portion of the env gene in clone A-12 and AKR MCF247 MuLV DNAs was determined. Nucleotide sequence comparison of AKR MCF247, NFS xenotropic, and AKR ecotropic MuLV DNAs permitted the identification of sequences unique to MCF247 in the 3' pol and 5' env regions. The MCF247specific sequence features were conserved in the endogenous MuLV DNA present in clone A-12, which distinguished the latter DNA from both xenotropic and ecotropic MuLV proviruses. In fact, almost total nucleotide sequence identity was seen in the 3' pol and 5' env regions between MCF247 MuLV provirus and the cloned endogenous MuLV DNA. In addition, the nonecotropic sequences present in recombinant Moloney- (M) (2) and Friend- (F) (A. Adachi, K. Sakai, N. Kitamura, S. Nakanishi, O. Niwa, and A. Ishimoto, submitted for publication) MCF MuLVs were also virtually identical in the 3' pol and 5' env regions to the endogenous MuLV provirus present in clone A-12.

## MATERIALS AND METHODS

MuLV DNA clones used for nucleotide sequencing. A recombinant plasmid DNA clone containing a 6.8-kilobasepair (kb) DNA segment from AKR MCF247 MuLV provirus which extended from an EcoRI site in the 5' LTR to an EcoRI site in the env region (at 6.9 kb) was used for nucleotide sequencing. The construction and characterization of this DNA clone (previously designated as pMCF-1) has been described previously (16). The nucleotide sequence of an endogenous MuLV DNA segment was determined with a recombinant lambda Charon 4A DNA clone designat-



FIG. 1. Strategy used for determining nucleotide sequences in the 3' pol and 5' env regions of AKR MCF247 DNA and endogenous MuLV DNA in clone A-12. Arrows indicate regions sequenced from minus strand ( $\leftarrow$ --) or plus strand ( $\rightarrow$ ) DNA. Restriction enzyme designations are as follows: Pv, PvuII; B, BamHI; S, SmaI; E, EcoRI.

ed as A-12. The A-12 DNA clone, which was isolated from AKR/J mouse DNA as described previously (17), consisted of a 15-kb DNA segment and contained endogenous MuLV proviral sequences extending from the 5' LTR to the end of the clone, which was an EcoRI site at 6.9 kb in the *env* region.

**DNA sequence determination.** Due to the presence of identical restriction sites in the 3' *pol* and 5' *env* regions of AKR MCF247 provirus and the endogenous MuLV (clone A-12) DNA, a similar nucleotide sequencing strategy could be used for both DNAs (shown in Fig. 1). Recombinant lambda and plasmid DNAs or fragments isolated from preparative agarose gels were restricted with enzymes and labeled at their 5' termini with  $[\gamma^{-32}P]$ ATP and polynucleotide kinase. Labeled segments were recleaved and isolated on agarose gels and their nucleotides sequenced by the partial degradation method of Maxam and Gilbert (19).

## RESULTS

Identification of sequences unique to AKR MCF247 MuLV DNA. The sequence of 363 nucleotides near the 3' end of the *pol* gene and 564 nucleotides from the 5' terminus of the *env* gene in a cloned AKR MCF247 MuLV DNA segment designated previously as pMCF-1 (16) was determined by the strategy shown in Fig. 1. A comparison of the nucleotide sequence in MCF247 MuLV DNA with that previously reported for NFS-Th-1 xenotropic (22) and AKR ecotropic (12) MuLV proviruses permitted the identification of MCF247-specific sequences.

pol region. Nucleotide sequence comparison in the 3' end of the pol gene of MCF247, NFS xenotropic, and AKR ecotropic MuLV DNAs is shown in Fig. 2. A splice acceptor site for env mRNA (27), located from nucleotides 16 to 24, is conserved in the three MuLV proviruses. The MCF247 MuLV DNA could be distinguished from xenotropic and ecotropic MuLV proviral sequences due to the insertion of 12 bp (designated as X in Fig. 2) extending from nucleotides 292 to 303 and located 60 bp from the 3' end of the *pol* gene. It is interesting to note that the adenine residue, which is the last nucleotide in the MCF247-specific 12-bp insert, becomes the first nucleotide of an ATG codon (indicated by an asterisk in Fig. 2), which is a potential amino terminus of an env gene product. The location of the 12-bp sequence in the pol reading frame predicts that the 3' pol gene-coded protein product of MCF247 DNA would be four amino acids larger than the protein encoded from the corresponding regions in

xenotropic or ecotropic MuLVs. The nucleotide sequence on either side of this 12-bp insert in MCF247 DNA is more homologous to the NFS xenotropic provirus (96%) than to AKR ecotropic MuLV DNA (85%). In addition, single base substitutions in the 3' region of *pol* result in only two heterologous amino acids (positions 43 and 78) between xenotropic and MCF247 MuLV DNAs (resulting in 95% homology), whereas 12 amino acids differ between ecotropic and MCF247 MuLV DNAs (resulting in 87% homology). It should be noted that in the 3' end of the *pol* gene, the region of greatest sequence heterology in MCF247, NFS xenotropic, and AKR ecotropic MuLV DNAs extends from nucleotides 280 to 312 and includes the 12-bp insert.

env region. The nucleotide sequence of the 5' portion of the env genes of AKR MCF247, NFS xenotropic, and AKR ecotropic MuLV proviruses is compared in Fig. 3. MCF247 MuLV DNA could be distinguished from xenotropic and ecotropic MuLV proviruses in the 5' env region based on the following unique features: (i) 12 bp encoding four amino acids were deleted in MCF MuLV DNA (designated as Y in Fig. 3), (ii) 3 bp (from 499 to 501) were inserted which resulted in the insertion of an MCF247-specific alanine at position 167 (indicated as Z in Fig. 3), and (iii) other single base substitutions resulted in 23 MCF247-specific amino acids, 3 in the leader peptide (8) and 20 in the gp70-coding region. Comparison of nucleotide sequences in the 5' env regions of MCF247, NFS xenotropic, and AKR ecotropic MuLV DNAs indicated 89% nucleotide sequence homology between MCF247 and NFS xenotropic MuLV DNAs and only 48% polynucleotide sequence homology between MCF247 and AKR ecotropic MuLV proviruses. The MCF247 MuLV DNA is further related to NFS xenotropic proviral DNA on the basis of conservation of a series of nucleotide deletions (such as the region between nucleotides 315 and 316 in Fig. 3) and insertions (such as the sequence between nucleotides 477 and 523 in Fig. 3) relative to the AKR ecotropic MuLV provirus in the 5' env region (22). A glycosylation site (20, 24) at nucleotide positions 127 to 135 (Fig. 3) is conserved between the three classes of MuLV DNAs; in addition, a second potential glycosylation site located at positions 172 to 180 is shared between MCF247 and NFS xenotropic MuLV DNAs. It is interesting to note that the greatest nucleotide sequence divergence in the 5' env regions of MCF247, NFS xenotropic, and AKR ecotropic MuLV proviruses occurs in the region of the 12-bp deletion (Y in Fig. 3) and the 3-bp insertion (Z in Fig. 3) present in MCF247 DNA.

Nucleotide sequence analysis of the 3' pol and 5' env regions of endogenous MuLV DNA. The env region present in a cloned endogenous MuLV DNA, A-12, was previously shown by restriction enzyme mapping and Southern blot analysis to be a potential progenitor of AKR MCF247 MuLV DNA (17). To establish this relationship, the nucleotide sequence in the 3' pol and 5' env regions of the cloned endogenous MuLV DNA was determined by the strategy shown in Fig. 1 and compared with sequences in an analogous region in AKR MCF247 DNA. In the 3' pol region, total nucleotide sequence identity existed between the endogenous MuLV DNA segment in clone A-12 and MCF247 MuLV DNA (Fig. 4). This included conservation of the splice acceptor site located from nucleotides 16 to 24 and the 12-bp MCF247-specific insert designated as X (nucleotides 291 to 303) and located 60 bp from the 3' end of the *pol* gene. In addition, the two amino acids specific to the MCF247 MuLV provirus (positions 43 and 78) were also conserved in the cloned endogenous MuLV DNA.

J. VIROL.

Nucleotide sequence comparison of the 5' env regions of the cloned endogenous and AKR MCF247 MuLV DNAs is shown in Fig. 5. All the MCF247-specific sequence hallmarks (Fig. 3) could be identified in the endogenous MuLV DNA. These included the 12-bp deletion with respect to xenotropic and ecotropic proviruses between nucleotides 261 and 262 (indicated as Y in Fig. 5) and the 3-bp insertion (nucleotides 499 to 501) encoding alanine at position 167 (designated as Z in Fig. 5). In fact, nucleotide sequences in the 5' env region of the endogenous MuLV provirus in clone A-12 and AKR MCF247 MuLV DNA were virtually identical; only 4 of 564 bases were different, resulting in three amino substitutions, two of which (positions 34 and 157) were present in the gp70-coding region and one of which (position 22) was present in the leader region. Moreover, 21 of the 23 MCF247-specific amino acids in the 5' env region were also present in the cloned endogenous MuLV DNA.

Nucleotide sequences in the 3' pol and 5' env regions of

the endogenous MuLV DNA in clone A-12 were also compared with those present in the analogous regions in M-MCF (2) and F-MCF (Adachi et al., submitted for publication) MuLV proviruses (Fig. 4 and 5). The MCF247-unique sequences found in the cloned endogenous MuLV DNA, which included the 12-bp insertion in the 3' region of the pol gene, the 12-bp deletion, and the 3-bp insertion in the 5' portion of the env gene, could also be identified in the same nucleotide positions in the M-MCF and F-MCF MuLV proviruses. Furthermore, a high degree of polynucleotide sequence homology was observed in the *pol* and *env* regions of M-MCF and F-MCF proviral DNAs and comparable segments of the cloned endogenous A-12 DNA. In the 3' pol region, M-MCF and the endogenous MuLV DNAs were different in only 8 of the 249 bp present in the nonecotropic portion of M-MCF DNA (Fig. 4). These nucleotide changes did not lead to amino acid alterations. Ten base substitutions were noted between F-MCF and the cloned endogenous

