Nucleotide sequence of the gag gene and gag-pol junction of feline leukemia virus.

I Laprevotte, A Hampe, C J Sherr and F Galibert
***J. Virol.* 1984, 50(3):884.**

JOURNAL OF VIROLOGY, June 1984, p. 884–894
0022-538X/84/060884-11$02.00/0

Vol. 50, No. 3

Nucleotide Sequence of the *gag* Gene and *gag-pol* Junction of Feline Leukemia Virus

IVAN LAPREVOTTE,[1]* ANNIE HAMPE,[1] CHARLES J. SHERR,[2] AND FRANCIS GALIBERT[1]

Laboratoire d'Hématologie Expérimentale, Centre Hayem, Hôpital Saint-Louis, 75475 Paris Cédex 10, France,[1] and Division of Human Tumor Cell Biology, St. Jude Children's Research Hospital, Memphis, Tennessee 38101[2]

Received 24 October 1983/Accepted 23 January 1984

The nucleotide sequence of the *gag* gene of feline leukemia virus and its flanking sequences were determined and compared with the corresponding sequences of two strains of feline sarcoma virus and with that of the Moloney strain of murine leukemia virus. A high degree of nucleotide sequence homology between the feline leukemia virus and murine leukemia virus *gag* genes was observed, suggesting that retroviruses of domestic cats and laboratory mice have a common, proximal evolutionary progenitor. The predicted structure of the complete feline leukemia virus *gag* gene precursor suggests that the translation of nonglycosylated and glycosylated *gag* gene polypeptides is initiated at two different AUG codons. These initiator codons fall in the same reading frame and are separated by a 222-base-pair segment which encodes an amino terminal signal peptide. The nucleotide sequence predicts the order of amino acids in each of the individual *gag*-coded proteins (p15, p12, p30, p10), all of which derive from the *gag* gene precursor. Stable stem-and-loop secondary structures are proposed for two regions of viral RNA. The first falls within sequences at the 5′ end of the viral genome, together with adjacent palindromic sequences which may play a role in dimer linkage of RNA subunits. The second includes coding sequences at the *gag-pol* junction and is proposed to be involved in translation of the *pol* gene product. Sequence analysis of the latter region shows that the *gag* and *pol* genes are translated in different reading frames. Classical consensus splice donor and acceptor sequences could not be localized to regions which would permit synthesis of the expected *gag-pol* precursor protein. Alternatively, we suggest that the *pol* gene product (RNA-dependent DNA polymerase) could be translated by a frameshift suppressing mechanism which could involve cleavage modification of stems and loops in a manner similar to that observed in tRNA processing.

Feline leukemia virus (FeLV) is a horizontally transmitted retrovirus which is a natural etiological agent of leukemia in domestic cats (13). FeLV contains a diploid RNA genome, each strand encoding the three genes (*gag*, *pol*, and *env*) necessary for viral replication. The *gag* gene encodes the structural proteins of the virion core; the *pol* gene encodes an RNA-dependent DNA polymerase; and the *env* gene encodes the envelope glycoproteins of the virion surface (2). Three subgroups of FeLV differ from one another in their host range, neutralization, and viral interference, all of which are properties of the *env* gene product (38). Like other retroviruses, FeLV replication depends on the formation of a DNA provirus which integrates into the DNA of the infected host cell (53). Although the order of genes in the DNA provirus is colinear with that of viral RNA, retroviral DNA intermediates differ at their ends from viral RNA by the presence of long terminally redundant sequences (LTRs), formed during the process of reverse transcription (reviewed in references 51 and 52).

In the present study, we determined the nucleotide sequence of 2,565 base pairs representing a segment of a previously cloned FeLV (subgroup B) DNA intermediate (46). The sequenced region includes a portion of the 5′ LTR, the *gag* leader, the complete *gag* gene, and 389 base pairs of the *pol* gene. The results were interpreted partly by comparison with those previously obtained with the Moloney strain of murine leukemia virus (Mo-MuLV) (8, 47) and with two previously sequenced strains of feline sarcoma virus (FeSV) (17).

* Corresponding author.

MATERIALS AND METHODS

The molecular clone of FeLV DNA was previously characterized by restriction enzyme mapping and heteroduplex analyses (46). The recombinant λ-WES phage was produced in liquid culture, and the cloned DNA insert was isolated after release from the vector with *Eco*RI. A subgenomic *Eco*RI-*Pst*I fragment representing a portion of the 5′ LTR, the *gag* gene, and a portion of the *pol* gene was purified, and its sequence was determined from the *Eco*RI site to an *Mbo*II site 2,565 base pairs downstream. The nucleotide sequence was established by the Maxam and Gilbert method by using five specific reactions of G, AG, CT, C, and AC (19, 27) (Fig. 1). Except for one case, 5′ end labeling was performed. For one of two hydrolyses done with *Hin*fI-*Sau*3A, 3′ end labeling was undertaken with a mixture of [α-^{32}P]dATP and unlabeled dGTP in the presence of DNA polymerase I. The complete sequence of both strands was determined by using a variety of different end labels such that all labeled termini appeared as internal sites within other sequenced fragments.

RESULTS AND DISCUSSION

Organizational landmarks. The nucleotide sequence of the cloned FeLV DNA fragment is presented in Fig. 2. By convention, the cap site corresponding to the 5′ end of FeLV RNA was numbered +1. Sequences upstream of the cap site include a portion of the U3 region of the viral 5′ LTR (16). These contain the CAAT box (positions −74 to −66) and the Hogness box (positions −31 to −25), characteristic of eucaryotic promotors (reviewed in reference 51). As previously reported (16), a polyadenylation signal and polyadenylate addition site are located downstream of the cap site within

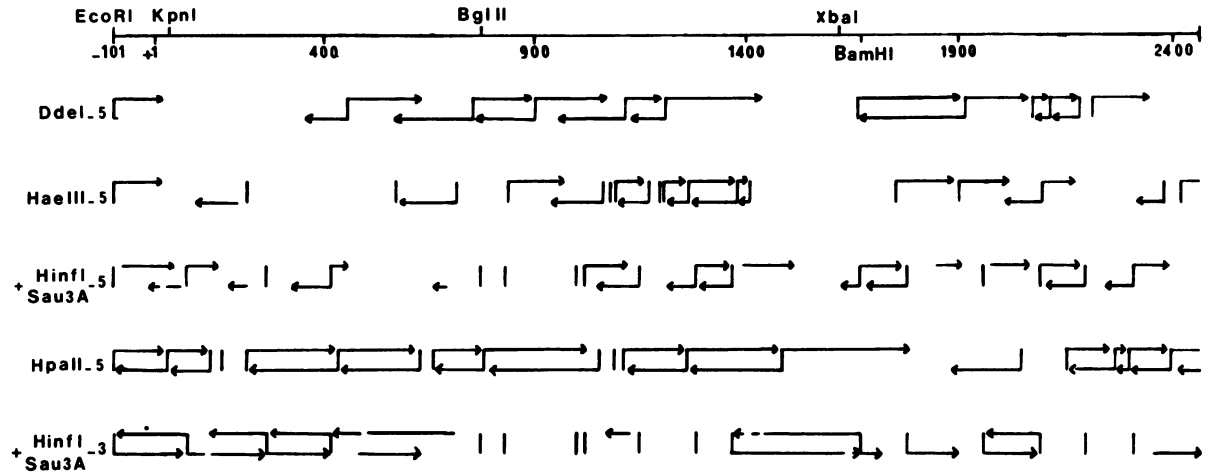

FIG. 1. Sequencing strategy. The upper line scheme represents the sequenced zone as a whole. Distances are given in base pairs (Fig. 2). Restriction sites indicated on this line are those previously known (46). Arrows give the length of the segments analyzed on each of the two strands (+ strand above, − strand below), starting from the labeled termini indicated by the vertical bars. -5 and -3 stand for 5′ and 3′ labeling, respectively.

the R region of the 5′ LTR and appear to function in terminating transcripts within the 3′ LTR near the other terminus of proviral DNA. The end of the inverted repeat which marks the end of the 5′ LTR appears at nucleotide 144 and is followed by the tRNA primer binding site at positions 145 to 162.

