Haematology and Blood Transfusion Vol. 28
Modern Trends in Human Leukemia V
Edited by Neth, Gallo, Greaves, Moore, Winkler

The Transforming Gene of Avian Myeloblastosis Virus (AMV): Nucleotide Sequence Analysis and Identification of Its Translational Product

T. S. Papas, K. E. Rushlow, D. K. Watson, J. P. Bader, D. Ray, and E. P. Reddy

A. Abstract

The genome of the avian myeloblastosis virus (AMV) has undergone a sequence substitution in which a portion of the region normally coding for the *env* protein has been replaced by cellular sequences. We have determined the complete nucleotide sequence of this region. Examination of the AMV oncogenic sequence revealed an open reading frame starting with the initiation codon ATG and terminating with the triplet TAG within the acquired cellular sequences and terminating with the triplet TAG at a point thirty-three nucleotides into helper viral sequences to the right of the helper-viral-cellular junction. The stretch of 795 nucleotides would code for a protein of 265 amino acids with a molecular weight of 30,000 daltons. The eleven amino acids at the carboxy terminus of such a protein would be derived from the *env* gene of helper virus. Antibodies were prepared against synthetic peptides derived from the predicted amino acid sequences. One such antibody precipitated two magnesium proteins of apparent nucleotide weight of 30,000 daltons and 51,000 daltons.

B. Introduction

The mechanism by which acute transforming retroviruses have acquired their ability to transform cells is closely associated with their capacity for synthesis of double-stranded DNA copies of their RNA genome. The DNA proviral forms cannot only integrate into host chromosomes, but have the ability to acquire host genetic material by a process of recombination. The replacement of viral genes by cellular ones usually results in the virus becoming defective, requiring the presence of non-defective helper virus for the maintenance of the acquired genes [1]. On rare occasions, the captured sequence contains a crucial control gene which, when inserted by the virus into cells which do not normally express it, or into sites in the cell where its expression is not regulated, can result in malignant transformation of the cells. In this report we present the nucleotide sequence of the transforming gene of AMV and identification of its translational product. This unique virus causes acute myeloblastic leukemia with a short latent period in chickens [2]. In vitro, AMV transforms a specific class of hematopoietic cells, but does not morphologically transform fibroblasts [3, 4]. Only certain target cells are responsive to the AMV *onc* gene product which induces proliferation of immature and altered hematopoietic cells, i.e., leukemic myeloblasts.

C. Results

I. Restriction Enzyme Map and Strategy of Sequencing the Transforming Gene of AMV

The upper portion of Fig. 1 shows the genetic map and orientation of the AMV provirus in the avian chromosome. An important structural feature of the integrated AMV genome is the occurrence of two large terminal repeats (LTRs) of 285 bases each at both the 5′ and 3′ ends of the pro-