																											_ Pv	u 1I	
A-MCF247	CCC Pro	TCT Ser	CTC Leu	O CAA GLN	GCT Ala	CAC HIS	2'0 TTA LEU	CAG GLN	GCC Ala	30 CTC LEU	CAA Gln	GCA Ala	GTA VAL	40 CAA GLN	CGA Arg	GAG GLU	50 GTC VAL	TGG Trp	AAG Lys	60 CCA PRO	C T G L E U	GCC Ala	GCT Ala	70 GCC Ala	T A T T Y R	C A G G L N	80 GAC ASP	CAG GLN	28
NFS XENO	CCC Pro	TCT SER	CTC LEU	CAA GLN	GC T A L A	CAC HIS	TTA LEU	CAG GLN	GCC Ala	C T C L E U	CAA GLN	GCA ALA	GTA VAL	CAA Gln	CGA ARG	GAG GLU	GTC Val	TGG TRP	A A G L Y S	CCA Pro	C T G L E U	GCC Ala	GCT Ala	GCT Ala	T A T T Y R	C A G G L N	GAC ASP	CAG GLN	
AKR ECO	CCA Pro	TCT SER	CTC LEU	CAA GLN	GCT Ala	CAC HIS	TTA LEŪ	CAG GEN	GCC Ala	CTC LEU	CAA GLN	ACG Thr	GT <b>G</b> Val	CA <b>S</b> Gln	CGA ARG	GA <b>A</b> Glu	ATT ILF	TGG TRP	AAA Lys	CCA Pro	CTG LEU	GCC Ala	G <b>as</b> Glu	GCC Ala	T A C T Y R	CGG Arg	GAC ASP	CAA Gln	
A-MCF247	<u>cie</u>	90 GAC	CAG	CCA	1 ( G T G	00 ATA	CCA	CAC	10 CCC	TTC	CGT	120 GTC	GGC	GAC		30 GTG	TGG	GTA	140 CGC	CGG	CAC	150 CAG	ACT	AAG	1 e AAC	60 ⊺⊺163	GAA	CCT	F /
NFS XENQ		GAT ASP	CAG GLN	CCA PRO	GTG VAL	ATA ILE	CCA PRO	CAC HIS	CCC PRO	TTC PHE	CGT ARG	GTC VAL	GGT GLY	GAC ASP	GCC ALA	GTG VAL	TGG TRP	GTA VAL	CGC ARG	CGG ARG	CAC HIS	CAG GLN	ACT THR	AAG LYS	ASN AAC ASN	TTA LEU	GAA GLU	CCC Pro	36
AKR ECO	C T.∰ L E U	GAC ASP	CA <b>A</b> Gln	CCA Pro	GTG Val	ATA ILE	CCA Pro	CAC HIS	CCC Pro	TTC Phe	CGS6 ARG	ATT ILE	GG <b>a</b> Gly	GAC ASP	TCC Ser	GTG VAL	TGG TRP	GT <b>S</b> Val	CGC ARG	CGG ARG	CAC HIS	C A G G L N	ACC Thr	A A A L Y S	AAC ASN	TTA Leu	GÀA Glu	CCT Pro	
																								Bai	n HI				
	170			180			1 1	90			200			210			2	20			230	_		240	_		2 !	50	
A-MCF247	CGC ARG	T G G T R P	AAA LYS	GGA Gly	CCC Pro	T A C T Y R	ACC Thr	GTC VAL	CTG LEU	C T G L E U	ACC Thr	ACC Thr	CCC PRO	ACC THR	GC T A L A	CTC LEU	A A A L Y S	GTA VAL	GAC ASP	GGC GLY	ATC ILE	GC T A L A	GCG Ala	T G G T R P	AIC	CAC HIS	GCC Ala	GCT Ala	84
NFS XENO	CGC ARG	T G G T R P	A A A L Y S	GGA Gly	CCC Pro	T A C T Y R	ACC Thr	GTÇ Val	C T G L E U	C T G L E U	ACC THR	ACC Thr	CCC Pro	ACC Thr	G C T A L A	CTC LEU	AAA LYS	GTA VAL	GAC ASP	GGC Gly	ATC ILE	TCC Ser	GCG Ala	TGG TRP	ATA Ile	C A C H I S	GCC Ala	GCT ALA	
AKR ECO	C G C A R G	TGG TRP	AA∯ LYS	GGA Gly	CCC PRO	T A C T Y R	ACC Thr	GTC VAL	CTA Leu	C T G L E U	ACC THR	ACC THR	CCC PRO	ACC Thr	GC T A L A	C T C L E U	4 4 <b>6</b> L Y S	GTA VAL	G A C A S P	GGC Gly	ATC ILE	TCT Ser	GCÅ Ala	TGG TRP	ATA ILE	C A C H I S	GCC Ala	GC <b>C</b> Ala	
															Х				_										
A-MCF247	C A C H I S	GTA VAL	260 AAAA LYS	GCG Ala	GCG Ala	270 ACA Thr	ACC THR	CCT Pro	2 CCG PRO	GCC ALA	GGA Gly	ACA THR	GCA ALA	TCA SER	GGA Gly	300 CCG PRD	ACA THR	TGG TRP	3 A A G L Y S	ст[С] VAL	C A G G L N	C G T A R G	S20 TCT SER	C A A G L N	AAC ASN	5 3 U CCC PRO	TTA LEU	A A G L Y S	112
NFS XENO	CA-C HIS	GTA Val	AAG Lys	GCG Ala	GCG Ala	A C A T H R	ACT Thr	CCT Pro	CCA Pro	GCC Ala	GGA Gly	ACA THR	# GCA Ala					TGG TRP	A A G L Y S	GTT VAL	C A G G L N	C G T A R G	T C T S E R	C A A G L N	AAC ASN	CCC Pro	TTA LEU	A A G L Y S	
AKR ECO	CAC HIS	GT <b>C</b> Val	A A 🛱 L Y S	GCÅ Ala	GCG Ala	ACC Thr	ACS Thr	CCC PRO	CCG Pro	ATA 1LE	AAA Lys	CCA Pro	¥ Tca Ser					T G G T R P	AGA Arg	GTA VAL	CAA GLN	CGC Arg	TCT SER	C A A G L N	AAC ASN	CCT Pro	T T A L E U	AAA Lys	
	34	40			350			360																					

		s•	+ 0			550			360		
A-MC	F247	ΑΤΑ	AGA	TTA	ACC	CGT	GGG	GCC	CCC	TAA	
		ILE	ARG	LEU	THR	ARG	GLY	ALA	PRO	END	120
NFS	XENO	ATA	AGA	TTA	ACC	CGT	GGG	GCC	ccc	TAA	
		ILE	ARG	LEU	THR	ARG	GLY	ALA	PRO	END	
AKR	ECD	ATC	AGG	TTA	ACC	CGT	GGG	GCC	ccc	TAA	
		ILE	ARG	LEU	THR	ARG	GLY	ALA	PRO	END	

FIG. 2. Identification of MCF247-specific sequences in the 3' *pol* region. Comparison of the nucleotide and deduced amino acid sequences of AKR MCF247 (A-MCF247) MuLV DNA with those previously published for NFS-Th-1 xenotropic (NFS XENO) (22) and AKR ecotropic (AKR ECO) (12) MuLV proviruses is presented. Sequences unique to MCF247 DNA are enclosed within solid lines; the MCF247-specific 12bp insert is enclosed within heavy solid lines and indicated as X; differences in nucleotides and deduced amino acids in the xenotropic and ecotropic MuLV DNAs with respect to MCF247 DNA are shaded; the gaps in sequences indicate absent bases; the splice acceptor site is indicated with a broken line; the restriction sites are indicated with wavy lines; the asterisk indicates the position of initiation codon for the translation of the envelope precursor polypeptide in the *env* gene reading frame (see Fig. 3).