Two ATG codons at positions 344 to 346 and 566 to 568 are in frame with one another and are separated by 222 base pairs. The second ATG codon corresponds in position to the amino terminus of the *gag* gene precursor, Pr65^{gag} (17). The *gag* reading frame ends at a TAG codon starting at nucleotide 2084. A second open reading frame begins after the TAG codon starting at position 2076 and continues to the *Mbo*II cleavage site in the *pol* gene at nucleotide 2464. The latter segment, which corresponds to the 5′ end of the *pol* gene, differs in frame from that of the *gag* gene (see below).

The individual *gag* gene polypeptides. The known order of processed polypeptides in the *gag* precursor, Pr65^{gag}, is NH_2-p15-p12-p30-p10-COOH (22). By comparison with protein sequences (17), the p15-p12, p12-p30, and p30-p10 junctions occur at nucleotides 947, 1157, and 1901, respectively. The carboxy terminus of p10 occurs at nucleotide 2071 and is followed by a 12-nucleotide coding sequence before the TAG stop codon at positions 2084 to 2086. The sites of proteolytic cleavage of individual *gag* peptides have serine or threonine at a position four amino acids to their amino-terminal side: (p15) NH-SerSerLeuTyr ↓ (p12) ProValVal-OH; (p12) NH-SerGlnAlaLeu ↓ (p30) ProLeuArgGlu-OH; (p30) NH-ThrLysValLeu ↓ (p10) AlaThrValVal-OH; (p10) NH-SerThrLeuLeu ↓ AsnLeuGluAsp-OH (end of *gag* gene). The immediate neighborhood of the junctions is hydrophobic: on the average, the four amino acids which border the junctions have a hydrophobicity index (41) of +2 to +3.1 and a hydrophilicity value (H.V.) (20) of −1 to −1.4. The corresponding values for Mo-MuLV are +2.1 to +2.75 and −1.1 to −1.45.

The calculated molecular weights of individual *gag*-coded peptides (Table 1) are compatible with values obtained previously (18, 57). For p12 and p10, the values were overestimated by gel filtration in the presence of 6 M guanidine hydrochloride (50). Existence of 31.5% proline in p12 of FeLV (Fig. 2) might account for the discrepancy, as a result of effects on secondary structure (26). For p10, the discrepancy was probably within experimental error since more recent estimates for the size of this polypeptide are closer to 7,000 daltons (18).

Approximate calculations of the degrees of acidity or basicity (Table 1) are compatible with the published FeLV isoelectric point values (50) for p15, p30, and p10; for p12 there is disagreement (pI = 5.4), probably because of the phosphorylation of this polypeptide (34, 43).

Glycosylated and nonglycosylated *gag* gene products. The RNA leader sequence upstream of the FeLV *gag* gene includes a portion of the 5′ LTR (the R region: nucleotides 1 to 145), a noncoding sequence of 199 additional nucleotides (nucleotides 146 to 344), and a potential coding sequence (nucleotides 345 to 566) 5′ to the *gag* p15 amino terminus. The open reading frame in the FeLV leader sequence is in frame with the predicted *gag* gene sequence and can code for a polypeptide of about 7.6 kilodaltons. The open reading frame has also been conserved in different strains of FeSV which contain the same FeLV-derived sequences (see Fig. 2) and contains signal sequences thought to be responsible for directing nascent polypeptide chains to membranes of the rough endoplasmic reticulum (12; reviewed in reference 16). Thus, we propose that the ATG at positions 344 to 346 specifies the amino terminus of a glycosylated *gag* precursor, gPr80^{gag}, whereas that at positions 567 to 569 specifies the amino terminus of the nonglycosylated *gag* precursor Pr65^{gag}. Although both glycosylated and nonglycosylated forms of the *gag* precursor are indeed detected in cells infected with FeLV (32, 45), the mechanism for regulating the translation of both proteins remains unknown. One possibility is RNA splicing between a donor site at nucleotides 203 to 209 and a potential acceptor sequence just 3′ of the AG at position 456.

These results are consistent with data obtained in other systems. Existence of a *gag* glycoprotein is well documented for Mo-MuLV (10, 25, 39, 56; reviewed in reference 9). This protein is found on the plasma membrane.

Glycosylation sites within gPr80^{gag} were experimentally determined for Mo-MuLV and were identified in p15 and p30

(39; reviewed in reference 9). In the case of FeLV, there are only two possible sites which comply with the sequence Asn-X-Ser/Thr, one within p15 at position 743 (Asn-Ile-Ser) and the other near the end of p10 at position 2057 (Asn-Ser-Thr). Only the latter site would be contained in the extracellular *gag* glycoprotein fragment with a molecular weight of 40,000 (32). By analogy with Mo-MuLV, it appears equally likely that the site within FeLV p15 is also glycosylated.

Hydrophilic sectors. Different hydrophilic zones have been located; some of them could be antigenic sites (20). Among them, two stretches at the p30 level have a maximum hydrophilicity (positions 1823 to 1838 and 1856 to 1874; H.V. = +3); they appear to be duplicated zones, as has been described in murine retroviruses (33, 47). At the p15 level, the sector of maximum hydrophilicity (positions 650 to 665; H.V. = +2.25) is situated in a block of homology (Fig. 2); another hydrophilic zone (positions 740 to 764; H.V. = +1.04) indeed contains the hypothetic glycosylation site. At the p12 level the maximum hydrophilicity sequence (positions 1109 to 1127; H.V. = +2.3) is also situated in a block of homology and could be connected with the phosphorylation site (see below). p10 also has a sector of maximum hydrophilicity (positions 1922 to 1940; H.V. = +3) in a zone which is only partially homologous with the corresponding one for Mo-MuLV (Fig. 2).

Comparison of the FeLV and FeSV *gag* genes. The Snyder-Theilen (ST) and Gardner-Arnstein (GA) strains of FeSV arose by recombination between FeLV and protooncogene sequences in cat cellular DNA (14, 48). Both sarcoma viruses contain a truncated *gag* gene which is fused in frame to the viral oncogene sequences required for transformation (17). The FeLV subgroup B analyzed in these studies was molecularly cloned as the natural helper virus of ST-FeSV. The natural FeLV helper of GA-FeSV has been cloned by others (29), and portions of its sequence, not including the *gag* gene, have been reported (11).

