The EMBO Journal Vol.2 No.12 pp.2239-2246, 1983

Sequence analysis of Gardner-Arnstein feline leukaemia virus envelope gene reveals common structural properties of mammalian retroviral envelope genes

Martin Wünsch, Ansgar S. Schulz, Werner Koch, Roland Friedrich* and Gerhard Hunsmann[1]

Institut für Immunbiologie, Universität Freiburg, D-7800 Freiburg, and [1]Deutsches Primatenzentrum, D-3400 Göttingen, FRG

Communicated by G. Hobom
Received on 8 August 1983

We have sequenced the envelope (*env*) gene and most of the adjacent 3' long terminal repeat (LTR) of Gardner-Arnstein feline leukaemia virus subtype B. The LTR of this virus contains, at corresponding positions, all signal sequence elements known from other retroviral LTRs. The deduced amino acid sequence of the longest open reading frame was compared with *env* polypeptide sequences of several murine leukaemia viruses. This allowed us to predict the positions of both p12/15^{env} and gp70 polypeptides as well as a hydrophobic leader polypeptide. The *env* polypeptides of the different viruses show long stretches of homology and similar hydrophilicity profiles in the p12^{env} region and in the carboxy-terminal half of gp70 (constant region). The most extensive variations are confined to certain parts of the amino-terminal half of gp70 (differential region). In this region, however, feline leukaemia virus and murine mink cell focus forming viruses are still closely related. A correspondingly spaced pattern of identical, short amino acid sequences appears in three different parts of the *env* polyprotein, suggesting its evolution from a primordial *env*-related precursor by tandem duplications.

Key words: antigenic determinants/evolution of retroviral *env* genes/feline leukaemia virus/Friend MCF virus/long terminal repeat

Introduction

Envelope (*env*) genes of retroviruses code for multifunctional polypeptides. These *env* polypeptides play an important role in several steps of the viral cycle, especially during the extracellular phase. They take part in maturation, when the particle is budding from the plasma membrane of the host cell. They protect the virion against the host's immune system and they select for infectible target cells by interaction with specific receptor molecules (Robinson *et al.*, 1980). By cooperation with such receptors, the viral core can penetrate into host cells. It is our working hypothesis that these specific functions may be mediated by defined domains of *env* gene polypeptides in analogy to other multifunctional polypeptides, e.g., immunoglobulins (Klein, 1982). To identify such domains, it appeared reasonable to compare sequences of *env* polypeptides of retroviruses with distinct biological characteristics possibly attributable to their *env* polypeptides.

Since the *env* genes of several murine leukaemia viruses have been sequenced recently, the *env* gene of a feline leukaemia virus (FeLV) seemed to be of special interest for a comparison. FeLVs are a group of horizontally transmitted type C retroviruses causing neoplastic and degenerative diseases of the haematopoietic system, such as lymphosarcomas or severe immunosuppression, in domestic and free ranging cats (Jarret, 1982). FeLVs are classified into three distinct subgroups A, B and C on the basis of interference (Sarma and Log, 1973). Subtype A shows the widest distribution within cat populations (Hardy *et al.*, 1976). Subtype B is always found together with A and subtype C is isolated only rarely (Jarret, 1982). Whereas replication of subtype A appears to be restricted to cells of feline origin, subtype B shows a broader host range, reminiscent of that of murine mink cell focus forming viruses (MCFV) (Sarma *et al.*, 1975). A more complete knowledge of the functional anatomy of retroviral envelope polypeptides may help to understand these specificities. This may enable effective interaction with viral propagation by immunization to specific determinants. In addition, the comparative approach may elucidate the evolution of these molecules.

Here we report the DNA sequence of the Gardner-Arnstein FeLV subtype B (GA-FeLV B) envelope gene, which allows us to deduce the amino acid sequence of the *env* polypeptides. When we compare it with those of the ecotropic AKR mouse virus (Akv; Lenz *et al.*, 1982), the Friend murine leukaemia virus (F-MuLV; Koch *et al.*, 1983) and the Friend mink cell focus forming virus (F-MCFV; Koch,W., Zimmermann,W., Oliff,A. and Friedrich,R., in preparation), we recognize a region of high variability (differential region) and another one of high conservation (constant region). A detailed analysis of these sequences suggests their evolution from a primordial *env*-related gene by two tandem duplications. An exchange of modular genetic blocks might have conferred new characteristics to the various gene products and the viruses carrying them.

Results and Discussion

Sequencing strategy

The smaller *Xho*I/*Eco*RI DNA fragment (Figure 1A) of the molecular clone pKHR4, containing the envelope gene and 3' long terminal repeat (LTR) of GA-FeLV B (J. Mullins, personal communication) was sequenced according to the strategy shown in Figure 1B-D. The nucleotide sequence and the deduced amino acid sequence of the proposed envelope precursor polypeptide is shown in Figure 2.

Structure of the 3' LTR

Retroviral LTRs are known to contain sequences that control reverse transcription, integration of the viral DNA into the host chromosome, RNA synthesis and other functions of the viral cycle (Temin, 1982; Coffin, 1982). By comparison with known sequences, we have localized the sequences U_3-R-U_5, where U_3 and U_5 designate unique sequences originating from the 3' and 5' ends of viral RNA, respectively, and R designates a sequence redundant at both ends of viral RNA. Furthermore, we have localized many of the putative signal sequences necessary for the above-mentioned functions in the 3' LTR of GA-FeLV B.

The GA-FeLV B LTR turned out to be similar to murine leukaemia virus LTRs (Shinnick *et al.*, 1981; Lenz *et al.*, 1982; Koch,W., Zimmermann,W., Oliff,A. and Friedrich,R., in preparation) and nearly identical to the LTR

*To whom reprint requests should be sent.