A-MCF247		# Atg Met	GAA Glu	GGT Gly	O CICA PRO	GCG Ala	T T C Phe	20 TCA SER	AAA Lys	CCC Pro	30 CTT LEU	AAA Lys	GAT ASP	AAG Lys	ATT ILE	AAC ASN	CCG Pro	50 TGG TRP	GGC Gly	CCC Pro	60 CTA LEU	ATA ILE	GTC VAL	CTG Leu	70 GGA GLY	ATC Ile	Т Т <mark>А</mark> L E U	80 ATA ILE	AGG ARG	GCA Ala	29
NFS XENO		# ATG MET	GAA Glu	GGT Gly	TCA SER	GCG ALA	T T C Phe	TCA SER	AAA LYS	CCC Pro	CTT LEU	AAA LYS	GAT ASP	AAG LYS	ATT ILE	AAC ASN	CCG Pro	TGG Trp	GGC Gly	CCC Pro	CTA LEU	ATA Ile	GTT Val	ATG NET	GG <b>e</b> Gly	ATC Ile	TT <b>\$</b> Leu	STS VAL	AGG ARG	GCA Ala	
AKR ECO		# ATG MFT	GAG	AGT	ACA The	ACG The	CTC LEU	TCA Ser		CCC PRO	TTT	AAA	AAT ASN	CAG GLN	GTT VAL	AAC ASN	CCG Pro	TGG TRP	GGC Gly	CCC Pro	CTA LEU	ATT ILE	GTC Val	CT¥ Leu	CTG LEU	ATT ILE	CTC LEU	SGA SLY	GGG GLY	GTC Val	
				<b></b>	70סו																										
MCF247	90 GGA		ŢĊĄ	GTA	dGA		GAC	AGC	CCT	C AT	120 CAG	GTC	TTC	1 AAT ASN	GTT	ACT	TGG	AGA	GTT	ACC	150 AAC				ATG MFT	1 ACA THR	60 GGA GLY		ACA THR	170 GCT ALA	57
XENO	GGA	GCC	TCG	GTA	CAA	CGT	GAC	AGC	CCT	CAC	CAG	ATC	TTC	AAT	GTT	ACT	TGG	AGA	GTT	ACC	AAC			CTA	ATG	ACA	GGA		ACA	GCT	•
ECO	AAC	CCC	GTT	ACG	TTG	AGA	AAC	AGC	cc¢	CAC	CAG	GTT	TTT	AAC	CTC	ACC	TGG	GAA	GTG	ACT	AAT	GGA	GAC	CGA	GAA	ACG	GTG	TEG	SCA	ATA	
	ACA	PRU		gp70	LEV	***		JEK	FRU .	<b>HI3</b>	GLN	VAL	PHE	ASN	1.20			919	VAL		ASN	ULT	ear	ARO		10.6	TAL		-	***	
MCF247	A AT ASN	GCT ALA	180 ACC THR	TCC Ser	CTC LEU	CTG LEU	GGG GLY	ACA THR	ATG MET	ACC THR	G A[] A S P	GC C ALA	210 TT PHE	CCT Pro	AAA LYS	СТ П LEU	TAC TYR	TTT Phe	GAC ASP	230 []] [EU	тыр суз	GAT ASP	TTA LEU				ATA ILE	GGG Gly	GAC ASP	GAC ASP	84
XENO	AAC ASN	GCC Ala	ACC THR	TCC Ser	CTC LEU	CTG LEU	GGG Gly	ACS Thr	ATG MET	ACA Thr	GAC Asp	ACC Thr	TT <b>C</b> Phe	CCT Pro	AAA Lys	CTA Leu	TAT Tyr	TTT Phe	GAC ASP	¢tg Leu	TG <b>t</b> Cys	GAT ASP	TTA LEU				STA Val	GG <b>&amp;</b> Gly	GAC ASP	TAC TYR	
ECO	ACC Thr	G <b>GC</b> Gly	AAT Asn	CAC HIS	CCT PRO	C T G L E U	TGG TRP	ACT Thr	TGG TRP	TGG TRP				CCT Pro	GAC ASP	C T€ L E U	ACA THR	CCA PRO	GAT Asp	CTC LEU	TGT Cys	A'TG Net	TT <b>G</b> Leu	GCC ALA	CTC LEU	CAC HIS	666 61 y	CCG PRO	TCC SER	TAT TYR	
													Sm	al													Sm	al			
MCF247	TGG TRP													GAT ASP	260 GAG GLU		Y	/		АЮТ [ТН R]	GGA Gly	270 CTC LEU	GOGG GLY	⊺G∏ CYS	24 CGC ARG	BO ACT THR	<u>CCC</u> Pro	<u>666</u> 617	290 GGA GLY	AGA Arg	98
XENO	TGG TRP													GAT ASP	GAC ASP	CCA PRO	GAA Glu	CCC PRO	GAT ASP	ATT	GGSE GLY	GAT	GG¶ Gly	т G <b>C</b> с Y S	CGC ARG	ACT THR		666 61 y	GGA Gly	AGA ARG	
ECO	TGG TRP	SGC GLY	CTA LEU	GAA GLU	TAT	CGG	GCT ALA	CCT	TTT	TCT	CCT	222	<u>CC6</u> P80	6 <b>66</b>	CCC	CCC	TSC	TGT	TCA		GGA GLY	AGC	AGC Ser	SAC ASP	TCC See	ACSS THR	CCA PRD	GG€C GLY	TGT CVS	TCC	
MCF247	A AA L Y S	300 AGG ARG		AGA ARG	ACA THR	१० गीग शिमही	GAC ASP																								105
XENO	464	AGG	*CA																												
	ARG	ARG	THE	AGA	LEU	TYP	4SP																								
ECO	ARG	ARG GAT	THE	AGA ARG GAO	CIG LEU GAG	TYR	CTC	ACT	TCA	TAT	ACT	000	CGG	TEC	AAT	ACS	GCC	TGG	AAC	AGA	CTT	AAG	TTA LEU	TCT	444 1 VS	STG	ACA	CAT	GCA	CAC	
ECO	ARG AGA ARG	ARG GAT ASP	THR TGT CYS	AGA ARG GAO GLU	CIG LEU GAG GLU	TYR CCC PRD	ASP CTE	ACT THR Sm	TCA SER	TAT TYR	ACT THR	CCC PRO	CGE ARG	TBC CYS	AAT Asn	ACG THR	GCC ALA	TGG TRP	AAC ASN	AGA ARG	CTT LEU	AAG LYS	TTA LEU	TCT SER	AAA LYS	GTG VAL	ACA Thr	CAT HIS	GCA ALA	CAC HIS	
ECO MCF247	ARG	ARG GAT ASP	THR TGT CYS	AGA ARG GAS GLU TTC PHE	CIG LEU GAG GLU 320 TAT TYR	TYR CCC PRD	CTS CTS LEV	ACT THR 330 CCC PRO	TCA SER a I GGG GLY	TAT TYR CAT HIS	ACT THR 3' ACT THR	CCC PRO GTA VAL	CCA PRO	TSC CYS ACA	AAT ASN 350 GGG GLY	ACG THR	GCC ALA	TGG TRP	AAC ASN TGT CYS	AGA ARG GGA GLY	CTT LEU 360 GGG GLY	AAG LYS CCG Pro	TTA LEU AGA	TCT SER 37 GAG GLU	AAA LVS GGC GLY	STS VAL TAC TYR	ACA THR TGT CYS	CAT HIS 380 GGC GLY	GCA ALA	CAC HIS TGG TRP	129
ECO MCF247 XENO	ARG	ARG GAT ASP	THE TOT CYS	AGA ARG GAS SLU TTC PHE TTC PHE	CIG LEU GAG SLU 320 TAT TYR TAT TYR	GTT VAL	CTE LEV TGC CYS TGC CYS	ACT THR 330 CCC PRO CCC PRO	TCA SER a I GGG GLY GGT GLY	TAT TYR CAT HIS CAT	ACT THR 3' ACT THR ACT THR	CCC PRO GTA VAL GTA VAL	CCA PRO CCA PRO	ACA THR ATA	AAT ASN 350 GGG GLY GGG GLY	ACG THR	GCC ALA	TGG TRP	AAC ASN TGT CYS TGT CYS	AGA ARG GGA GLY GGA	CTT LEU 360 GGG GLY GGG	AAG LYS CCG PRO CCG PRO	TTA LEU Aga Arg Gga Gly	TCT SER GAG GLU GAG GLU	AAA LYS GGC GLY GGC	STG VAL TAC TYR TAC	ACA THR TGT CYS TGT CYS	CAT HIS 380 GGC GLY GGC	GCA ALA LYS AAA	CAC HIS TGG TRP TGG TRP	129
ECO MCF247 XENO ECO	ARG AGA ARG ARG	ARG GAT ASP	THE TGT CYS	AGA ARG SAS SLU TTC PHE TTC PHE	CIG LEU GAG SLU 320 TAT TYR TAT TYR	GTT VAL GTC VAL	CTC CTC LEU TGC CYS TGC CYS	ACT THR 330 CCC PRO CCC PRO	ICA SEI a I GGG GLY GCI GLY GGI GLY	TAT TYR CAT HIS CAT HIS CCA	ACT THR ACT THR ACT THR ACT THR GAT HIS	CCC PRO GTA VAL GTA VAL CCC	CCA PRO CCA PRO CCA PRO	ACA THR ATA ILE CGG	AAT ASN 350 GGG GLY GGG GLY TGG	ACG THR GCC	GCC ALA CGG	TGG TRP TCA	AAC ASN TGT CYS TGT CYS TGT	AGA ARG GGA GLY GGA GLY GGT	CTT LEU 360 GCC GLY GCC GLY GCT	AAG LYS CCG PRO CCG PRO CCA	TTA LEU Aga Arg Gga Gly Gaa	TCT SER GAG GLU GAG GLU TCC SEP	AAA iys GGC GLY GGC GLY ITC PH#	GTG VAL TAC TYR TAC TYR TAC	ACA THR TGT CYS TGT CYS TGT CYS	CAT HIS 380 GGC GLY GGC GLY GCC	GCA ALA LYS AAA LYS TCT	CAC HIS TGG TRP TGG TRP	129
ECO MCF247 XENO ECO	ARG ARG ARG ARG	ARG GAT ASP GGA GLY	THR TGT CYS GGA GLY	AGA ARG SLU TTC PHE TTC PHE TTC PHE	CIG LEU GAG SLU 320 TAT TYR TAT TYR TAT TYR	GTT VAL GTC VAL	CYS CYS TGC CYS TGC CYS TGC CYS	ACT THR 330 CCC PRO CCC PRO CCC PRO	ICA   SER   a   GGG   GLY   GGY   GLY   GGY   GLY	TAT TYR CAT HIS CAT HIS CCA PRO	ACT THR ACT THR ACT THR ACT THR ACT THR ACT THR ACT	CCC PRO GTA VAL GTA VAL CSC ARS	CCA PRO CCA PRO CCA PRO CCC PRO	ACA CYS THR ATA ILE CGG ARG	AAT ASN 350 GGG GLY GGG GLY TGG T2P	ACG THR GCC ALA	GCC ALA CGG ARG	TGG TRP TCA SER	AAC ASM TGT CYS TGT CYS TGT CYS	AGA ARG GLY GLY GLY GGIY GLY	<b>CTT</b> <b>LEU</b> 360 GGG GLY GGG GLY GGT GLY	AAG LYS CCG PRO CCG PRO CCA PRO	TTA LEU Aga Arg Gga Gly Gaa Gly Gaa	TCT SER GAG GLU GAG GLU TCC SER	AAA LYS GGCC GLY GGCY GLY TTC PHE	STG VAL TAC TYR TAC TYR TAC TYR	ACA THR TGT CYS TGT CYS TGT CYS	CAT HIS 380 GGC GLY GCC GLY GCC ALA	GCA ALA LYS AAA LYS TCT SER	CAC HIS TGG TRP TGG TRP TGG TRP	129
ECO MCF247 XENO ECO MCF247	ARG AGA ARG ARG ARG ARG ARG ARG ARG ARG	ARG GAT ASP GGA GLY TGT CYS	THE THE CVS CVS	AGA ARG CA ELU TTC PHE TTC PHE TTC PHE CA CC THR	CIG LEU GAG SLU 320 TAT TYR TAT TYR TAT TYR TAT TYR OO ACT THR	GTT VAL GTT VAL GTC GGA GLY	TGC CYS TGC CYS TGC CYS TGC CYS	ACT THR Sm 330 CCC PRO CCCC PRO	TCA SER a I GGG GLY GGG GLY TAC TYR	TAT TYR CAT HIS CAT HIS CAT HIS CCA TRP	ACT THR 3 ACT THR ACT THR ACT THR 420 AAG LYS	CCC PRO GTA VAL GTA VAL CCCA PRO	CCA PRO CCA PRO CCA PRO CCA PRO	ACA CYS ACA THR ATA ILE CGG ARG CGG SER	AAT ASN 350 GGG GLY TGG GLY TGG TRP 30 TCA SER	ACG THR GCC ALA	GCC ALA CGG ARG	TGG TRP TCA SER 440 CTA	AAC ASN TGT CYS TGT CYS TGT CYS	AGA ARG GGA GLY GGT GLY TCCC SER	<b>CTT</b> <b>LEU</b> 3600 GLY GGG GLY GGT GLY 4500 CTT LFU	AAG LYS CCCG PRO CCCA PRO CCA	TTA LEU Aga Arg Gga Gga Gly Gaa Gly Cga Arg	TCT SER 3: GAG GLU GAG GLU TCC SER 4, GGA GLY	AAA iYS GGC GLY GGC FHE 60 AACC	STS VAL TAC TYR TAC TYR TAT TYR	ACA THR TGT CYS TGT CYS TGT CYS	CAT HIS 380 GGC GLY GGC GLY GGC GLY 470 GCAG GIN	GCA ALA LYS AAA LYS TCT SER	CAC HIS TGG TRP TGG TRP TGG TRP CAG GLN	129
ECO MCF247 XENO ECO MCF247 XENO	ARGA ARGA ARGA ARGA ARGA ARGA ARGA ARGA	ARG GAT ASP GEA GLY TGT CYS TGT CYS	THE TOTECYS Cars Gaag Glu Gaag Glu	AGA ARG GAS GLU TTC PHE TTC PHE TTC PHE TTC PHE ACC THR ACC	CIG LEU SAG SLU 320 TAT TYR TAT TYR TAT TYR TAT TYR DO ACT THR	GTT VAL GTT VAL GTC GTC VAL GTC GGA GLY GGA	CAC ASP CTS LEV TGC CYS TGC CYS TGC CYS CAG GLN CAG GLN	ACT THR SM 3300 CCC PRO CCC PRO CCC PRO	TCA SER a I GGC GLY GGT GLY TAC TYR	TAT TYR CAT HIS CAT HIS CAT HIS CCA TRP	ACT THR 3 ACT THR ACT THR ACT THR 420 AAG LYS	CCC PRU GTA VAL GTA VAL CCC ARG CCCA PRO	CCA ARG CCA PRO CCA PRO CCC PRO TCA SER TCA	TISC CYS ACA THR ATA ILE CGG ARG CGG SER TCA SER	AAT ASN 350 GGG GLY TGG GLY TGG TTG STCA SER	ACS THR GCC ALA TGG TRP	GACC ALA CGG GACC ARG GACC ASP	TGG TRP TCA SER 440 CTA LEU CTA	AAC ASN TGT CYS TGT CYS TGT CYS ATT ILE	AGA ARG GLY GGA GLY TCCC SER TCCC SER	<b>CTT</b> <b>LEU</b> 360 GGG GLY GGG GLY 450 CTT LEU CTT	AAG LYS CCCG PRO CCCA PRO CCCA PRO AAG LYS	TTA LEU Agga Gga Gga Gly Gaa Arg Cga Arg Cga Cga Cga	TCT SER GAG GLU GAG GLU TCC SER GGA GLY GGA	AAA <b>LYS</b> 70 GGC GLY FHE 60 AASN AACN	STG VAL TAC TYR TAC TYR TAT TYR ACCC THR ACT	ACA THR TGT CYS TGT CYS TGT CYS CCTS PRO	CAT HIS 380 GGC GLY GGC GLY GGC GLY GGC GLY GGC GLY	GCA ALA LVS AAA LVS TCT SER AAAT ASN GAT	CAC HIS TGG TRP TGG TRP TGG TRP CAG GLN CAG	129
ECO MCF247 XENO ECO MCF247 XENO ECO	ARGA ARGA ARGA ARGA ARGA ARGA ARGA ARGA	ARG GAT ASP GEA GEA TGT CYS TGT CYS	THE THE CYS GAG GLU GAA GLU GAA GLU	AGA ARG GAG GAG GAG GAG C PHE TTC PHE TTC PHE C C C THR ACCC THR	CIG LEU GAG SLU 320 TAT TYR TAT TYR TAT TYR TAT TYR CO ACT THR ACT	GTT VAL GTT VAL GTC VAL GGGA GLY GGGA GLY	CACCASS CYS TGCCYS TGCCYS CAGGLN CAGGLN CAGGLN	ACT THR Sm 330 CCC PRO CCC PRO CCC PRO CCC CCC CCC CCC CCC CCC CCC CCC CCC C	TAC TAC GG GG GG GG TAC TAC TAC TAC TAC TAC TAC TAC	TAT TYP CAT HIS CAT HIS CAT HIS CCA TRP TGG TRP TGG TRP	ACT THR 3.7 ACT THR ACT THR <b>GAT</b> HIS 420 AAG LYS AAA	CCCC PRO GTA VAL GTA VAL CCCC ARS CCCA PRO CCCA PRO	¢GS ARG PRO CCA PRO CCA PRO CCA PRO CCA PRO TCA SER TCA SER TCA	ACA CYS ACA THR ATA ILE CGG ARG CGG SER TCA SER TCA SER	AAT ASN 350 GGG GLY GGG GLY TGG GLY TGG SC TCA SER TCA SER TCA	ACS THR GCCC ALA TGG TRP TGG TRP	GCC ALA CGG ARG GAC ASP GAC GACP	TGG TRP TCA SER G400 CTA LEU CTA LEU CTA LEU TAC	AAC ASM TGT CYS TGT CYS TGT CYS ATT ILE ATT ILE	AGA ARG GLY GGA GLY TCCC SER TCCC SER ACA	<b>CTT</b> <b>3</b> 60 <b>G</b> GG <b>G</b> LY <b>G</b> GG <b>G</b> LY <b>4</b> 50 <b>CTT</b> <b>LEU</b> <b>CTT</b> <b>LEU</b> <b>CTT</b> <b>LEU</b> <b>CTT</b>	AAG LYS CCCG PRO CCCG PRO CCCA PRO CCCA LYS AAG LYS	TTA LEU Aga Arg Gga Gly Gaa Gly Cga Arg Cga Arg Cga Arg	TCT SER 33 GAG GLU GAG GLU TCC SER 4 GGA GLY GGA GLY AAT	AAA LYS GGCC GGCY GGCY TTCC ASN AACC ASN AACC ASN	STS VAL TAC TYR TAC TYR TYR ACC THR ACC THR	ACA THR TGT CYS TGT CYS TGT CYS CCT PRO CCT PRO	CAT HIS BGCC GLY GGCC ALA GAG GLN GLN LYS	GCA ALA AAA LYS AAA LYS TCT SER AAA LYS SER AAA LYS SER AAA SOAT ASP GAT	CAC HIS TGG TRP TGG TRP TGG TRP CAG GLN CAG GLN CAG	129
ECO MCF247 XENO ECO MCF247 XENO ECO	ARG ARAT ARG ARG ARG ARG G G C C C C C C C C C C C C C C C C C	ARG GAT ASP GEA GEY TGT CYS TGC CYS	THR TGYS GAAG GLU GAA GLU	AGA ARG SAS TTC PHE TTC PHE TTC PHE C C THR ACC THR ACC THR	CIE LEU GAG SLU TAT TYR TAT TYR TAT TYR TAT TYR ACT THR ACA THR	GTT YR GTT VAL GTT VAL GTC VAL GTC VAL GGGA GLY GGGA GLY	CACCASE CYS TGCCYS TGCCYS TGCCYS CAG GLN CAG GLN CGA	ACT THR Sm 330 CCC PRO CCC PRO CCC PRO CCC PRO GCA ALA GCA ALA	TAC GGE GLY GGT GLY TAC GLY TAC TYR TAC TYR TAC TYR TAC TYR TAC TYR TAC	TAT TYP CAT HIS CAT HIS CAT HIS CAT HIS CAT HIS TGG TRP TGG TRP	ACT THR 3 ACT THR ACT THR 420 AAG LYS AAAA LYS	CCCC PRO GTA GTA VAL GTA VAL CCCC ARG CCCA PRO CCCA PRO	CCA PRO CCA PRO CCA PRO TCA SER TCA SER TCC SER	TEC CYS AGA THR ATA ILE CGG ARG CARG CARG CARG CARG CARG CARG C	AAT ASH 350 GGG GLY GGG GLY TGG GLY TTRP 30 TCA SER TCA SER TCC SER	ACS THR GCC ALA TGG TRP TGG TRP TGG TRP	GCC ALA GAC GAC GAC ASP GAC ASP GAC ASP	TGG TRP TCAR SER LEU CTAU LEU CTAU LEU TTYR	AAC ASN TGT CYS TGT CYS ATT ILE ATT ILE	AGA ARG GLY GGA GLY GGA GLY TCCC SER TCCC SER ACA THR	<b>CTT</b> <b>3</b> 60 <b>GGG</b> <b>GLY</b> <b>GGG</b> <b>GLY</b> <b>GGG</b> <b>GLY</b> <b>GGG</b> <b>CTT</b> <b>LEU</b> <b>CTT</b> <b>LEU</b> <b>STA</b>	AAG LYS CCCG PRO CCCG PRO CCA PRO AAG LYS AAG SER	TTA LEU AGA ARG GLY CGA ARG ARG ARG ARG ARG ARG ARG	TCT SER 33 GAG GLU GAG GLU SER 44 GGA GLY GGA GLY AAT	AAA ivs GGCY GGCY GGCY PHE 60 AAACN AAACN AAACN AAACN AAACN EEU	GTG VAL TAC TYR TAC TYR ACCC THR ACC THR ACC	ACA THR TGT CYS TGT CYS TGT CYS CYS CYS CYS CYS CYS CYS CYS CYS CYS	CAT HIS GGC GLY GGC GLY GGC GLY GGC GLY GAG GLY ALA AAG LYS	GCA ALA LVS TCT SER AAAT ASP GAT ASP	CAC HIS TGG TRP TGG TRP TGG TRP CAG GLN CAG GLN CAG GLN	129
ECO MCF247 XENO ECO MCF247 XENO ECO	ARA ARA ARA ARA ARA ARA ARA ARA ARA ARA	ARG GAT ASP GEA GEY TGT CYS TGT CYS TGC CYS	THR TGT CYS GAG GLU GAG GLU GAG GLU TGT	AGA ARG SAS TTCC PHE TTCC PHE TTCC PHE TTCC PHE 4( ACCC THR ACCC THR ACCC THR ACCC THR ACCC THR ACCC THR ACCC THR ACCC TTCC THR ACCC ACCC ACCC ACCC ACCC ACCC ACCC AC	CIE LEU GAG GAG GAG GAG TYR TYR TYR TYR TYR TYR TYR TYR	GTT YAL GTT VAL GTC GGA GLY GGA GLY TCCC	CASP CERE TGCS CCSS CASG CASG CASG CASG CASG CASG C	ACT THR Sm 3300 CCC PRO CCC PRO GCA ALA GCA ALA GCA ALA	TCA SER a l GGG GLY GGT GLY TAC TYR TAC TYR TAC TYR TAC SER	TAT TYP CAT HIS CAT HIS CCA TRP TGG TRP TGG TRP TGG TRP TGG TRP	ACT THR 3 ACT THR ACT THR ACT THR 420 AAG LVS AAG LVS S10 AGT	CCCC PRO GTA VAL GTA VAL CCCA PRO CCA PRO CCA PRO CCA	CCA ARG CCA PRO CCA PRO CCC PRO TCA SER TCA SER TCA SER	ACLA CYS ACLA THR ATA JLE CGG ARG 4. TCA SER TCA SER TCA SER TCA SER TCA SER	AAT ASN GGG GLY GGG GLY TGG GGLY TTGG SER TCA SER TCA SER TCA SER TCA	ACG THR GCCC ALA TGG TRP TGG TRP TGG TRP	GCC ALA GACA GACC ASP GACC ASP GACC ASP GACC ASP	TGG TRP TCA SER 440 CTA LEU CTA LEU TAC TYR 530	AAC ASN TGT CYS TGT CYS ATT ILE ATT ILE	AGA ARG GLY GGA GLY GGLY TCC SER TCC SER ACA THR	<b>CTT</b> <b>J</b> <b>G</b> <b>G</b> <b>G</b> <b>G</b> <b>G</b> <b>G</b> <b>G</b> <b>G</b>	AAG LYS CCCG PRO CCCG PRO CCCA PRO CCCA PRO CCCA LYS AAG LYS AAG SER	AGA ARG GGA GGA GGA ARG GLU CGA ARG CGA ARG ARG ARG ARG ARG CGA ARG CGA ARG CGA CGA ARG CGA CGA CGA CGA CGA CGA CGA CGA CGA CG	TCT SER GAG GLU GAG GLU TCC SER GGA GLY GGA GLY AAT	AAA LYS GGCY GGCY GGCY TTC GGCY TTC GGCY AAC ACN AACN AACN ACN CTA LEU 540	GTG VAL TAC TYR TAC TYR TAT TYR ACC THR ACC THR	ACA THR TGT CYS TGT CYS TGT CYS CCT PRO CCT PRO CCT PRO CCT PRO	CAT HIS GGC GLY GGC GLY GGC GLY GGC GLY GGC GLY GGC GLN AAG LYS	GCA ALA LYS AAA LYS TCT SER ASP GAT ASP GAT ASP GAT ASP GAT ASP	CAC HIS TGG TRP TGG TRP TGG GLN CAG GLN CAG GLN GTC	129