Figure 2 shows that the ST-FeSV *gag* sequence ends after nucleotide 1457 and that of GA-FeSV ends after nucleotide 1611. In the region where ST-FeSV, GA-FeSV, and FeLV *gag* sequences could be compared, only six base changes (transitions) between FeLV subgroup B and ST-FeSV were found, four of which lead to amino acid differences. By contrast, 84 base differences (with 69 transitions) were detected between FeLV subgroup B and GA-FeSV, leading to 16 amino acid changes. The latter include 10 alterations in the signal peptide, 3 within p15 and 3 within p12.

Although ST- and GA-FeSV represent independent isolates, they appear to have arisen from a single recombination event between FeLV and c-*fes* protooncogene sequences, since the 3′ FeLV/v-*fes* recombination sites are identical in both strains (17). Subsequent deletion and mutational divergence and recombination with the replication competent FeLV helper could explain the strain-specific differences between the two sarcoma viruses. Another hypothesis remains, as suggested by hybridization and immunological studies (36): independent FeSV isolates could have arisen from different FeLV subgroups. If this were the case, the unique 3′ FeLV/v-*fes* recombination site would be an intriguing datum which could be explained by additional work on the mechanisms of recombination. In any case, the answers to these questions deserve further investigation.

Comparison with the Mo-MuLV sequence. In regions where there is no ambiguity in the relative position of the FeLV and MuLV sequences (Fig. 2), ca. 330 transitions and 365 transversions were noted. Between the cap site and the beginning of the gPr80^{gag} coding region, there are variations in sequence homology between 42 and 86%. From the beginning of gPr80^{gag}, the percentage of nucleotide homology remains close to the percentage of amino acid homology. In general, the percentage of nucleotide homology exceeds 50% for p15, p12, and p10 and 60% for p30 and the part of the *pol* gene studied. The data are consistent with the hypothesis that FeLV was ancestrally derived from viruses of rodent origin which were horizontally transmitted between different species (3).

The overall percentages of amino acid homology with Mo-MuLV are ca. 43.5% for the signal peptide, 66.5% for p15, 56% for p12, 68% for p30, 55% for p10, and 77% for the first 128 *pol* residues. The blocks of complete homology are not very great for the signal peptide (Fig. 2). Apart from the first three amino acids and other groups of two or three residues, there are entire portions without homology. The only extended homology within the signal peptide (90%) runs from the serine at nucleotide position 467 to the valine at position 500. However, besides these changes in amino acid sequence, the hydrophobicity of the sequence coded by nucleotides 497 to 556 is highly maintained. There are four blocks of extended amino acid homology within p15 which include the first uncharged 12 amino acid residues and FeLV sequences 632 to 674, 704 to 731, and 764 to 812. Comparison with the polypeptide sequences of the Rauscher strain of MuLV and Mo-MuLV p12 shows at this level a first block of 85% homology (positions 1007 to 1043) which ends with ProProProTyr. This tyrosine might be the only phosphorylation-mediated p12 nucleotidylation point (57). There is another block of homology (75%; positions 1094 to 1127) including two serines and an intriguing hydrophilic zone seven amino acids long (from serine at position 1106) which includes four arginines (H.V. = +2.3, the maximum for p12); although not already described (reviewed in references 24, 34, 43, and 49), a phosphorylation sector could be hypothesized at this level. Figure 2 shows that the entire p30

TABLE 1. Molecular weights of the different polypeptide sectors and the degrees of acidity or basicity[a]

Peptide	Calculated mol wt	b − a	b − a/b + a	ΣpK/ΣfRd	pI
Leader	7,662	+2	+0.2	8.2	
p15	14,155	+6	+0.25	7.85	7.5
p12	7,366	−1	−0.06	7.5	5.4
p30	28,898	+4	+0.055	8	7.6
p10	6,591	+8	+0.35	8.5	9.1
Pr65	57,429	+15	+0.1	8	
P75	65,073	+17	+0.11	8.1	

[a] Dry molecular weights were calculated from the polypeptide sequence deduced from the nucleotide sequence (Fig. 2). b − a is the difference between the sum of His, Lys, and Arg and the sum of Glu and Asp (6); b + a is the sum of the whole. ΣpK/ΣfRd is the ratio of the sum of the free acid and basic radical pK to the total number of these radicals (it is a case of global appreciation of the degrees of acidity or basicity, taking into account the dissociation constants of the free radicals). The isoelectric points indicated are values published for FeLV (50). The methionine at the beginning of P75 (gPr80^{gag}), Pr65^{gag}, and the leader peptide, and that observed at the beginning of p15, are excluded from the calculation of the molecular weights of the latter polypeptides. The methionine at the beginning of p15 has been included in the leader peptide. The molecular weights of p15, p12, p30, and p10 do not equal Pr65 because the carboxy terminus of the predicted *gag* polyprotein is four amino acids longer than the carboxy terminus of p10, as it is in Mo-MuLV (47).