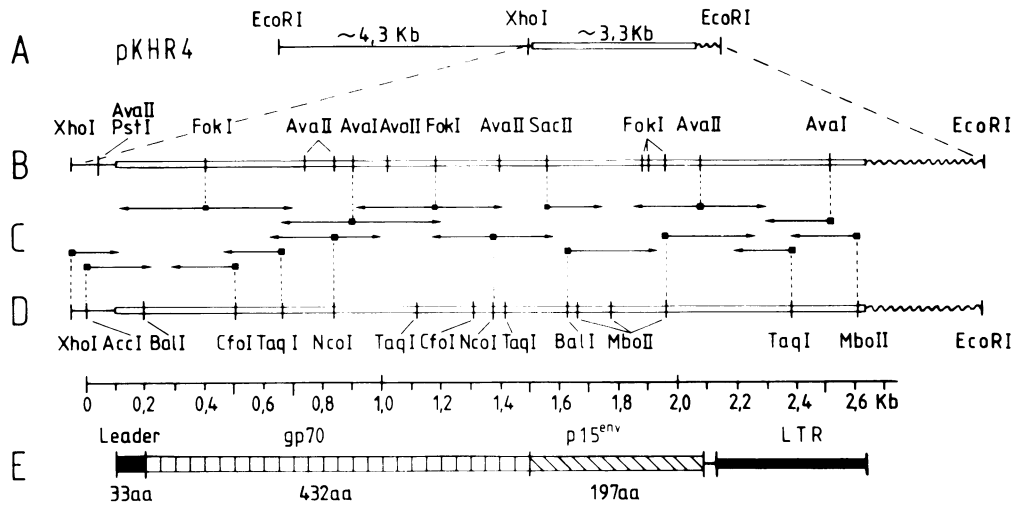

Fig. 1. Physical map and sequencing strategy for GA-FeLV B *env* gene and 3′ LTR region. Double lines represent proviral DNA, thin lines vector DNA and wavy lines cellular sequences picked up during DNA cloning. **(A)** The Ga-FeLV B *env* gene is located between the *Xho*I and *Eco*RI site of pKHR4. **(B,D)** The endonuclease cleavage sites used for DNA sequencing within the *Xho*I/*Eco*RI fragment. **(C)** Sequencing strategy. Solid boxes represent 5′ ends labelled with [γ-^{32}P]ATP. **(E)** Representation of the *env* polypeptide and LTR region as deduced from the DNA sequence.

of GA-FeSV, recently described by Hampe *et al.* (1983). The only difference to GA-FeSV is a G at position 2260 in GA-FeLV B instead of an A in GA-FeSV.

U_3 is concluded to start at position 2148 proximal to the inverted repeat (IR) AATGAAAGACCCC (Figure 2) 48 nucleotides 3′ to the stop codon of p15^{env}, immediately 3′ to a polypurine tract. Since our sequence ends shortly before the right border of U_3, the right IR is not given. The R region has a length of 72 nucleotides extending from position 2486 to 2557.

The following signal sequences could be localized within the LTR: the CAT box was found at position 2419–2423. A Goldberg-Hogness (TATA) box was identified at residue 2460–2466. A possible Z-DNA segment (Nordheim and Rich, 1983) of the form TGTGCGCGCGC has been observed at position 2444–2452. The polyadenylation signal AATAAA was found at position 2536–2541. Unlike many other retroviruses (Temin, 1982; Coffin, 1982) the U_3 segment of GA-FeLV B contains no large direct repeat.

Open reading frames and localization of env polypeptides

The single major open reading frame is 2097 nucleotides long and able to code for a polyprotein of 699 amino acids (Figure 2). In contrast to other published *env* sequences of retroviruses, a stop codon does not precede its 5′ end. Similarly to F-MuLV (Koch *et al.*, 1983), the complementary strand also contains several longer open reading frames up to a length of 126 amino acids. Their significance remains to be investigated.

In analogy to other retroviruses, the gene product of the GA-FeLV B *env* gene is probably translated from a spliced mRNA (Hayward, 1977; Mellon and Duesberg, 1977) and consists of a glycosylated protein precursor of ~90 000 daltons (Pinter and Fleissner, 1979) including a leader sequence of 33 amino acids. The precursor is then cleaved (Velicer and Graves, 1974; Pinter and Fleissner, 1977) into a glycosylated gp70 and an unglycosylated p15^{env} which further matures to p12^{env} (Figure 2). Gp70 and p12^{env} are linked together by disulfide bridges (Leamnson *et al.*, 1977; Schneider and Hunsmann, 1978). Three points of polyprotein cleavage were defined by aligning published sequences of murine leukaemia viruses to the GA-FeLV B *env* sequence. By analogy to other signal peptides (Blobel and Dobberstein, 1975; Shinnick *et al.*, 1981; Koch *et al.*, 1983) we assume that the leader begins at the methionine residue in position −33 relative to the maturation cleavage point (Figure 2). The N-terminal amino acid sequences of gp70 and p15^{env} for several mouse viruses have been determined (Oroszlan *et al.*, 1980). From these sequences we deduced the start of gp70 to be the alanine residue at position 1 and the start of p15^{env} to be the glutamine residue at position 433.

Table I. Calculated mol. wts. of proposed GA-FeLV B *env* polypeptides

Polypeptide	No. of amino acid residues	Mol. wt. of the apopolypeptide
Total *env* precursor polypeptide	662	73 120
N-terminal signal peptide	33	3714
gp70	432[a]	47 406[a]
p15^{env}	197	22 036
p12^{env}	180	19 922

[a]Since some amino acids might be cleaved off the carboxy terminus of gp70, the actual number might be slightly smaller.