ECO MCF247 XENO ECO MCF247 XENO ECO	ARGA ARGA ARGA ARGA ARGA ARGA ARGA ARGA	ARG GAT ASP GEA GEV TGT CYS TGC CYS CCS CYS CCC CYS	GAG GLU GAG GLU TGT CYS TGT	AGA ARA ARA ARA ARA ARA ARA ARA ARA ARA	GAGE GAGE	GTT TYR PRD GTT VAL GTT VAL GTT VAL GTT VAL GTT VAL GTT VAL GTT VAL GTT VAL GTT VAL CTT CTT CTT CTTT CTTTT CTTTTT CTTTTTTTT	CASP CTGCS TCCSS TCCSS CASP CCSS CASP CASP CASP CASP CASP	ACT THR SM0 CPR0 CCCC PR0 CCCC PR0 CCCC PR0 CCCC PR0 CCCC PR0 CCCC CCCC	TCA SER al GGUY GGUY TAC TVR TAC TVR TAC TVR GGUY TAC TVR GGUY GGUY TAC TVR GGUY TAC TVR TAC TVR GGUY TAC TVR GGUY	TAT TYR CAT HIS CAT HIS CAT HIS CCA PRO TGG TRP TGG TRP TGG TRP TGG TRP TGG TRP	ACT THR 3 ACT THR ACT THR ACT THR 4 20 GAT HIS 4 20 GAT LYS 510 AGT SER AGT	CCC PRU GTA VAL GTA VAL CCCA PRO CCCA PRO CCCA PRO CCCA PRO CCCA CCCA CCCA CCCA CCCA CCCA CCCA CC	CGG ARG PRO CCA PRO CCA PRO CCC PRO TCA SER TCA SER TCA SER TCA SER TCA	TEC CYS ACA TTHR ATA TTHR ATA SER CARG SER SER SER SER SER SER SER SER SER SER	AAT ASN 350 GGGG GGG GGG TGG GGLY TGG SER TCA SER TCA SER TCA SER TCA SER TCA SER TCA SER GGG GLY GGT	ACG THR GCCC ALA TGG TRP TGG TRP TGG TRP GCCC ALA GCC	GCC ALA CGGG CARG GACC ASP GACC ASP GACC ASP ACA THR ACA	TGG TRP TCA SER 4400 LEU LEU LEU LEU LEU TAC TYR 5300 CCCG CCCG	AAC ASN TGT CYS TGT CYS TGT CYS ATT ILE ATT ILE ATT ILE	AGA ARG GLY GGA GLY TCCC GLY TCCC SER TCCC SER ACA THR	360   GGG	AAG LYS CCG PRO CCG PRO CCA PRO CCA PRO CCA LYS AAG LYS AAG SER	TTA LEU AGA ARG GGA GGA GGA CGA ARG CGA ARG CGA ARG ARG ARG CGA ARG CGA ARG CGA CGA CGA CGA CGA CGA CGA CGA CGA CG	TCT SER 3 GAG GLU GAG GLU TCC SER 4,A GGA GLY GGA GLY AAT ASN	AAA LYS GGC GGC GGC GGC GGC TTC FPHE 60 AASN AASN AASN AASN AASN AASN AASN AAS	GTG VAL TAC TYR TAC TYR ACC THR ACC THR TGC CYS	ACA THR TGTS TGYS TGYS TGYS CYS CYS CYS CYS CYS PRO CCT PRO TCA SER AATA	CAT HIS 380 GGC GGC GGC GGC GGC GGC GGC GGC GGC GG	GCA ALA LYS AAA LYS TCT SER ASP GAT ASP 50 CTA LEU CTG	CAC HIS TGG TRP TGG TRP TGG TRP CAG GLN CAG GLN GTC VAL GTC	1 2 9
ECO MCF247 XENO ECO MCF247 XENO ECO ECO	ARGA ARGA ARGA ARGA ARGA ARGA ARGA ARGA	ARG GAT ASP GLY TGT CVS TGT CVS TGC CVS TGC CVS CCS PRO	GAG GAGU GAGU GAA GLU TGT CYS	AGA ARG GAS FLU TTC PHE TTC PHE TTC PHE 4( ACC THR ACC THR ACC THR 4( TTYR TAT TYR	GAG GAG GAG GAU TYR TYR TYR TAT TYR TAT TYR ACT THR ACT THR ACA THR GGAT ASP	GTT TYR GCC PRO GCT VAL GTT VAL GTT VAL GCC VAL GCC CVAL GCC CCC SER TCCC SER	TGC SC	ACT THR S S S S S C C C C C C C C C C C C C	TACA SER GLY GGT GLY TAC TYR TAC TYR SER GTC VAL	TAT TYR CAT HIS CAT HIS CAT HIS CAT TRP TGG TRP TGG TRP TGG TRP TCC SER TCC SER	ACT THR 3-CT ACT HR ACT HR 4 20 GAT S 4 20 CAT S 4 20 CAT S 5 10 1 S 5 10 1 S CAT S C	CCCA GTA GTA VAL GTA VAL CCCA PRO CCCA PRO CCCA PRO GAC CCA GAC CCA CCA CCA CCA CCA CCA CCA	CCA PRO CCA PRO CCA PRO CCC PRO CCC PRO TCA SER TCA SER TCCA SER ATCC SER	ACA THR ATA ATA ATA ATA AAGA SER CAAGE CAGE CAGE CAGE	AAT ASM GGG GGG GGG GGC GGC TGG GGC TTGG TCA SER TCA SER TCA SER TCA SER CGGC GGC GGC GGC GGC GGC GGC GGC GGC G	ACS THR GCC ALA TGG TRP TGG TRP GCCC ALA GCC ALA	GCC ALA CGG ARG GACC ASP GACC ASP GACC ASP ACA THR ACA THR ACA	TGG TRP TCA SER 440 CTA LEU CTA LEU CTA LEU CTA LEU CTA CCA TYR 530 CCCG PRO CCA	AACC ASN TGT CYS TGT CYS TGT CYS ATT ILE ATT ILE ILE	AGA ARG GLY GGA GLY TCCC SER ACA THR	3600 GGG   3600 GGG   GGG	AAG PRO CCG PRO CCG PRO CCA PRO CCA PRO CCA PRO CCA PRO CCA PRO CCA PRO CCA PRO CCA PRO CCA PRO CCA CA PRO CCA CA CA CA CA CA CA CA CA CA CA CA CA	TTA LEU AGA ARG GAA GLU CGAA ARG CGAA ARG CGAA ARG CGAA ARG CGAA ARG CGAA ARG CGAA ARG CGAA ARG CGAA ARG CGAA ARG CGAA CGAA	TCT SER 3: GAG GLU GAG GLU TCC SER 4, GGA GLY AAT ASN GAG	AAA LYS GGC GGC GGC GGC GGC GGC GGC GGC GGC GG	GTG VAL TACC TYR TATTYR ACCC THR ACCC THR ACCC THR TGC CYS TGCC	ACA THR TGT CYS TGT CYS TGT CYS CCT PRO CCT PRO CCT <b>TCR</b> ASN AAC	CAT HIS GGC GGC GGC GGC GGC GGC GGC GGC GGC GG	GCA ALA ALA LYS AAAA LYS <b>TCT</b> SER AAST GAT AST GAT ASP GAC ASP CTA LEU CTG LEU CTG LEU	CAC HIS TGG TRP TGG TRP TGG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG CAC	129
ECO MCF247 XENO ECO MCF247 XENO ECO	ARA AGA ARA ARA ASA 390CG LY 480CGLY 480CGLY GGLY	ARG GAT ASP GLY TGT CYS TGT CYS TGC CYS CCS CYS CCS CYS CCC PRO	GAAGUU GAAGUU GAAGUU GAAGUU TGTTCYS	AGA AGA AGA AGA AGA AGA AGA TTTC PHE TTC PHE TTC PHE AGC THR ACC THR ACC TTAT TYR	GAG SLU 320 TAT TYR TAT TYR TAT TYR TAT TYR ACT THR ACT THR ACT THR ACA SP GAT ASP	GTT TYR GTT VAL GTT VAL GTT VAL GGA GLY GGA GLY TCC SER TCC SER	ASP CYS TGCS TGCS TGC CYS TGC CYS CAG GLN CAR CAR CAR CAR CAR CAR CAR CAR	ACT THR Sm S300 CCCC PRO CCCC PRO CCCC PRO CCCC PRO SCA ALA ALA ALA SCA ALA CCA ALA	TAC SER GLY GGT GLY TAC TYR TAC TYR TAC TYR SER GTC VAL GTC VAL	TAT TYR CAT HIS CAT HIS CAT HIS CAT HIS CCA TRP TGG TRP TGG TRP TGG TRP TGG SER	ACT THR ACT THR ACT THR ACT THR 420 AAG LYS AAAA LYS S10 ACT SER AGT SER	CCC PRO GTA VAL GTA VAL CCC ARG PRO CCA PRO CCA PRO CCA PRO GAC CCA PRO GAC	¢¢¢ ARG PRO CCA PRO CCA PRO CCC PRO TCA SER TCA SER TCA SER TCA SER TCA	AGA AGA ITHR ATA ATA ILE CGG ARG CGG SER TCA SER TCA SER TCA SER TCA SER CAG GLM	AAT ASM GGGG GLY GGGG GLY GGGG TGG GGG SER TCA SER TCA SER TCA SER TCA SER CGGG GLY GGGY GGT GLY	ACG THR GCCC ALA TGG TRP TGG TRP TGG TRP GCCC ALA GCCA ALA	GCC ALA CCGG ALA GACAASP GACCAASP GACCAASP ACAASP ACAASP ACAASP ACAASP ACAASP ACAASP ACAASP ACAASP ACAASP ACAASP THR	TGG TRP TCA SER 440 LEU CTA LEU CTA LEU CTA CCA PRO CCG PRO CCA PRO	AAC ASN TGT CYS TGT CYS TGT CYS ATT ILE ATT ILE ATT ILE STA	AGA ARG GLY GGA GLY GGA GLY TCCC SER TCCC SER ACA THR TECCYS	360 <td>AAG LYS CCG PRO CCG PRO CCA PRO CCA LYS AAG LYS AAG GLY GGLY GGT GLY</td> <td>TTA LEU AGA ARG GGA GGA CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG ARG ARG ARG ARG ARG ARG ARG CGA ARG C C C C C C C C C C C C C C C C C C C</td> <td>TCT SER GAG GLU GAG GLU GAG GLY GGA GLY AAT ASN GAG GLU</td> <td>AAA LYS GGCY GGCY GGCY FHE GGCY FHE GGCY</td> <td>GTGCYS</td> <td>ACA THR TGT CYS TGT CYS TGT CYS CCT PRO CCT PRO CCT PRO CCT PRO CCT SER AAM</td> <td>CAT HIS GGC GLY GGC GLY GGC GLY GGC GLY ALA CCCC PRO CCCC PRO SER</td> <td>GCA ALA ALA LYS AAA LYS TCT SER AASN GAT ASN GAT ASN CTA LEU</td> <td>CAC HIS TGG TRP TGG TRP TGG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG TC CAG TC CAG TC CAG TRP</td> <td>129</td>	AAG LYS CCG PRO CCG PRO CCA PRO CCA LYS AAG LYS AAG GLY GGLY GGT GLY	TTA LEU AGA ARG GGA GGA CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG CGA ARG ARG ARG ARG ARG ARG ARG ARG CGA ARG C C C C C C C C C C C C C C C C C C C	TCT SER GAG GLU GAG GLU GAG GLY GGA GLY AAT ASN GAG GLU	AAA LYS GGCY GGCY GGCY FHE GGCY FHE GGCY	GTGCYS	ACA THR TGT CYS TGT CYS TGT CYS CCT PRO CCT PRO CCT PRO CCT PRO CCT SER AAM	CAT HIS GGC GLY GGC GLY GGC GLY GGC GLY ALA CCCC PRO CCCC PRO SER	GCA ALA ALA LYS AAA LYS TCT SER AASN GAT ASN GAT ASN CTA LEU	CAC HIS TGG TRP TGG TRP TGG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG TC CAG TC CAG TC CAG TRP	129
ECO MCF247 XENO ECO MCF247 XENO ECO MCF247	ARGA ARGA ARGA ARGA ARGA ARGA ARGA ARGA	ARG GAT ASP GLY TGT CYS TGT CYS TGC CYS TGC CYS CCS CYS CCS CYS CCS CYS CCS CYS CCS CYS CCS CYS CCS CYS CCS CYS CCS CC	GAGG GLU GAAGGLU GAAGGLU GAAGGLU GAAGGLU TGT CYS TGT CYS TGT CYS	AGA GAS GAS TTC PHE TTC PHE TTC PHE TTC PHE CACC THR ACC THR ACC THR TAT TYR	CIE	GTT TYR GTT VAL GTT VAL GTT VAL GGA GLY GGA GLY TCCC SER TCCC	ASP CYS TGCS TGCS TGCS TGCS CAG GLN CAR CAR CAR SER TCAR	ACT THR Sm Signal Signa	TAC SER GLY GGT GLY TAC TYR TAC TYR GTC GTC VAL GTC VAL	TAT TYR CAT HIS CAT HIS CAT HIS CCA TRP TGG TRP TGG TRP TGG TRP TGG TRP TGG TRP TGG TRP	ACT THR ACT THR ACT THR ACT THR ACT THR 420 AAG LYS S10 AAG SER	CCC PRO GTA VAL GTA VAL GTA VAL CCC ARG PRO CCA PRO CCA PRO GAC CCA PRO GAC CCA	¢¢¢ ARG PRO CCA PRO CCA PRO CCA PRO TCA SER TCA SER TCA SER TCA SER TCA	AGA AGA ITHR ATA ATA SER CGG ARG SER TCA SER TCA SER CAG GLN	AAT ASM 350 GGG GLY GGG GLY GGG TGG GGG SER SER TCA SER TCA SER TCA SER CGG GLY GGG GLY GGG GLY	ACG THR GCCC ALA TGG TRP TGG TRP TGG TRP GCCC ALA GCCA ALA	GCC ALA CCGG ALA GACA GACC ASP GACC ASP GACC ASP ACAA ASP ACAA THR ACA THR ACA THR	TGG TRP TCA SER 440 LEU CTA LEU CTA LEU CTA CCG PRO CCG PRO	AAC ASN TGT CYS TGT CYS TGT CYS ATT ILE ATT ILE ATT ILE ATT CYS	AGA ARG GLY GGA GLY GGA GLY TCCC SER TCCC SER TCCC SER TCCC SER TCCC SER TCCC SER TCCC SER	360 <td>AAG LYS CCG PRO CCG PRO CCA PRO CCA LYS AAG CLYS GGG GLY GGG GLY</td> <td>TTA LEU AGA GGA GGA GGA GGA CGA ARG GGA CGA ARG CGA ARG CGA ASN GGT CGA SN CGA SN</td> <td>TCT SER GAG GLU GAG GLU GAG GLY GGA GLY AAT ASN GAG GLU</td> <td>AAA LYS GGCY GGCY GGCY FHE GGCY FHE GGCY</td> <td>GTG VAL TAC TYR TAC TYR TAT TYR ACC THR ACC THR TAC THR TGC CYS TGC CYS</td> <td>ACA THR TGT CYS TGT CYS TGT CYS CCT PRO CCT PRO CCT PRO CCT PRO CCT SER AAM</td> <td>CAT HIS GGC GLY GGC GLY GGC GLY GGC GLY GGC GLY GGC GLY ALA SCCCC PRO CCCC PRO SER</td> <td>GCA ALA ALA LYS AAAA LYS <b>TCT</b> SER AASN GAT ASN GAT ASN CTA ASP CTA LEU TTA LEU</td> <td>CAC HIS TGG TRP TGG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG CAC TRP</td> <td>129</td>	AAG LYS CCG PRO CCG PRO CCA PRO CCA LYS AAG CLYS GGG GLY GGG GLY	TTA LEU AGA GGA GGA GGA GGA CGA ARG GGA CGA ARG CGA ARG CGA ASN GGT CGA SN CGA SN	TCT SER GAG GLU GAG GLU GAG GLY GGA GLY AAT ASN GAG GLU	AAA LYS GGCY GGCY GGCY FHE GGCY FHE GGCY	GTG VAL TAC TYR TAC TYR TAT TYR ACC THR ACC THR TAC THR TGC CYS TGC CYS	ACA THR TGT CYS TGT CYS TGT CYS CCT PRO CCT PRO CCT PRO CCT PRO CCT SER AAM	CAT HIS GGC GLY GGC GLY GGC GLY GGC GLY GGC GLY GGC GLY ALA SCCCC PRO CCCC PRO SER	GCA ALA ALA LYS AAAA LYS <b>TCT</b> SER AASN GAT ASN GAT ASN CTA ASP CTA LEU TTA LEU	CAC HIS TGG TRP TGG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG GLN CAG CAC TRP	129
ECO MCF247 XENO ECO MCF247 XENO ECO MCF247 XENO	ARGA ARGA ARGA ARGA ARGA ARGA ARGA ARGA	ARG GAT ASP GEA GEV TGT CVS TGT CVS CCS CCS CCS CCS CCC PRO CCC PRO CCC CPRO CCC CPRO CCC CCC CPRO	GAGU GAGU GAGU GAGU GAU GAU GAU GAU GAU	AGA AGA AGA AGA AGA ACC THR ACC THR ACC THR ACC THR ACC THR ACC THR ACC THR ACC THR ACC THR ACC THR AGA ACC ACC ACC ACC ACC ACC ACC ACC ACC	CIEU GAG SLU 320 TAT TYR TAT TYR TAT TYR TAT TYR TAT TYR ACT THR ACT THR ACT THR ACT ASP GAT ASP	GTT TYR GTT VAL GTT VAL GTT VAL GGA GLY GGA GLY TCCC SER	ASP SASP SEE TGCSS TGCSS TGCSS CAG CAG CAG SER TCAS SER	ACT THR Small S330 PRO CCCC PRO GCA ALA GCA ALA GCA ALA CCCC Z	TAC SER GLY GGT GLY TAC TYR TAC TYR TAC TYR SER GTC VAL	TAT TYR CAT HIS CAT HIS CAT HIS CAT HIS CCA TRP TGG TRP TGG TRP TGG TRP TGG TRP TCCC SER	ACT THR ACT THR ACT THR ACT HIS 420 AAG LYS AAG LYS S10T AAGT SER AGT SER	CCC PRO GTA VAL CCA PRO CCA PRO CCA PRO CCA PRO CCA GAC CCA	CCA PRO CCA CCA CCA CCA CCA CCA CCA CCA CCA CC	TGC CYS ACA THR ATA ATA SER CGG CAG SER SER SER SER CAG SEN	AAT ASN 350 GGG GLY GGG GLY TGG GGLY 30 A SER TCA SER TCA SER CGG GLY GGG GLY	ACG THR GCCC ALA TGG TRP TGG TRP TGG TRP TGG CCC ALA GCCA ALA	GCC ALA CGG ALA GACA GACP GACA ASP GACA ASP GACA ASP ACAA THR ACA THR	TGG TRP TCA SER 440 CTA LEU CTA LEU CTA LEU CTA CCG PRO CCG PRO	AAC ASN TGT CYS TGT CYS ATT ILE ATT ILE ATT ILE STA	AGA ARG GLY GGA GLY TCCC SER TCCC SER TCCC SER TCCC SER TCCC SER TCCC SER TCCC SER	360 <td>AAG LYS CCG PRO CCG PRO CCA PRO CCA LYS AAG CLYS GGG GLY GGGY GLY</td> <td>TTA LEU AGA GGA GGA GGA CGA ARG GLU CGA ARG CGA ARG GLU CGA ARG CGA ASN</td> <td>TCT SER GAG GLU GAG GLU GAG GLU GGA GLY GGA SEN GAS GLU</td> <td>AAA LYS GGCY GGCY TTCC GGCY TTCC GGCY FPHE SO AACA ASN ASN ASN ACCEL S400 ACCAG ACC</td> <td>GTG VAL TACC TYR TACTYR TATTYR ACCTTHR ACTTHR TGC CYS TGCC CYS</td> <td>ACA THR TGT CYS TGT CYS TGT CYS CCT PRO CCT PRO CCT PRO CCT FRO SER AAD AACA ASN</td> <td>CAT HIS GGC GL GGC GGC GGC GGC GGC GGC GGC GGC</td> <td>GCA ALA ALA LYS TCTTSER AAAA LYS SCAC ASP SOCTA LEU CTG LEU TTA LEU</td> <td>CAC HIS TGG TRP TGG GLN CAG GLN CAG GLN CAG GLN GTC VAL GTC VAL ACT THR</td> <td>129</td>	AAG LYS CCG PRO CCG PRO CCA PRO CCA LYS AAG CLYS GGG GLY GGGY GLY	TTA LEU AGA GGA GGA GGA CGA ARG GLU CGA ARG CGA ARG GLU CGA ARG CGA ASN	TCT SER GAG GLU GAG GLU GAG GLU GGA GLY GGA SEN GAS GLU	AAA LYS GGCY GGCY TTCC GGCY TTCC GGCY FPHE SO AACA ASN ASN ASN ACCEL S400 ACCAG ACC	GTG VAL TACC TYR TACTYR TATTYR ACCTTHR ACTTHR TGC CYS TGCC CYS	ACA THR TGT CYS TGT CYS TGT CYS CCT PRO CCT PRO CCT PRO CCT FRO SER AAD AACA ASN	CAT HIS GGC GL GGC GGC GGC GGC GGC GGC GGC GGC	GCA ALA ALA LYS TCTTSER AAAA LYS SCAC ASP SOCTA LEU CTG LEU TTA LEU	CAC HIS TGG TRP TGG GLN CAG GLN CAG GLN CAG GLN GTC VAL GTC VAL ACT THR	129