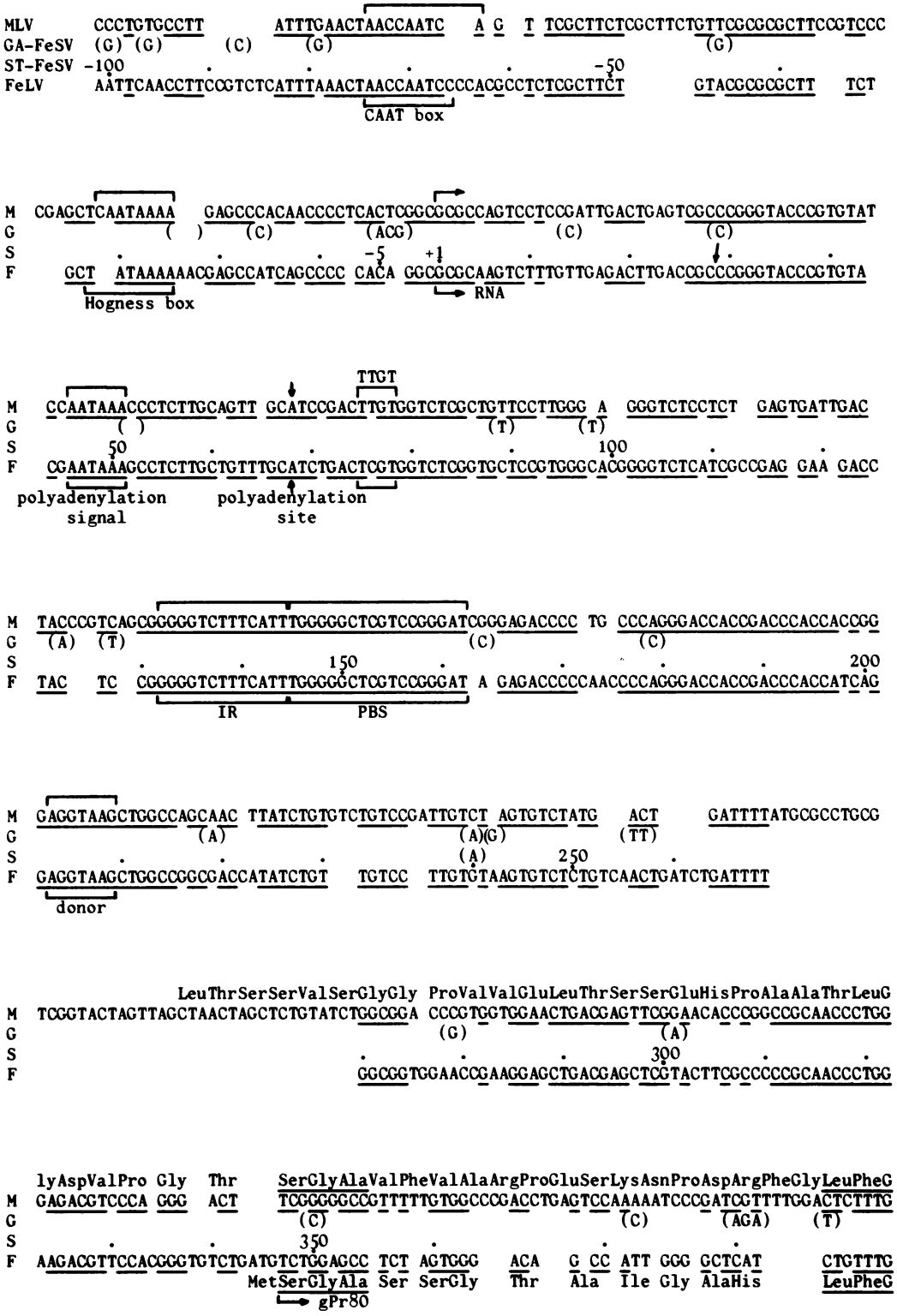

FIG. 2. The FeLV subgroup B sequence at the level of the first 2,565 base pairs starting from the *Eco*RI site of the LTR, 5′ side. Orientation is that of the viral RNA. Only the DNA (+) strand representing the viral RNA (starting from the capping site) is represented in this context. The numbering noted at this strand acts as a reference for the locations at this nucleotide chain, those with which it is compared, and the amino acids translated from the nucleotide chains (located by the position of the first base of the codon). The position +1 is the first base of viral RNA (capping site); from this position the nucleotides are numbered negatively on the 5′ side and positively on the 3′ side up to position 2464; the dots which are regularly placed upon some letters are 10 to 10 graduations of DNA (+) sequence. The sequence has been written below the aligned sequence of murine sarcoma virus (8) up to the beginning of the viral RNA and subsequently below the aligned sequence of Mo-MuLV (MLV; M) (47). Gaps have been inserted in the writing to maintain maximum agreement between the two sequences. In actual fact the nucleotides which border these gaps are contiguous. The same means of presentation has been adopted for the amino acids corresponding

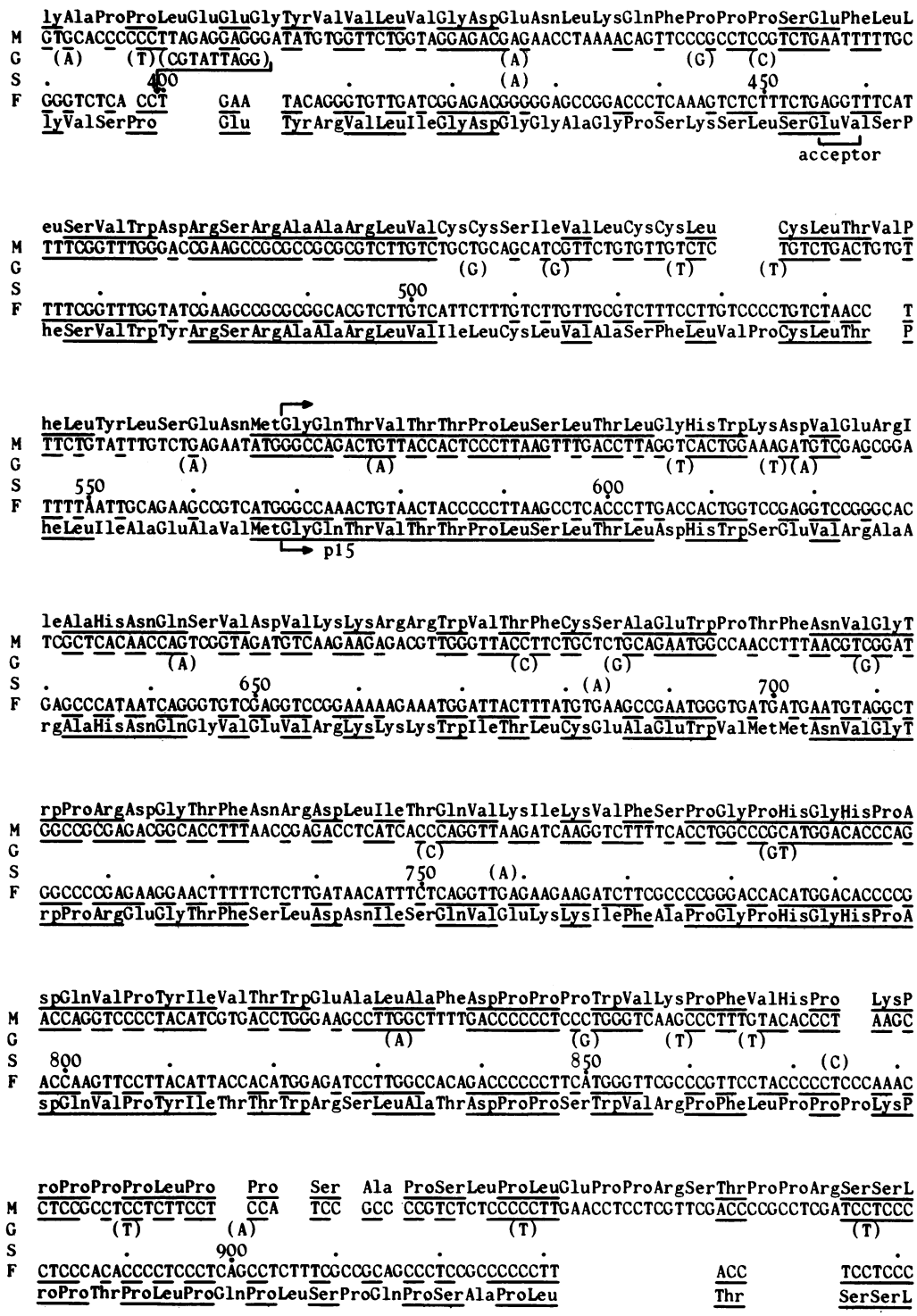

FIG. 2—*Continued*

to translation of the nucleotide chains in the known coding zones. Nucleotide or polypeptide identities between the two viruses have been underlined. The correspondences established are based on blocks of homology. In certain places several solutions can be proposed which do not contradict the greater part of the homologies. Between the two sequences of FeLV (F) and Mo-MuLV, a comparison has been made between FeLV and GA-FeSV (G) first and ST-FeSV (S) second, up to the two dots indicating divergence on the 5′ side. Nucleotides different from their correspondents in FeLV have been labeled with parentheses. Missing nucleotides are represented by empty parentheses. Excess nucleotides with regard to FeLV are between parentheses with an arrow indicating at what place they are inserted in the sequence. The locations of the different patterns, the locations of the divergence with ST-FeSV and with GA-FeSV, and the limits of the different polypeptide sectors are indicated (see the text). The TAG codons at the end of *gag* and the beginning of *pol* are indicated by three dots.