The C terminus of gp70 has not yet been defined for any retrovirus. It is possible that, in analogy to maturation of influenza A virus haemagglutinin (Garten *et al.*, 1981), the proteolytic cleavage between gp70 and p15^{env} may lead to removal of one or several basic amino acids at the C-terminus of gp70. Therefore, the carboxy terminus of gp70 might be located a few amino acids to the left. Mol. wts. of the different *env* proteins are given in Table I.

Comparison of the deduced amino acid sequences of GA-FeLV B, F-MCFV, F-MuLV and Akv

To discover which regions of the *env* polyprotein are characteristic for GA-FeLV B, we compared the deduced *env* polyprotein sequence with those of known sequences of several murine viruses. We have chosen for comparison the

Fig. 2. Nucleotide sequence and deduced amino acid sequence of the GA-FeLV B *env* gene and most of 3′ LTR. The nucleotide sequence has been determined between an *Acc*I and an *Mbo*II site marked in Figure 1B. Boundaries of codogenic regions are marked by arrows. The putative signal sequences in the LTR are underlined (see text). Nucleotide sequences are indicated on the left and amino acid positions relative to the start of gp70 on the right margin.

env polyprotein sequences of a polytropic murine leukaemia virus, F-MCFV (Koch,W., Zimmermann,W., Oliff,A. and Friedrich,R., in preparation), and two ecotropic murine leukaemia viruses, F-MuLV (Koch *et al.*, 1983) and Akv (Lenz *et al.*, 1982), both being of distinct subtype and oncogenic spectrum (Figure 3). Two other published sequences of *env* genes of murine leukaemia viruses, Moloney-MuLV (Shinnick *et al.*, 1981) and Moloney-MCFV (Bosselman *et al.*, 1982), are very similar to the ones shown and therefore not included.

Regions of homology. All four viruses are homologous in p15^{env} by >80%, and in the carboxy-terminal half (constant region, CR) of gp70 by >67%. These regions are conserved most extensively during evolution. Both constant regions are even more conserved within the group of mouse viruses (96% and 75%, respectively) than between murine and feline viruses. In both comparisons, the extreme carboxy terminus of p15^{env}, where seven positions lack any conservation, has not been considered, since 17 C-terminal residues are removed from most of p15^{env} molecules during maturation, yielding the final p12^{env} product (Naso *et al.*, 1976; Oroszlan *et al.*, 1980).

Regions of high variability. The amino-terminal half of gp70 shows considerably larger variations (differential region,