FIG. 3. Comparison of the nucleotide and deduced amino acid sequences in the 5' env regions of AKR MCF247 (A-MCF247 and MCF247), NFS-Th-1 xenotropic (NFS XENO and XENO) and AKR ecotropic (AKR ECO and ECO) MuLV proviruses. Nucleotide and predicted amino acid sequences which are unique to AKR MCF247 DNA are enclosed within solid lines; the MCF247-specific 12-bp deletion and 3-bp insertion (enclosed within heavy solid lines) are indicated as Y and Z, respectively; the nucleotide and deduced amino acid sequences in the shaded regions in NFS-Th-1 xenotropic and AKR ecotropic proviruses represent substitutions with respect to the sequence of AKR MCF247 DNA; the gaps represent absent nucleotides; the asterisk indicates the initiation codon for translation of the env gene-coded precursor polypeptide; the amino terminus of gp70 is indicated in AKR MCF247 DNA based on amino acid sequence homology to Rauscher MCF MuLV DNA (26); the potential glycosylation sites are underlined; and the restriction sites are indicated with wavy lines. MuLV DNAs in the 3' *pol* region which resulted in only a single substituted amino acid at position 29. In the 5' portion of the *env* gene, M-MCF and the endogenous MuLV DNAs were different in three nucleotides which resulted in two