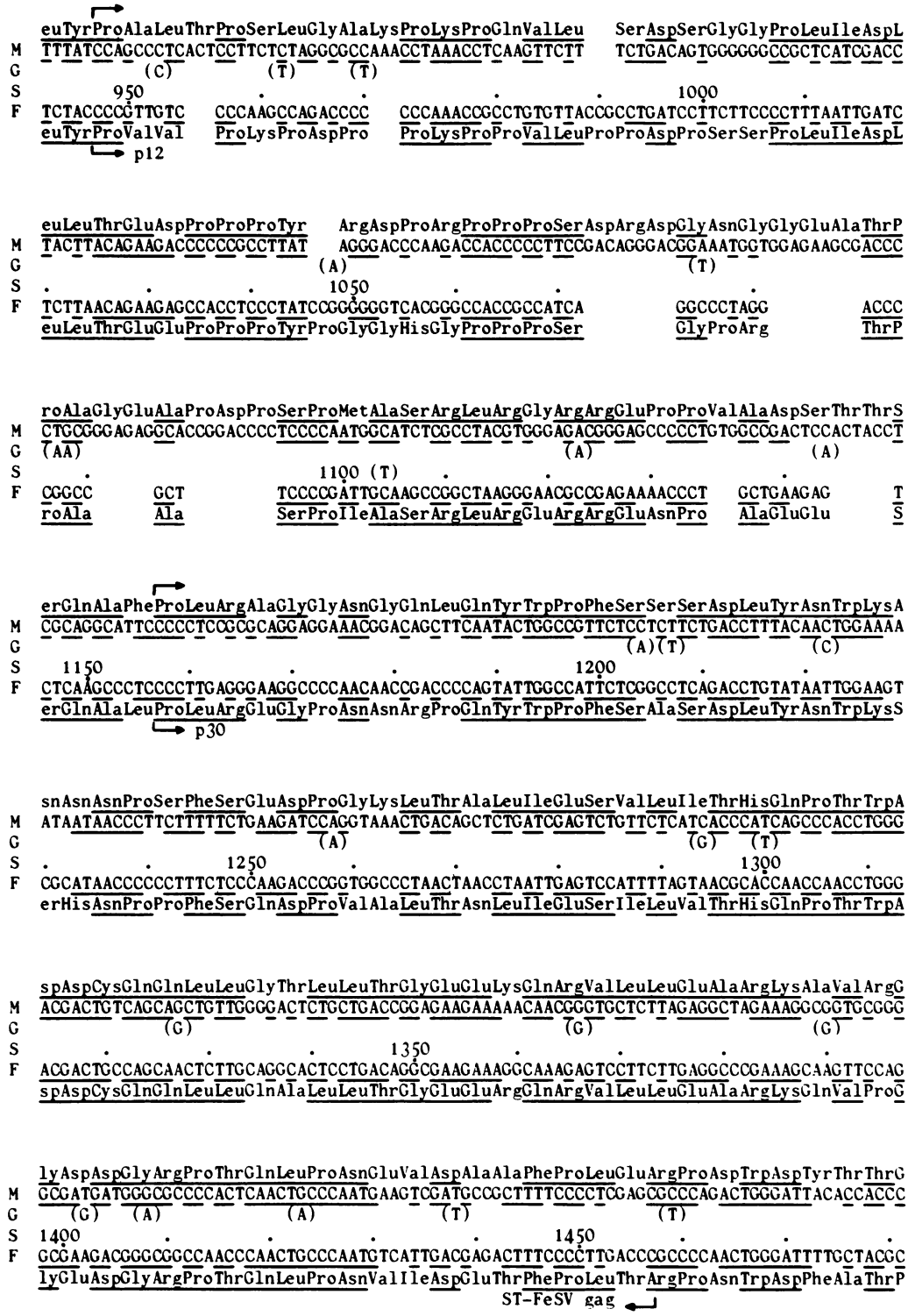

FIG. 2—*Continued*

sequence is made up of long blocks of homology. Among the homologous regions, the region which runs from glutamine at position 1187 to asparagine at position 1427 is particularly conserved among mammalian retroviruses (reviewed in references 21 and 33). A zone of extensive homology within p10 runs from the aspartic acid at position 1982 to the arginine at position 2042 (Fig. 2). The major antigenicity of p10 is group specific with strong interspecific determinants (reviewed in reference 9), possibly located within the conserved region. The conserved region could functionally correspond to the site of fixation of p10 with the viral RNA (18).