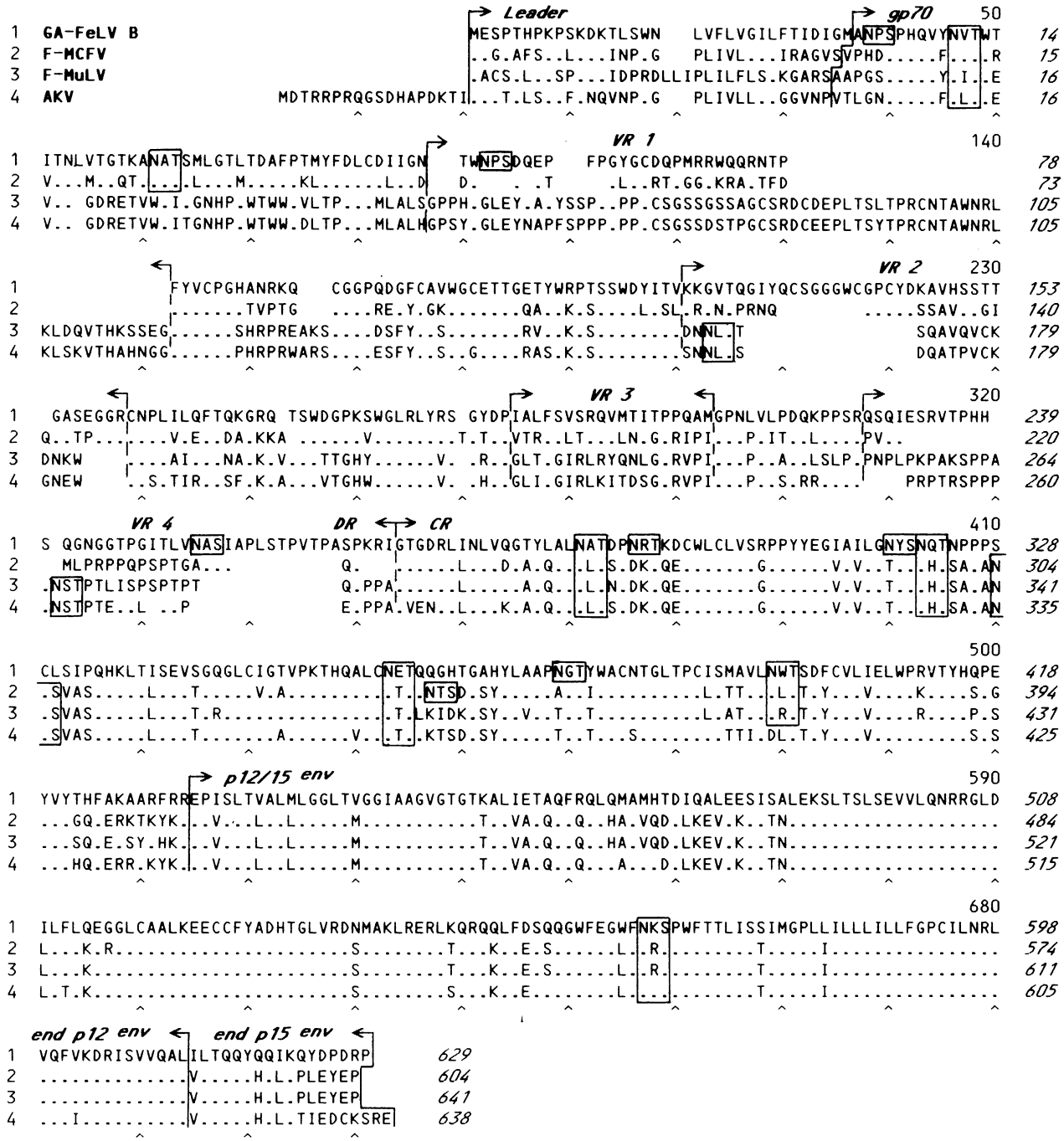

Fig. 3. Amino acid sequence comparison of proposed *env* polypeptides of GA-FeLV B (**1**), F-MCFV (**2**), F-MuLV (**3**) and Akv (**4**). Gaps have been introduced to allow optimal alignment. Positions of identical amino acids are denoted by dots. Potential glycosylation sites are shown in boxes. Roman numbers above the lines indicate relative amino acid positions of the polypeptides compared. Numbers in the text refer to these positions. Numbers in italics indicate amino acid positions of individual *env* sequences, starting wtih position 1 at the first amino acid of gp70. CR = constant region, DR = differential region, VR = variable region. The start of the leader peptide of Akv has not yet been determined (Lenz *et al.*, 1982). In analogy to the other viruses we assume it to be at the methionine in position 1.

DR; Figure 3). Four main variable regions (VR 1 – 4) can be pointed out, which are separated by short, conserved sequences. The ecotropic viruses F-MuLV and Akv, however, are more highly related in all four VRs (Table II). The degree of variation generally increases from VR 1 to VR 4. In VR 1 (position 87 – 152) the similarity of GA-FeLV B and F-MCFV is striking. It is also significant in VR 2 and VR 3. In VR 2 (position 201 – 238) differences between GA-FeLV B and the ecotropic mouse viruses are most obvious, suggesting some group-specific properties for this region. In VR 3 (position 275 – 293) the variations of F-MCFV relative to either GA-FeLV B or F-MuLV is intermediate. VR 4 of GA-FeLV B (position 308 – 353) shows very few homologies to any of the other viruses (hypervariable region; Koch *et al.*, 1983; Koch,W., Zimmermann,W., Oliff,A. and Friedrich,R., in preparation). It is part of the proline-rich region (Koch *et al.*, 1983). Only half as much proline residues are found in this sequence of GA-FeLV B (16%) compared with F-MuLV and Akv (32% each). This suggests a less rigid secondary structure of the GA-FeLV B gp70 in this region.

FeLV B is known to have a broader host range than FeLV A (Sarma *et al.*, 1975; Jarret *et al.*, 1982) similar to the MCFVs. This similarity is paralleled by a striking sequence homology, especially in VR 1, between GA-FeLV B and F-MCFV. These results may suggest recombinational events that occurred in the cat between a progenitor FeLV and acquired murine MCFV sequences. The extended host range of MCFVs might have favoured such an acquisition of part of their sequences by viruses of other species. Interestingly, an exchange of the N-terminal half of gp70 appears to be involved in recombinations between ecotropic- and xenotropic-like viruses in the formation of MCFVs (Bosselmann *et al.*, 1982; Koch,W., Zimmermann,W., Oliff,A. and Friedrich,R., in preparation). Since MCFVs are assumed to be essential for the induction of leukaemia in mice (Chattopadhyay *et al.*, 1982; Ruscetti *et al.*, 1981), MCFV-like sequences in GA-FeLV B could play a similar role in the cat.

Glycosylation sites

Asn-X-Thr/Ser sequences are known to be possible sites for glycosylation of retroviral polypeptides (Montreuil, 1980; Geyer *et al.*, 1982). Oligosaccharide side chains may help to expose certain antigenic epitopes or receptor recognition sites to the surface of the molecule. The *env* polyprotein of GA-FeLV B carries 13 potential glycosylation sites (Figures 3,4), four at the amino terminus, eight in the carboxy-terminal half of gp70 and one in p15^{env}. In the comparative analysis, five of these sites occur in homologous positions of all four viruses with the asparagine and the threonine/serine residue conserved, but varying in position X. This argues for a strong selective pressure in favour of their conservation suggesting that they are actually of functional importance. Although conserved, the sequence of p15^{env} is not glycosylated (Moennig *et al.*, 1974; Pinter and Fleissner, 1977; Witte and Wirth, 1979).

One constant and three glycosylation sites, not found in murine viruses, cluster (Figure 3) around the N terminus of gp70 of GA-FeLV B (positions 38, 46, 61, 92). In VR 4 (position 335) this glycosylation site is shifted towards the C-terminal end compared with the sites of the ecotropic mouse viruses (position 322), while F-MCFV lacks this site. Three additional sites appear in the constant region of gp70 of GA-FeLV B (positions 376, 400, 459). Two glycosylation sites of the ecotropic murine viruses are not shared by GA-FeLV B (positions 203, 410). Consistent with its high homology to GA-FeLV B, F-MCFV shares one of the additional sites at the N terminus with GA-FeLV B (position 61).