altered amino acids: one (position 59) was located in the second potential glycosylation site and the second was located at position 171. In the same region, F-MCF DNA contained 18 different nucleotides with respect to the cloned

ENDOG A-12	ccc	тст	стс	10 CAA	GCT	CAC	20 <u>TTA</u>	CAG	GCC	30 CTC	C A A	GCA	GTA	40 CAA	CGA	GAG	50 Gtc	TGG	AAG	60 CCA	CTG	GCC	GCT	70 GCC	ΤΑΤ	CAG	80 Gac	Pvu II cag	
	PRO	SER	LEU	GLN	ALA	HIS	LEU	GLN	ALA	LEU	GLN	ALA	VAL	GLN	ARG	GLU	VAL	TRP	LYS	PRO	LEU	ALA	ALA	ALA	TYR	GLN	ASP	ĞĔŇ	28
A-MCF247	•••	•••	•••	•••	•••	<u></u>	<u></u>	<u></u>	•••	• • •	•••	•••	•••	•••	• • •	•••	•••	•••		•••	•••	•••	•••	•••	•••	•••	•••	ىنىن	
M-MCF	•••	•••	•••	•••	•••	•••		222		•••			•••				•••	•••		т	•••			•••	¢				
F-MCF	•••	•••	•••	•••	c	<u></u>	<u></u>	<u></u>		т		•••	•••			•••	•••			•••		т	G						
ENDOG A-12	<u>Ctg</u> Leu	90 GAC ASP	CAG GLN	CCA Pro	1 GTG VAL	DO ATA ILE	CCA Pro	CAC HIS	110 CCC PRD	TTC PHE	CGT ARG	120 GTC VAL	GGC Gly	GAC ASP	1 ACC THR	30   G T G   V A L	TGG TRP	GTA Val	140 CGC ARG	C G G A R G	CAC HIS	150 Cag Gln	ACT Thr	AAG Lys	1 AAC ASN	60 Ttg Leu	GAA Glu	CCT Pro	56
A-MCF247	ننن	•••								•••		•••						•••			•••								
M-MCF	•••	•••				ç, Val		•••		tři		•••	•••					c				•••				C.A			
F-MCF	. AA Gln	•••				•••						•••				]	•••					•••							
																								Ba	ım H	I			
ENDOG A-12	170 CGC ARG	TGG TRP	AAA Lys	180 GGA GLY	CCC Pro	T A C T Y R	1 ACC THR	90 GTC VAL	C T G L E U	CTG Leu	200 ACC THR	ACC Thr	CCC Pro	210 ACC Thr	GCT Ala	C T C L E U	2 444 L Y S	20 GTA VAL	GAC ASP	GGC Gly	230 ATC ILE	GCT	GCG	240 T <u>GG</u> TRP	<u>ATC</u> ILE	CAC HIS	2 GCC Ala	50 GCT ALA	84
A-MCF247				•••						•••	•••													ند.	ننن				
M-MCF				•••	T	•••					•••			• • • •						•••					ىنىن				
F-MCF	T			•••		т				• • •	•••						•••						]		ننن				
ENDOG A-12	CAC HIS	GTA VAL	260 AAA LYS	GCG Ala	GCG Ala	2 7 0 A C A T H R	ACC Thr	CCT Pro	28 CCG Pro	30 GCC ALA	GGA Gly	ACA THR	290 GCA ALA	TCA SER	GGA GLY	300 CCG PRO	ACA THR	TGG TRP	3 AAG LYS	10 G≓C VAL	C A G G L N	C G T A R G	320 TCT SER	C A A G L N	AAC ASN	330 CCC PRO	TTA LEU	AAG LYS	112
A-MCF247				•••								•••					*			• • •	· · ·								
M-MCF			•••														*												
F-MCF			G														*												
ENDOG A-12	34 ATA ILE	O AGA ARG	TTA LEU	ACC THR	350 CGT ARG	GGG GLY	GCC Ala	360 CCC PRO	T A A E N D	1;	20							-											
A-MCF247			•••																										
M-MCF																													
F-MCF									.G.																				