Stems, loops, and palindromes. A search for potentially

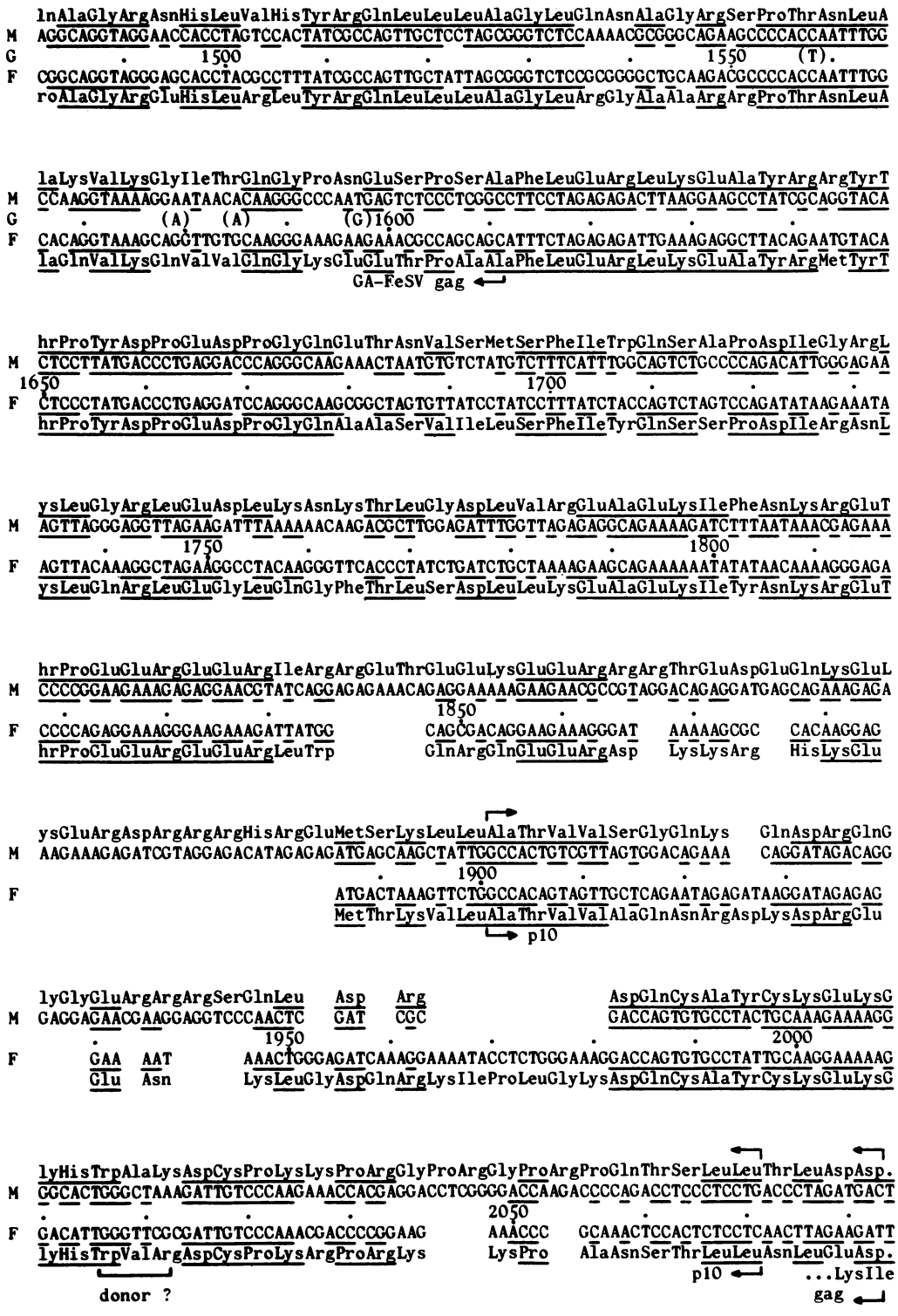

FIG. 2—*Continued*

stable (<−10 and <−0.4 kcal [ca. <−41.8 and <−1.7 kJ] per nucleotide for stems and loops, respectively; <−19.2 kcal [ca. <−80.3 kJ] for palindromes) secondary structural features was performed by using previously described methods (37, 54, 55).

These features are more numerous within the 500 first bases of the viral RNA. Among the 5 palindromes (out of 11 in the whole sequence) situated in this sector, which could be hypothesized as being possible linkage points of the two RNA subunits (5, 31), the more stable (−35.8 kcal [ca. −149.8 kJ]) is at position 483 (Fig. 3). A similar palindromic zone was observed for Rous sarcoma virus (40). A stem and

FIG. 2—*Continued*

loop of −13 kcal (ca. −54.4 kJ) (positions 102 to 129) is possible near the 5′ end of viral RNA and has previously been proposed for Moloney sarcoma virus (4). A larger stem and loop of −75.1 kcal (ca. −314.2 kJ) (positions 70 to 201) contains the LTR inverted repeat and the primer binding site, both of which are located in the double loop at the stem summit (Fig. 3). In the case of Mo-MuLV, a stem and loop visualized by electron microscopy was localized between the 5′ end of viral RNA and the stable linkage point of the two RNA subunits (31). This location appears to correspond to that predicted for FeLV. Overall, the first 500-base region of FeLV RNA could generate a secondary structure similar to that described for Rous sarcoma virus, physically linking the cap site, possible ribosome-binding site, and initiation codon of the *gag* gene (7).

A possible ribosome-binding site is encountered at position 23:

23-5′GA CC^{G} CC CGGG3′-33
3′$CU_{A}GG_{AA}GG_{C}$GUCC5′ ← 3′ end of 18S rRNA
(−19.5 and −0.78 kcal [ca. −81.6 and −3.26 kJ] per nucleotide)

It overlaps a set of two palindromes of −22.2 and −31.2 kcal (ca. −92.9 and −130.5 kJ) (Fig. 3), 100% homologous in Mo-MuLV, which include restriction sites *Kpn*I, *Ava*I, *Hpa*II, *Rsa*I, and *Sma*I. Another possible ribosome-binding site is encountered at position 325, a little before the ATG at the beginning of $gPr80^{gag}$ (Fig. 2):

325-5′GA $C^{G}TTCC^{A}$CGGGTGTCTGATG3′-346
3′$CU_{A}G_{G}AAGG_{C}$GUCC5′ ← 3′ end of 18S rRNA
(−15.85 and −0.59 kcal [ca. −66.31 and −2.47 kJ] per nucleotide)

One could hypothesize that these two possible ribosome-binding sites could be used in an alternating way and could be responsible for the discriminatory translations of $Pr65^{gag}$ and P75 ($gPr80^{gag}$) by yet unknown mechanisms in which secondary structures could take part.

A third stem-and-loop structure proposed between nucleotides 2060 and 2124 (Fig. 3) includes the *gag-pol* junction. It constitutes the most stable predicted structure (−37.9 kcal [ca. −158.6 kJ]) apart from the first 500 bases of viral RNA. This zone is very homologous with Mo-MuLV in nucleotides (79%) but less so in amino acids (50%).

The *gag-pol* junction. Reverse transcriptase is thought to arise from the posttranslational processing of a large polyprotein ($Pr180^{gag\text{-}pol}$) encoded by the *gag* and *pol* genes of full-genome-length mRNA. In the case of MuLV, the *gag-pol* polyprotein is presumed to be generated by in-frame suppression of the *gag* gene stop codon, allowing translation to continue into the *pol* sequences (30, 35). By contrast, in vitro translation of Rous sarcoma virus RNA in the presence of suppressor tRNA resulted in the synthesis of a polyprotein only slightly longer than the *gag*-coded precursor (58). Recent nucleotide sequencing indeed showed that the RSV *gag* and *pol* genes are in two different reading frames separated by 20 nucleotides (40).