A further site in the constant region may illustrate the importance of glycosylation sequences for immunogenicity since it coincides with the genomic position of G_{IX}^+ and G_{IX}^- oligonucleotides obtained by RNase T1 fingerprinting (Donis-Keller *et al.*, 1980). Viruses harbouring the G_{IX}^+ oligonucleotide lack this site but express antigen G_{IX} at their gp70, whereas G_{IX}^- viruses show a glycosylation site, but fail to express antigen (Stockert *et al.*, 1971; Tung *et al.*, 1975; Old and Stockert, 1977). Glycosylation of this position seems to restrict antibody binding to an otherwise antigenic determinant.

Hydrophobic and hydrophilic regions

Following the algorithm of Hopp and Woods (1981), we established hydrophilicity profiles for the *env* polypeptides of the four viruses compared (Figure 4). Hydrophobic regions help to integrate proteins into lipid membranes. Regions of at least 11 hydrophobic amino acids can span the membrane (Engelmann and Steitz, 1981). The largest hydrophobic regions are located at corresponding positions in the leader peptide, the constant region and p15^{env} of all four viruses. The apolar character of these regions is maintained in spite of major differences in the level of amino acids. The clustering of glycosylation sites in the constant region might prevent its hydrophobic stretches from re-integration into the membrane after the excretion of gp70. In GA-FeLV B the hydrophobic character of the constant region extends into VR 4. This was not observed in the other viruses. p15^{env} is probably anchored in the membrane by its two apolar N- and C-terminal ends forming α-helices (Lenz *et al.*, 1982), while its polar central part loops to the outside. This part carries the three cysteine residues probably involved in the binding of gp70 by disulfide bridges (Pinter and Fleissner, 1977; Schneider and Hunsmann, 1978).

Regions of strong hydrophilicity should be located at the outer side of a polypeptide and thus may be recognized either by the immune system as antigenic determinants (Hopp and Woods, 1981) or by endopeptidases for cleavage reactions. Hydrophilicity at the N-terminal end of the leader peptide and at the border between gp70 and p15^{env} (Figure 4) is maintained despite great differences in the amino acid sequences (Figure 3). At these sites, endopeptidases might cleave during glycoprotein maturation (see above).

The amino acid sequences of the hydrophilic peaks at positions 346–357 and 370–377 of the constant region and in the middle of p15^{env} are clearly conserved for all viral glycoproteins compared. Together with some smaller peaks they might well be part of the morphological substrate for interspecies-specific antigenic determinants of gp70 (Strand and August, 1973; Moennig *et al.*, 1973; Hunsmann *et al.*, 1974) and p15^{env} (Schäfer *et al.*, 1975). The comparison of the amino acid sequences (Figure 3) and the hydrophilicity profiles (Figure 4) suggests positions of group-specific antigenic determinants in the differential region of gp70. They should consist of hydrophilic peaks shared only by viruses of the same group. Type-specific determinants would be hydrophilic peaks unique for the differential region of a given virus subtype.

Table II. Amino acid homologies (in %) deduced for the four variable regions (VR) of GA-FeLV B, F-MCFV, F-MuLV and Akv

	VR1 (position 87–152)			VR 2 (position 201–238)			VR 3 (position 275–293)			VR 4 (position 308–353)		
	GA-FeLV	F-MCFV	F-MuLV	GA-FeLV	F-MCFV	F-MuLV	GA-FeLV	F-MCFV	F-MuLV	GA-FeLV	F-MCFV	F-MuLV
F-MCFV	32			39			37			15		
F-MuLV	9	6		3	11		11	32		4	15	
Akv	4.5	4.5	73	3	5	61	11	32	68	18	4	57