FIG. 4. Comparison of sequences in the 3' pol regions of cloned endogenous and MCF MuLV DNAs. Nucleotide and deduced amino acid sequences of an endogenous MuLV DNA segment in clone A-12 (17) (ENDOG A-12) are compared with sequences in an analogous region in AKR MCF247 (A-MCF247), M-MCF (2), and F-MCF (Adachi et al., submitted for publication) MuLV DNAs. Nucleotides which are identical in the endogenous A-12 and the MCF DNAs are represented by dots; the nucleotide differences are indicated; the 12-bp MCF247-specific insert is enclosed within heavy solid lines and designated as X; the sequences conserved in the cloned endogenous and MCF MuLV DNAs which are different from those present in ecotropic or xenotropic proviruses are enclosed within solid lines; the light-shaded region represents the nucleotides and amino acid unique to F-MCF; the dark-shaded region indicates M-MuLV ecotropic sequences (29), which are present in the M-MCF recombinant MuLV DNA; the splice acceptor site is indicated with a broken line; restriction sites are indicated with a wavy line; the asterisk marks the position of the initiation codon for translation of *env* gene precursor polypeptide.

ENDOG A-12	# 10 ATG GAA GGT C MET GLU GLY P	CA GCG TTC T RO ALA PHE S	0 3 CA AAA CCC CT ER LYS PRD LE	0 T AAA GAT U LYS ASP	40 AAG ATT LYS ILE	50 AAC CCG TGG ASN PRO TRP	GGC CCC GLY PRU	60 CTA ATA ATC LEU ILE ILE	70 CTG GGG ATC TTA ATA LEU GLY ILE LEU ILE	AGG ARG 28
A-MCF247	<b>*</b>							G Val		
M-MCF	•						••••	···· ··· <b>···</b>		
F-MCF	•							G VAL	···· ··· ··· ··· <u> </u>	J
		► gp70								
ENDOG A-12	90 GCA GGA GTA 1 ALA GLY VAL	TCA GTA CAA C	AT GAC AGC CC	120 T CAT CAG O HIS GLN	GTC TTC VAL PHE	AAT GTT ACT ASN VAL THE	140 TGG AGA TRP ARG	GTT ACC AAC VAL THR ASN	160 TTA ATG ACA GGA CAJ LEU MET THR GLY GLY	A ACA N THR 56
A-MCF247		e. 				<u></u>	<u> </u>			
M-MCF						<u></u>	<u>.</u>			
F-MCF			<u></u>	CA		<u></u>	<u>.</u>			
	170	180	190	200	210		220	230		250
ENDOG A-12	ALA ASN ALA	THR SER LEU I	EU GLY THR ME	T THR ASP	ALA PHE	PRO LYS LEU	J TYR PHE	ASP LEU CYS	ASP LEU ILE GLY ASI	PASP 84
A-MCF247	···· <u>····</u>	<u></u>			···· ···		· ··· ···			
	¥AL									Ţ
F-MCF	···· <u>···</u> ····	<u></u> 7			···· ···	NET				• • • • • •
ENDOG A-12		ACT GGA CTC	280 GGG TGT CGC AG	Smal	290 GGA AGA			0 TTT GAC TIC PHE ASP PHE	Smal 320 330 TAT GTT TGC CCC GG( TYR VAL CYS PRO GL	G CAT Y HIS 112
A-MCF247				ىنى نىن ·	,				ىن نىن	. · · ·
M-MCF				ىند مىن .	,		s			: ···
F-MCF	··· ··· []	ст	••• ••• •••	 نند سند ·	· · · · · · ·				ىند ئىنە نىن	
ENDOG A-12	340 Act gta cca Thr val pro	350 ACA GGG TGT I Thr Gly Cys I	360 GGA GGG CCG AI GLY GLY PRO AI	370 54 GAG GGC 86 GLU GLY	TAC TGT	380 GGC AAA TGG GLY LYS TRI	390 G GGC TGT P GLY CYS	400 GAG ACC ACT GLU THR THR	410 GGA CAG GCA TAC TG GLY GLN ALA TYR TRI	420 G AAG P LYS 140
A-MCF247										
M-MCF										
F-MCF			···· ··· ··· [.	<u></u>			• • • • • • • •			
	43	0 4	40 4	50	460	470		430	490 <b>5</b> 00	<u>,</u>
ENDOG A-12	CCA TCA TCA Pro ser ser	TCA TGG GAC SER TRP ASP	CTA ATT TCC C LEU ILE SER L	TT AAG CGA EU LYS ARG	GGA AAC Gly ASN	ACC CCT CG THR PRO AR	G AAT CAG G ASN GLN	GGC CCC TGT GLY PRO CYS	TAT GAT TCC TCA GC TYR ASP SER SER AL	G GTC A VAL 168
A-MCF247				••••••		···· ···	•			
M-MCF										
F-MCF				•••••		···· ··· [				.l
ENDOG A-12	510 TCC AGT GAC SER SER ASP	520 ATC AAG GGC	530 GCC ACA CCG G ALA THR PRO G	540 GG GGT CGA Ly Gly Arg	) TGC AAT G CYS ASN	550 CCC CTA GT PRO LEU VA	ECO 560 C CTG GAA L LEU GLU	RI TTC PHE 196		
A-MCF247								ىنىن		
M-MCF	A						ىنىن	ننن		
F-MCF	G. GLY	C.AT		•••••		•••• ••• ••	• نند	ىنىن		

FIG. 5. Comparison of sequences in the 5' env regions of cloned endogenous and MCF MuLV DNAs. Nucleotide and deduced amino acid sequences of AKR MCF247 (A-MCF247), M-MCF, and F-MCF MuLVs are compared with the sequences present in the cloned endogenous A-12 DNA (ENDOG A-12). The designations used in this figure are the same as described in the legend to Fig. 4. The MCF247-specific 12-bp deletion  $(\nabla)$  and 3-bp insertion (enclosed within heavy solid lines) are designated as Y and Z, respectively. The shaded areas indicate nucleotides and amino acids which are unique to each MCF isolate and different from sequences in corresponding positions in ecotropic and xenotropic proviruses; the amino terminus of the gp70 in the cloned endogenous and MCF DNAs is indicated based on amino acid homology to Rauscher MCF gp70 (26); potential glycosylation sites are underlined.

endogenous MuLV DNA sequences, which resulted in five substituted amino acids in the *env* region: one in the leader segment (position 22) and four in the gp70-coding region (positions 72, 167, 171, and 173). The substitution of alanine by a valine at position 167 in F-MCF DNA resulted from a single base change at nucleotide 500 in the 3-bp MCF247-specific insertion (designated as Z in Fig. 5).

It is interesting to note that no single alteration was shared between the three MCF DNAs with respect to the endogenous MuLV DNA in clone A-12. The high degree of nucleotide sequence conservation between the retroviral DNA in clone A-12 and the three MCF proviruses establishes the role of endogenous MuLV DNA segments in formation of recombinant MuLVs.