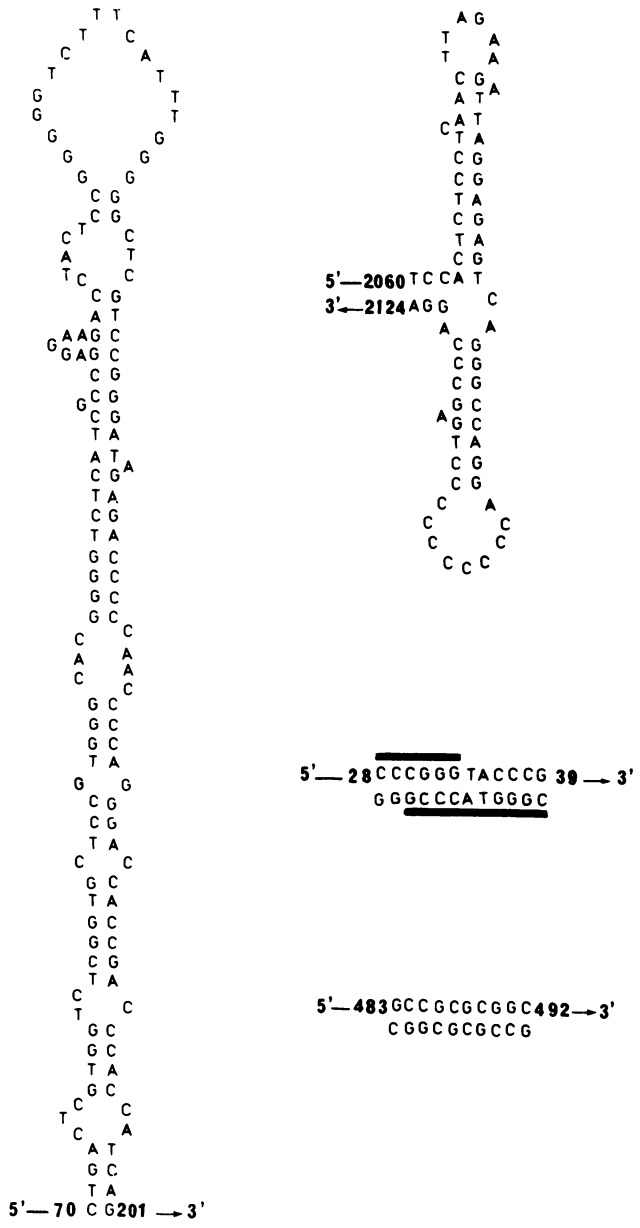

FIG. 3. The two most stable palindromic zones and stems and loops, from the energy point of view, in the viral RNA sequence. Numbered sequences correspond to the DNA (+) strand and its numbering (Fig. 2). Numbers give the position of the first and last letter of each pattern or group of patterns. For palindromes the unnumbered complementary chain has been added opposite. The structure which runs from positions 28 to 39 is made up of two overlapping palindromes, each of which is indicated by a thick line.

Like RSV, the FeLV *gag* and *pol* genes are in different reading frames, suggesting that *pol* is translated from a processed mRNA. One possibility is that the *pol* gene product is synthesized from a spliced mRNA molecule in which the *gag* and *pol* genes are fused in frame. A possible splice acceptor sequence around nucleotide 2117 (CCCCCCCCCCTGAG↓) (28, 42, 44) occurs close to the *gag-pol* junction. However, no typical donor groupings were found in the same region of the FeLV genome. The closest consensus sequence for a potential splice donor occurs around nucleotide 2017 within p10 (TGG↓GTTCGC-3') and shows only limited homology (four of nine bases) to expected donor sequences ($^{C}_{A}$AG↓GT$^{A}_{G}$AGT) (28). Taking into account the proposed splice acceptor, the latter donor site could place the reading frames into continuity without producing a major deletion of *gag* sequences.

Another possibility is that a tRNA-specific splicing enzyme might recognize a stem and loop like that proposed at the *gag-pol* junction (Fig. 3; see above) and produce an excision which shifts the reading frame (1, 23, 47), similar to what has been observed with other tRNA-specific enzymes (reviewed in reference 15). Although the latter represents a novel mechanism for frame suppression in translation of retroviral RNA, like splicing, it is compatible with the expected size of the *gag-pol* precursor and with the inability to detect subgenomic mRNA species which could account for a unique *pol* mRNA species. Regardless of the exact mechanism, the nucleotide sequences of both FeLV and RSV emphasize that the modes of translation of retroviral polymerases deserve further investigation.

ACKNOWLEDGMENTS

We thank Carmen Bernard and Christian-Jacques Larsen for helpful discussion, Michel Cartigny and Alan Kay for helping to draw up microcomputer programs, and Jacqueline Calliat for typing the manuscript.

This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), France, through the Service Commun N°15.

LITERATURE CITED

1. **Altman, S.** 1981. Transfer RNA processing enzymes. Cell **23**:3–4.
2. **Baltimore, D.** 1975. Tumor viruses: 1974. Cold Spring Harbor Symp. Quant. Biol. **39**:1187–1200.
3. **Benveniste, R. E., C. J. Sherr, and G. J. Todaro.** 1975. Evolution of type C viral genes: origin of feline leukemia virus. Science **190**:886–888.
4. **Benz, E. W., Jr., R. M. Wydro, B. Nadal-Ginard, and D. Dina.** 1980. Moloney murine sarcoma proviral DNA is a transcriptional unit. Nature (London) **288**:665–669.
5. **Chien, Y.-H., R. P. Junghans, and N. Davidson.** 1980. Electron microscopic analysis of the structure of RNA tumor virus nucleic acids, p. 395–446. *In* J. R. Stephenson (ed.), Molecular biology of RNA tumor viruses. Academic Press, Inc., New York.
6. **Czernilofsky, A. P., A. D. Levinson, H. E. Varmus, J. M. Bishop, E. Tischer, and H. M. Goodman.** 1980. Nucleotide sequence of an avian sarcoma virus oncogene (*src*) and proposed amino acid sequence for gene product. Nature (London) **287**:198–203.
7. **Darlix, J.-L., M. Zuker, and P. F. Spahr.** 1982. Structure-function relationship of Rous sarcoma virus leader RNA. Nucleic Acids Res. **10**:5183–5196.
8. **Dhar, R., W. L. McClements, L. W. Enquist, and G. F. Vande Woude.** 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. U.S.A. **77**:3937–3941.
9. **Dickson, C., R. Eisenman, H. Fan, E. Hunter, and N. Teich.** 1982. Protein biosynthesis and assembly, p. 513–648. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses: molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Edwards, S. A., and H. Fan.** 1979. *gag*-related polyproteins of Moloney murine leukemia virus: evidence for independent synthesis of glycosylated and unglycosylated forms. J. Virol. **30**:551–563.
11. **Elder, J. H., and J. I. Mullins.** 1983. Nucleotide sequence of the envelope gene of Gardner-Arnstein feline leukemia virus B reveals unique sequence homologies with a murine mink cell