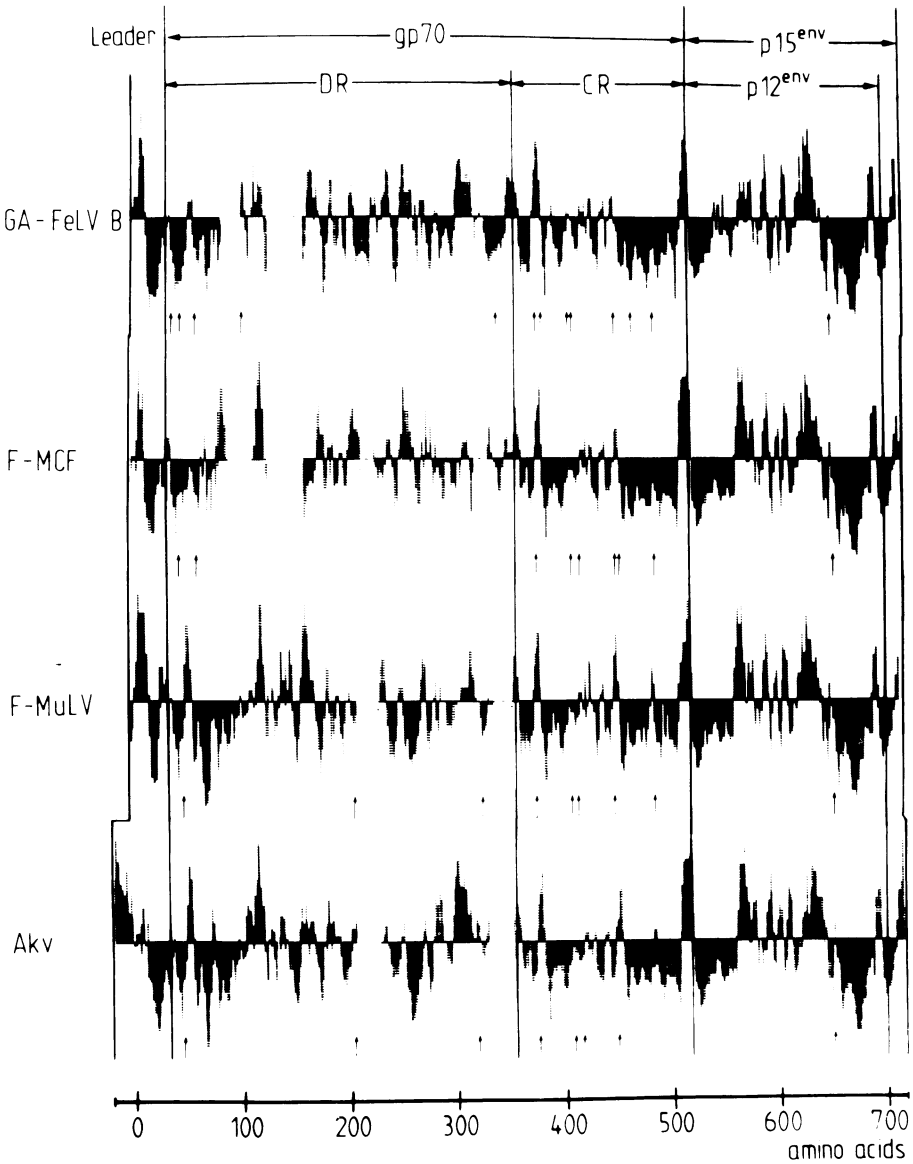

Fig. 4. Hydrophilicity profile of *env* gene polypeptides of GA-FeLV B, F-MCFV, F-MuLV and Akv. Hydrophilicity is shown by positive values. Arrows indcate possible glycosylation sites. Dotted lines refer to gaps introduced in the amino acid comparison of Figure 3. Peptide leader, gp70, differential (DR) and constant (CR) region and p12/15^{env} are marked as in Figure 3.

Short amino acid sequences are repeated in the env polypeptide

In earlier experiments, peptide maps of p15^{env} and gp70 of F-MuLV revealed overlapping spots suggesting a structural relatedness of both molecules (Schneider *et al.*, 1980). Within the amino acid sequence of GA-FeLV B, we noticed a remarkable similarity in the hydrophilicity profiles of the processed p15^{env} (p12^{env}; Figure 4), and the constant region of gp70. The extreme N terminus of gp70 is also partially reminiscent of this pattern. In these three regions, we found a similar pattern of four consecutive blocks of amino acid sequences (a,b,c, and d; Figure 5). They consist of up to 17 amino acid residues. Of these, up to five consecutive amino acid residues are identical (block a). In two other blocks (block b and c), equivalent exchanges (Ile/Leu) have occurred. Interestingly, block b contains a glycosylation site. The probability of such a consecutive order of distinctly spaced related sequences occurring repeatedly by chance is quite low. Underlying the amino acid comparison (Figure 3), we also found repeats in analogous positions of the other viruses, though these were less obvious (data not shown).

Based on these findings, we propose that two tandem duplications of a primordial *env*-related gene may have generated the framework for genes of current *env* polypeptides. A similar mechanism has been described for the evolution of immunoglobulin domains (Klein, 1982) and major histocompatibility antigens (Ohno *et al.*, 1982). This progenitor *env* gene may then have been diversified by deletions and recombinational exchanges of modular blocks of genetic information (VR 1–4; Figure 5). By these mechanisms, viruses may have acquired new characteristics influencing their host range, pathogenicity and escape from the host's immune surveillance.

If retroviral *env* genes in general have evolved in such a way it will be important to examine whether their genetic constituents described here are also found in the cellular genome separated from endogenous retroviral genes. Viral *env* polypeptides might interfere with products of such genes which

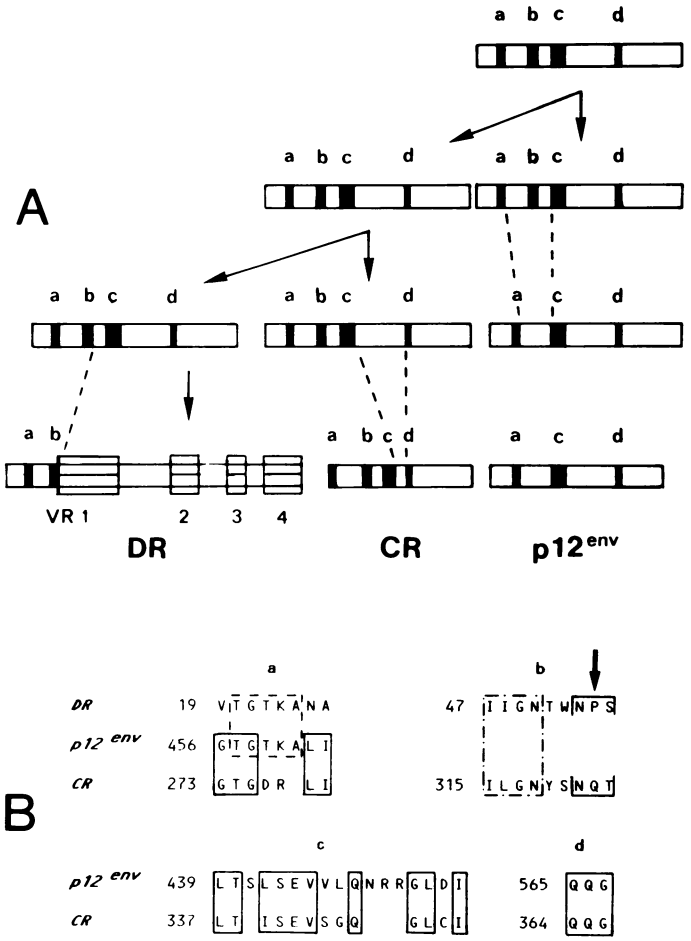

Fig. 5. Hypothesis of *env* gene evolution. (**A**) Model for the generation of p12/p15^{env}, constant region (CR) and differential region (DR, containing the four variable regions VR1 – VR4; striped boxes) from a primordial *env* gene. Black boxes (a,b,c and d) represent repeated amino acid sequences shown in (**B**). Bifurcated arrows indicate tandem duplications and dashed lines deletions. (**B**) Repeated amino acid sequences (blocks a,b,c and d) in the differential and constant region of gp70 and in p12^{env}. Homologous amino acids are shown in blocks. The arrow indicates a conserved glycosylation sequence in block b.

could have a role in differentiation of the haematopoietic system. This interaction could then lead to leukaemia.