#### DISCUSSION

Oligonucleotide fingerprinting and restriction enzyme mapping analyses of MCF MuLV isolates indicate that they differ from one another in the amount of nonecotropic sequences incorporated into their *pol* and *env* regions (4, 23). These nucleotide substitutions invariably involve the 5' portion of the env gene. It has previously been demonstrated that the novel sequences which are ultimately incorporated into MCF viruses preexist in mouse chromosomal DNA (3, 17). Some of these segments, which contain MCF-related env regions on the basis of restriction mapping and Southern blot analyses, have recently been cloned from AKR/J and BALB/c mouse DNAs (17). One of the cloned endogenous MuLV DNA segments, which was isolated from AKR/J mouse DNA and designated A-12, was identified as a potential env gene progenitor of AKR MCF247 MuLV. To confirm this relationship between the endogenous and infectious MuLV DNAs, the nucleotide sequence of 363 bases near the 3' end of the pol gene and 564 bases in the 5' region of the env gene in the cloned endogenous and AKR MCF247 MuLV DNA was determined.

Recently, Holland et al. have reported the nucleotide sequence of the env gene in cloned AKR MCF247 MuLV DNA (14); that sequence differs by one nucleotide from the one reported here (position 249; Fig. 3). This single base change does not result in an amino acid alteration. The nucleotide sequence comparison of the cloned endogenous MuLV DNA segment and MCF247 MuLV DNA reported here indicates that all of the hallmarks of MCF proviruses, identified after the sequencing of MCF247, xenotropic, and ecotropic MuLV proviral DNAs (Fig. 2 and 3), were present in the cloned AKR mouse cellular DNA segment. These unique structural features, which were lacking in both xenotropic and AKR ecotropic MuLV proviruses, included (i) a 12-bp insertion (designated as X in Fig. 2) located 60 bp from the 3' end of the pol gene (and immediately preceding the 5' terminus of the env gene) and (ii) a 12-bp deletion and a 3-bp insertion in the 5' env region (indicated with a Y and Z, respectively, in Fig. 3). The comparison of the cloned endogenous and MCF247 MuLV DNAs revealed total nucleotide sequence identity in the 3' pol region (Fig. 4) and only 4 of 564 mismatched bases in the 5' env region, resulting in the substitution of three amino acids (Fig. 5). In addition, the unique structural features present in the cloned endogenous and MCF247 MuLV DNAs could also be identified in analogous regions of M- and F-MCF MuLV proviruses. The high degree of nucleotide sequence conservation between the cloned endogenous MuLV DNA and MCF proviruses corroborates the role of mouse cellular DNA sequences in the generation of env gene recombinant MuLVs.

Southern blot analyses of restricted mouse DNAs have

indicated the presence of 18 to 28 xenotropic MuLV-related DNA segments (13). Of the six env-containing cloned endogenous MuLV proviral segments isolated from BALB/c and AKR/J mouse DNAs, only one contained purely xenotropic env specificity (17). The remaining five cloned endogenous MuLV DNAs had env regions possessing restriction maps and hybridization properties typical of MCF proviruses. Such results suggest that the majority of MuLV-related DNAs in the mouse genome contain an MCF-type env region. It is interesting to note that each of the MCF-related env segments associated with cloned endogenous MuLV DNAs could be distinguished from one another after AluI or HpaII digestion (17). It is unclear at this time whether these endogenous MCF-type env DNA segments constitute a family of highly related MuLV DNAs or whether each isolate represents a different class of endogenous MuLV DNA. The relationship of the endogenous MCF-type env DNA segments to each other and to known MCF MuLV proviral env regions remains to be elucidated by nucleotide sequence analysis. In view of the existence of a large number of endogenous MuLV DNAs containing distinctive env regions, it was quite unexpected to learn that the env regions of three different MCF MuLV (AKR MCF247, M-MCF, and F-MCF) were nearly identical to the env region present in a single endogenous MuLV DNA (namely, in clone A-12). Several possibilities can be entertained to explain the incorporation of *pol* and *env* sequences similar to those present in clone A-12 into three different MCF proviruses. (i) The proviral DNA in clone A-12 is located in a region of the chromosome in which a high frequency of recombination occurs; such a "hot spot" for recombination could even be located within the retroviral sequences in clone A-12. (ii) The endogenous MuLV provirus associated with clone A-12 may represent a member of a highly related multimembered MuLV family which contains env genes capable of donating a functional gp70 to an ecotropic MuLV to generate MCF viruses.

It has been demonstrated in viral interference studies that the host range of MuLVs is defined by the recognition of distinct cell surface receptors by the viral env gene-coded protein products, such as gp70, of the three classes of MuLVs, namely, ecotropic, xenotropic, and dualtropic MCFs (1, 21). The results presented here indicate that the amino acid sequence of the 5' portion of the env gene in MCF247 MuLV is significantly different from that in the analogous region of AKR ecotropic MuLV. Furthermore, MCF247 MuLV can also be distinguished from xenotropic MuLV in its 5' env region based on 23 amino acid substitutions, deletion of 4 amino acids, and insertion of 1 amino acid (Fig. 3). Thus, the amino-terminal portions of the gp70s associated with the three classes of MuLVs may very well be involved in determining host range by recognizing different receptors on the cell surface. Recombinant MCF MuLVs have been shown to infect mouse cells via receptors distinct from those used by ecotropic MuLVs (21). The MCFencoded gp70 might, thereby, allow the entry of MCF viruses into tissues that are inaccessible to prototype ecotropic MuLVs. After entry into such cells, other regions of the viral genome may then function to initiate the leukemogenic process. For example, in the case of MuLV-induced thymic lymphomas (25), the LTR associated with MCF proviruses may facilitate enhanced expression of viral genes in thymus cells. The resulting increase in the number of progeny virions could then lead to multiple integration events, one of which might activate a "leuk" gene in a manner analogous to the promoter-insertion model (11). Alternatively, enhanced

#### Vol. 50, 1984

expression of a viral gene product (e.g., gp70) mediated by an LTR now associated with a recombinant MuLV *env* region and located in a cell type previously inaccessible to it could be pathogenic.

It is worth noting that the rate of viral integration can be affected by mutations in the 3' region of the *pol* gene (S. Goff, personal communication). Thus, the insertion of 12 bp in the 3' end of the *pol* gene in MCF MuLV DNAs might affect its integration into cellular DNA and activation of a potential *leuk* gene. A role for this 12-bp *pol* insert in leukemogenesis is supported by the fact that both M-MuLV (29) and F-MuLV (18), which are leukemogenic ecotropic viruses, also have sequences (9 bp) inserted into this portion of their *pol* genes, whereas the nonleukemogenic ecotropic MuLVs, such as AKv, lack sequences in this region.

### ACKNOWLEDGMENTS

I am grateful to Paul E. Steele for introducing me to the Maxam and Gilbert sequencing procedure. I also thank Willard W. Davis for excellent technical assistance, Malcolm A. Martin and Paul E. Steele for critical review of the paper, and Laura A. Wilders for typing the manuscript.

#### 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.
- 2. 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.
- 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, S. Gupta, E. Rands, and D. R. Lowy. 1981. Origin of mink cytopathic focus-forming (MCF) viruses; comparison with ecotropic and xenotropic murine leukemia virus genomes. Virology 113:465–483.
- Chien, Y.-H., I. M. Verma, T. Y. Shih, E. M. Scolnick, and N. Davidson. 1978. Heteroduplex analysis of the sequence relations between the RNAs of mink cell focus-inducing and murine leukemia viruses. J. Virol. 28:352–360.
- 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.
- Donoghue, D. J., E. Rothenberg, N. Hopkins, D. Baltimore, and P. A. Sharp. 1978. Heteroduplex analysis of the nonhomology region between Moloney MuLV and the dual host range derivative HIX virus. Cell 14:959–970.
- 8. 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.
- 9. Fischinger, P. J., S. Normura, and D. P. Bolognesi. 1975. A novel murine oncornavirus with dual eco- and xenotropic properties. Proc. Natl. Acad. Sci. U.S.A. 72:5150-5155.
- Hartley, J. W., N. K. Wolford, L. J. Old, and W. P. Rowe. 1977. A new class of murine leukemia virus associated with development of spontaneous lymphomas. Proc. Natl. Acad. Sci. U.S.A.

74:789-792.

- 11. 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.
- Herr, W., V. Corbin, and W. Gilbert. 1982. The nucleotide sequence of the 3' half of Akv. Nucleic Acids Res. 10:6931– 6944.
- Hoggan, M. D., C. E. Buckler, J. F. Sears, W. P. Rowe, and M. A. Martin. 1983. Organization and stability of endogenous xenotropic murine leukemia virus proviral DNA in mouse genomes. J. Virol. 45:473-477.
- 14. Holland, C. A., J. Wozney, and N. Hopkins. 1983. Nucleotide sequence of the gp70 gene of murine retrovirus MCF 247. J. Virol. 47:413-420.
- Khan, A. S., and M. A. Martin. 1983. Endogenous murine leukemia proviral long terminal repeats contain a unique 190base-pair insert. Proc. Natl. Acad. Sci. U.S.A. 80:2699-2703.
- Khan, A. S., R. Repaske, C. F. Garon, H. W. Chan, W. P. Rowe, and M. A. Martin. 1982. Characterization of proviruses cloned from mink cell focus-forming virus-infected cellular DNA. J. Virol. 41:435-448.
- 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.
- Koch, W., G. Hunsmann, and R. Friedrich. 1983. Nucleotide sequence of the envelope gene of Friend murine leukemia virus. J. Virol. 45:1-9.
- Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. 74:560-564.
- Montreuil, J. 1980. Primary structure of glycoprotein glycans. Basis for the molecular biology of glycoproteins. Adv. Carbohydr. Chem. Biochem. 37:157-223.
- Rein, A. 1982. Interference grouping of murine leukemia viruses: a distinct receptor for the MCF-recombinant viruses on mouse cells. Virology 120:251-257.
- Repaske, R., R. R. O'Neill, A. S. Khan, and M. A. Martin. 1983. Nucleotide sequence of the *env*-specific segment of NFS-Th-1 xenotropic murine leukemia virus. J. Virol. 46:204–211.
- 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.
- Rosner, M. R., L. S. Grinna, and P. W. Robbins. 1980. Differences in glycosylation patterns of closely related murine leukemia viruses. Proc. Natl. Acad. Sci. U.S.A. 77:67-71.
- Rowe, W. P., M. W. Cloyd, and J. W. Hartley. 1979. Status of the association of mink cell focus-forming viruses with leukemogenesis. Cold Spring Harbor Symp. Quant. Biol. 44:1265–1268.
- Schultz, A., A. Rein, L. Henderson, and S. Oroszlan. 1983. Biological, chemical, and immunological studies of Rauscher ecotropic and mink cell focus-forming viruses from JLS-V9 cells. J. Virol. 45:995-1003.
- 27. Sharp, P. 1981. Speculations on RNA splicing. Cell 23:643-646.
- Shih, T. Y., M. O. Weeks, D. H. Troxler, J. M. Coffin, and E. M. Scolnick. 1978. Mapping host range-specific oligonucleotides within genomes of the ecotropic and mink cell focus-inducing strains of Moloney murine leukemia virus. J. Virol. 26:71-83.
- Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) 293:543-548.