focus-forming virus. J. Virol. **46**:871–880.
12. **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.
13. **Essex, M.** 1980. Feline leukemia and sarcoma viruses, p. 205–229. *In* G. Klein (ed.), Viral oncology. Raven Press, New York.
14. **Gardner, M. B., R. W. Rongey, P. Arnstein, J. D. Estes, P. Sarma, R. J. Huebner, and C. G. Rickard.** 1970. Experimental transmission of feline fibrosarcoma to cats and dogs. Nature (London) **226**:807–809.
15. **Haenni, A. L., S. Joshi, and F. Chapeville.** 1982. tRNA-like structures in the genomes of RNA viruses. Prog. Nucleic Acid Res. Mol. Biol. **27**:85–104.
16. **Hampe, A., M. Gobet, J. Even, C. J. Sherr, and F. Galibert.** 1983. Nucleotide sequences of feline sarcoma virus long terminal repeats and 5′ leaders show extensive homology to those of other mammalian retroviruses. J. Virol. **45**:466–472.
17. **Hampe, A., I. Laprevotte, F. Galibert, L. A. Fedele, and C. J. Sherr.** 1982. Nucleotide sequences of feline retroviral oncogenes (*v-fes*) provide evidence for a family of tyrosine-specific protein kinase genes. Cell **30**:775–785.
18. **Henderson, L. E., T. D. Copeland, R. C. Sowder, G. W. Smythers, and S. Oroszlan.** 1981. Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses. J. Biol. Chem. **256**:8400–8406.
19. **Hérissé, J., G. Courtois, and F. Galibert.** 1980. Nucleotide sequence of the EcoRI D fragment of adenovirus 2 genome. Nucleic Acids Res. **8**:2173–2192.
20. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:3824–3828.
21. **Hughes, S.** 1982. Appendixes, p. 1285–1386. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses: molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. **Khan, A. S., and J. R. Stephenson.** 1977. Feline leukemia virus: biochemical and immunological characterization of *gag* gene-coded structural proteins. J. Virol. **23**:599–607.
23. **Knapp, G., R. C. Ogden, C. L. Peebles, and J. Abelson.** 1979. Splicing of yeast tRNA precursors: structure of the reaction intermediates. Cell **18**:37–45.
24. **Krebs, E. G., and J. A. Beavo.** 1979. Phosphorylation-dephosphorylation of enzymes. Annu. Rev. Biochem. **48**:923–959.
25. **Ledbetter, J., and R. C. Nowinski.** 1977. Identification of the Gross cell surface antigen associated with murine leukemia virus-infected cells. J. Virol. **23**:315–322.
26. **Lehninger, A. L.** 1970. Proteins: conformation, p. 109–127. *In* A. L. Lehninger (ed.), Biochemistry. Worth Publishers, Inc., New York.
27. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
28. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
29. **Mullins, J. I., J. W. Casey, M. O. Nicolson, K. Bauman Burck, and N. Davidson.** 1981. Sequence arrangement and biological activity of cloned feline leukemia virus proviruses from a virus-productive human cell line. J. Virol. **38**:688–703.
30. **Murphy, E. C., Jr., J. J. Kopchick, K. F. Watson, and R. B. Arlinghaus.** 1978. Cell-free synthesis of a precursor polyprotein containing both *gag* and *pol* gene products by Rauscher murine leukemia virus 35S RNA. Cell **13**:359–369.
31. **Murti, K. G., M. Bondurant, and A. Tereba.** 1981. Secondary structural features in the 70S RNAs of Moloney murine leukemia and Rous sarcoma viruses as observed by electron microscopy. J. Virol. **37**:411–419.
32. **Neil, J. C., J. E. Smart, M. J. Hayman, and O. Jarrett.** 1980. Polypeptides of feline leukemia virus: a glycosylated *gag*-related protein is released into culture fluids. Virology **105**:250–253.
33. **Oroszlan, S., and R. V. Gilden.** 1980. Primary structure analysis of retrovirus proteins, p. 299–344. *In* J. R. Stephenson (ed.), Molecular biology of RNA tumor viruses. Academic Press, Inc., New York.
34. **Pal, B. K., R. M. McAllister, M. B. Gardner, and P. Roy-Burman.** 1975. Comparative studies on the structural phosphoproteins of mammalian type C viruses. J. Virol. **16**:123–131.
35. **Philipson, L., P. Andersson, U. Olshevsky, R. Weinberg, and D. Baltimore.** 1978. Translation of MuLV and MSV RNAs in nuclease-treated reticulocyte extracts: enhancement of the *gag-pol* polypeptide with yeast suppressor tRNA. Cell **13**:189–199.
36. **Robbins, K. C., M. Barbacid, K. J. Porzig, and S. A. Aaronson.** 1979. Involvement of different exogenous feline leukemia virus subgroups in the generation of independent feline sarcoma virus isolates. Virology **97**:1–11.
37. **Salser, W.** 1978. Globin mRNA sequences: analysis of base pairing and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. **42**:985–1002.
38. **Sarma, P. S., and T. Log.** 1973. Subgroup classification of feline leukemia and sarcoma viruses by viral interference and neutralization tests. Virology **54**:160–169.
39. **Schultz, A. M., S. M. Lockhart, E. M. Rabin, and S. Oroszlan.** 1981. Structure of glycosylated and unglycosylated *gag* polyproteins of Rauscher murine leukemia virus: carbohydrate attachment sites. J. Virol. **38**:581–592.
40. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
41. **Segrest, J. P., and R. J. Feldmann.** 1974. Membrane proteins: amino acid sequence and membrane penetration. J. Mol. Biol. **87**:853–858.
42. **Seif, I., G. Khoury, and R. Dhar.** 1979. BKV splice sequences based on analysis of preferred donor and acceptor sites. Nucleic Acids Res. **6**:3387–3397.
43. **Sen, A., C. J. Sherr, and G. J. Todaro.** 1977. Phosphorylation of murine type C viral p12 proteins regulates their extent of binding to the homologous viral RNA. Cell **10**:489–496.
44. **Sharp, P. A.** 1981. Speculations on RNA splicing. Cell **23**:643–646.
45. **Sherr, C. J., L. Donner, L. A. Fedele, L. Turek, J. Even, and S. K. Ruscetti.** 1980. Molecular structure and products of feline sarcoma and leukemia viruses: relationship to FOCMA expression, p. 293–307. *In* W. D. Hardy, Jr., M. Essex, and A. J. McClelland (ed.), Feline leukemia virus. Elsevier/North-Holland Biomedical Press, Amsterdam.
46. **Sherr, C. J., L. A. Fedele, M. Oskarsson, J. Maizel, and G. Vande Woude.** 1980. Molecular cloning of Snyder-Theilen feline leukemia and sarcoma viruses: comparative studies of feline sarcoma virus with its natural helper virus and with Moloney murine sarcoma virus. J. Virol. **34**:200–212.
47. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**:543–548.
48. **Snyder, S. P., and G. H. Theilen.** 1969. Transmissible feline fibrosarcoma. Nature (London) **221**:1074–1075.
49. **Stephenson, J. R.** 1980. Type C virus structural and transformation-specific proteins, p. 245–297. *In* J. R. Stephenson (ed.), Molecular biology of RNA tumor viruses. Academic Press, Inc., New York.
50. **Stephenson, J. R., R. K. Reynolds, S. G. Devare, and F. H. Reynolds.** 1977. Biochemical and immunological properties of *gag* gene-coded structural proteins of endogenous type C RNA tumor viruses of diverse mammalian species. J. Biol. Chem. **252**:7818–7825.
51. **Temin, H. M.** 1981. Structure, variation and synthesis of retrovirus long terminal repeat. Cell **27**:1–3.
52. **Temin, H. M.** 1982. Function of the retrovirus long terminal repeat. Cell **28**:3–5.
53. **Temin, H. M., and D. Baltimore.** 1972. RNA-directed DNA synthesis and RNA tumor viruses. Adv. Virus Res. **17**:129–186.
54. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
55. **Tinoco, I., Jr., O. C. Uhlenbeck, and M. D. Levine.** 1971. Estimation of secondary structure in ribonucleic acids. Nature (London) **230**:362–367.

56. **Tung, J.-S., A. Pinter, and E. Fleissner.** 1977. Two species of type C viral core polyprotein on AKR mouse leukemia cells. J. Virol. **23**:430–435.
57. **Versteegen, R. J., T. D. Copeland, and S. Oroszlan.** 1982. Complete amino acid sequence of the group-specific antigen gene-encoded phosphorylated proteins of mouse leukemia viruses. J. Biol. Chem. **257**:3007–3013.
58. **Weiss, S. R., P. B. Hackett, H. Oppermann, A. Ullrich, L. Levintow, and J. M. Bishop.** 1978. Cell-free translation of avian sarcoma virus RNA: suppression of the *gag* termination codon does not augment synthesis of the joint *gag/pol* product. Cell **15**:607–614.