Materials and methods

DNA preparation

The plasmid pKHR 4 containing the *env* gene and the 3′ LTR of a non-defective GA-FeLV B virus (Figure 1A) was kindly provided by J. Mullins (Mullins *et al.*, 1981). Plasmid-containing bacteria were grown in M9 medium as described (Koch *et al.*, 1983). DNA, prepared by a cleared lysate procedure (Clewell and Helinski, 1969) was further purified by banding twice in caesium chloride gradients. Plasmid DNA was then treated with 100 μg/ml proteinase K (E. Merck AG, Darmstadt, FRG) for 30 min at 37°C in the presence of 10 mM Tris/HCl (pH 7.0), 1 mM EDTA, and 0.2 mg/ml SDS. Finally, the DNA was extracted three times with phenol-chloroform (1:1) and precipitated with ethanol.

Restriction enzyme analysis and DNA sequencing

Restriction endonucleases were used as recommended by the manufacturers (Bethesda Research Laboratories, Bethesda, MD and New England Biolabs, MA). Restriction endonucleases requiring low or medium salt concentrations were used in 33 mM Tris/HCl (pH 7.4), 66 mM potassium acetate, 10 mM magnesium acetate, and 0.5 mM dithiothreitol. The resulting DNA fragments were separated for analytical and preparative purposes by electrophoresis either on agarose gels (0.8 – 1.2 g/100 ml) using a buffer of 40 mM Tris/HCl (pH 8.3), 20 mM sodium acetate, 1 mM EDTA or on polyacrylamide gels (5.0 – 7.5 g/100 ml) in 0.1 M Tris/HCl (pH 7.8), 0.1 M boric acid, 1 mM EDTA. For electrophoresis, samples were adjusted to 4 g/100 ml Ficoll (Pharmacia, Uppsala, Sweden), 0.1 g/100 ml bromphenol blue and 0.2 g/100 ml SDS. The DNA bands in the gels were stained with ethidium bromide (0.5 μg/ml) and visualized under u.v. light at 302 nm. DNA restriction fragments were eluted from agarose with glass powder by using the method of Vogelstein and Gillespie (1979). The phosphate at the 5′ end of recovered fragments was removed by alkaline phosphatase from calf intestine (Boehringer Mannheim, FRG). The 5′ ends were then labelled with [γ-^{32}P]ATP by polynucleotide kinase (New England Biolabs). Strand separation and sequence analysis were carried out by using the method of Maxam and Gilbert (1980). All parts of the genome were sequenced at least twice. Autoradiograms of sequencing gels were read independently by two workers.

Amino acid comparison and calculation of homologies

The amino acid sequence deduced from the GA-FeLV B *env* gene sequence was matched to those of Akv (Lenz *et al.*, 1982), F-MuLV (Koch *et al.*, 1983) and F-MCFV (Koch,W., Zimmermann,W., Oliff,A. and Friedrich,R., in preparation; Figure 3). Gaps were assumed to render the best homologies in differential regions. To quantify the homologies in certain regions for two viruses, the number of homologous positions was expressed as a percentage (Table II). Deletions in only one of the viruses were scored as non-homologues.

Acknowledgements

We thank James Mullins for generously providing the clone pKHR4, Paul MacIsaac, Rita Rossa and Heiko Taube for excellent technical assistance. Josef Schneider and Wolfgang Zimmermann contributed valuable comments. We are grateful to Manfred Kröger and Anneliese Kröger-Block for making available to us their computer programs used in this study (Kröger and Kröger-Block, 1982). The computer program for establishment of the hydrophilicity profiles was kindly provided by Walter Gruber. This research was supported by the Deutsche Forschungsgemeinschaft.

References

Blobel,G. and Dobberstein,B.D. (1975) *Int. J. Cell Biol.*, **67**, 835-851.
Bosselman,R.A., Van Straaten,F., Van Beveren,C., Verma,I.M. and Vogt, M. (1982) *J. Virol.*, **44**, 19-31.
Chattopadhyay,S.K., Cloyd,M.W., Linemeyer,D.L., Lander,M.R., Rands, E. and Lowry,D.R. (1982) *Nature*, **295**, 25-31.
Clewell,D.B. and Helinski,D.R. (1969) *Proc. Natl. Acad. Sci. USA*, **62**, 1159-1166.
Coffin,J. (1982) in Weiss,R., Teich,N., Varmus,H. and Coffin,J. (eds.), *RNA Tumor Viruses*, Cold Spring Harbor Laboratory Press, NY, pp. 261-368.
Donis-Keller,H., Rommelaere,J., Ellis,R.W. and Hopkins,N.H. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1642-1645.
Elder,J.H. and Mullins,J.I. (1983) *J. Virol.*, **46**, 871-880.
Engelman,D.M. and Steitz,T.A. (1981) *Cell*, **23**, 411-422.
Garten,W., Bosch,F.X., Linder,D., Rott,R. and Klenk,H.-D. (1981) *Virology*, **115**, 361-374.
Geyer,R., Geyer,H., Hunsmann,G., Schneider,J. and Stirm,S. (1982) *Biochim. Biophys. Acta*, **117**, 491-501.
Hampe,A., Gobbet,M., Even,J., Sherr,C.J. and Galibert,F. (1983) *J. Virol.*, **45**, 466-472.
Hardy,W.D.,Jr., Hess,P.W., MacEwen,E.G., McClelland,A.J., Zuckerman,E.E., Essex,M. and Cotter,S.M. (1976) *Cancer Res.*, **36**, 582-588.
Hayward, W.S. (1977) *J. Virol.*, **24**, 47-63.
Hopp,P. and Woods,K.R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3824-3828.
Hunsmann,G., Moennig,V., Pister,L., Seifert,E. and Schäfer,W. (1974) *Virology*, **62**, 307-318.
Jarret,O. (1982) in Yohn,D.S. and Blakeslee,J.R. (eds.), *Advances of Comparative Leukaemia Research*, Elsevier North Holland, Inc., NY, pp. 31-37.
Klein,J. (1982) in Klein,J. (ed.), *Immunbiology*, J. Wiley and Sons, NY, pp. 227-231.
Koch,W., Hunsmann,G. and Friedrich,R. (1983) *J. Virol.*, **45**, 1-9.
Kröger,M. and Kröger-Block,A. (1982) *Nucleic Acids Res.*, **10**, 229-236.
Leamnson,R.N., Shander,M.H.M. and Halpern,M.S. (1977) *J. Virol.*, **18**, 956-968.
Lenz,J., Crowther,R., Straceski,A. and Haseltine,W.A. (1982) *J. Virol.*, **42**, 519-529.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.
Mellon,P. and Duesberg,P.M. (1977) *Nature*, **270**, 631-634.
Moennig,V., Hunsmann,G. and Schäfer,W. (1973) *Z. Naturforsch.*, **28c**, 785-788.

Moennig,V., Frank,H., Hunsmann,G., Schneider,G. and Schäfer,W. (1974) *Virology,* **61**, 100-111.
Montreuil,J. (1980) *Adv. Carbohydr. Chem. Biochem.,* **37**, 157-223.
Mullins,J.I., Casey,J.W., Nicolson,M.O., Baumann Burck,K. and Davidson,N. (1981) *J. Virol.,* **38**, 688-703.
Naso,R.B., Arcement,L.J., Karshin,W.L., Jamjoon,G.A. and Arlinghaus, R.B. (1976) *Proc. Natl. Acad. Sci. USA,* **73**, 2326-2330.
Nordheim,A. and Rich,A. (1983) *Nature,* **303**, 674-679.
Ohno,S., Matsunaga,T., Epplen,J.T., Itakura,K. and Wallace,R.B. (1982) *Proc. Natl. Acad. Sci. USA,* **79**, 6342-6346.
Old,L.J. and Stockert,E. (1977) *Annu. Rev. Genet.,* **11**, 127-160.
Oroszlan,S., Henderson,L.E., Copeland,T.D., Schultz,A.M. and Rabin,E.M. (1980) in Koch,G. and Richter,D. (eds.), *Biosynthesis, Modification and Processing of Cellular and Viral Polyproteins,* Academic Press, Inc., NY, pp. 219-232.
Pinter,A. and Fleissner,E. (1977) *Virology,* **83**, 417-422.
Pinter,A. and Fleissner,E. (1979) *J. Virol.,* **30**, 157-165.
Robinson,P.J., Hunsmann,G., Schneider,J. and Schirrmacher,V. (1980) *J. Virol.,* **36**, 291-294.
Ruscetti,S., Davis,L., Field,J. and Oliff,A. (1981) *J. Exp. Med.,* **154**, 907-920.
Sarma,P.S. and Log,T. (1973) *Virology,* **54**, 160-169.
Sarma,P.S., Log,T., Jain,D., Hill,P.R. and Huebner,R.J. (1975) *Virology,* **64**, 438-46.
Schäfer,W., Hunsmann,G., Moennig,V., Noronha,F., Bolognesi,D.P., Hüper,G. and Green,R.W. (1975) *Virology,* **63**, 48-59.
Schneider,J., Falk,H. and Hunsmann,G. (1980) *J. Virol.,* **33**, 597-605.
Schneider,J. and Hunsmann,G. (1978) *Int. J. Cancer,* **22**, 204-213.
Shinnick,T.M. Lerner,R.A. and Sutcliffe,J.G. (1981) *Nature,* **293**, 543-548.
Stockert,E., Old,L.J. and Boyse,E.A. (1971) *J. Exp. Med.,* **133**, 1334-1355.
Strand,M. and August,J.T. (1973) *J. Biol. Chem.,* **248**, 5627-5633.
Temin,H.M. (1982) *Cell,* **28**, 3-5.
Tung,J.G., Vitetta,E.S., Fleissner,E. and Boyse,E.A. (1975) *J. Exp. Med.,* **141**, 198-205.
Velicer,D.F. and Graves,D.C. (1974) *J. Virol.,* **14**, 700-703.
Vogelstein,B. and Gillespie,D. (1979) *Proc. Natl. Acad. Sci. USA,* **76**, 615-619.
Witte,O.N. and Wirth,D.F. (1979) *J. Virol.,* **29**, 735-743.

Note added in proof

While this manuscript was in preparation, the sequence of the *env* gene and LTR of another DNA clone of GA-FeLV B has been published (Elder and Mullins, 1983). Both sequences are identical with the exception of a single nucleotide in the LTR at position 2287, where our sequence contains T instead of A.