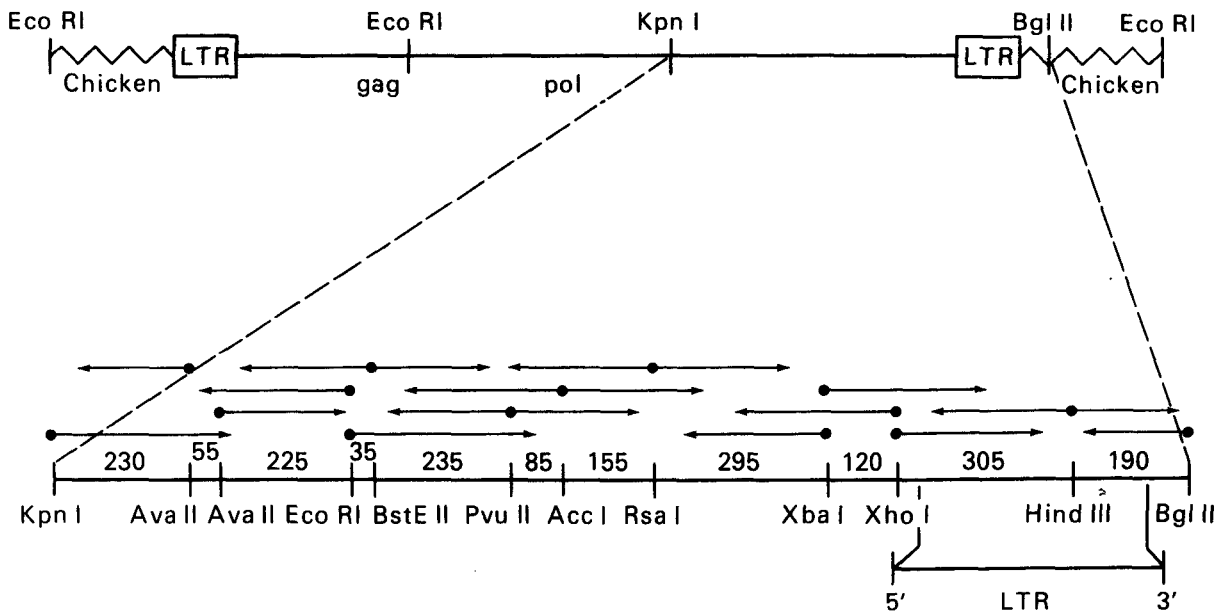

Fig. 1. Restriction enzyme map and strategy of sequencing the transforming gene of AMV. The genome is sequenced using the restriction sites indicated as the diagrammatic map. The 5′ ends were labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase. The labeled end of each fragment is indicated by the *filled circle* and the extent and direction of sequencing are indicated by *arrows*

viral sequence [6]. Within the proviral sequences, we can identify the *gag* gene proximal to the 5′ end of the viral RNA, followed by the polymerase gene and the oncogenic sequences. The oncogenic sequences extend to the 3′ end of the viral RNA and beyond the *Kpn* site of the cloned provirus (Fig. 1).

The lower portion of Fig. 1 provides a summary of the strategy employed to determine the nucleotide sequence. After digestion of DNA with appropriate restriction enzymes, the fragments to be sequenced were isolated on agarose gels or polyacrylamide gels and sequenced in either the 5′→3′ or 3′→5′ direction. The fragment is oriented from left to right, 5′ to 3′, with respect to the viral RNA. The *arrows* below the DNA strand indicate the length and direction of sequencing.

II. Nucleotide Sequence of the Transforming Gene of AMV

The nucleotide sequence of the 3′ end of the integrated AMV provirus is shown in Fig. 2. Within this sequence we can identify several domains: (1) the terminal portion of the polymerase gene, identified by an open reading frame extending from position 1 and terminating with a TAG codon at position 162; (2) a region of 350 bases without on apparent open reading frame extending between positions 165–515; (3) an open reading frame of 795 bases extending from positions 516–1310; and (4) the 3′-LTR adjacent to the host sequences.

Earlier studies have revealed that the AMV genome has undergone recombination in which the entire helper virus *env* gene has been replaced by cellular sequences [7, 8]. In order to localize the points of recombination, we have compared the carboxy terminal sequence of the AMV polymerase gene with that of the nondefective Prague strain of Rous sarcoma virus (PR-RSV) (Schwartz D, personal communication). From position 1 to position 78 of the AMV DNA fragment sequenced here, the nucleotide sequence is identical to that of PR-RSV. From position 78 to the termination signal, TAG at position 1313, the sequences of AMV and PR-RSV are entirely different, thus localizing the 5′ end of the cellular insertion sequences.

It is interesting to note that the host-helper virus junction occurs at a region which constitutes a potential splice acceptor site. In general, splicing acceptor sites (at the 3′ end of the intervening sequence) contain a pyrimidine-rich nucleo-

tide tract followed by the sequence AGG. The junction point between the cellular insertion sequence and the helper viral sequence fits the consensus acceptor splice sequence [9].

The product of the AMV transforming gene has yet to be identified. Examination of the cellular-derived *amv* sequences (Fig. 2) reveals an open reading frame starting with the initiation codon ATG at position 516 and terminating with the triplet TAG at position 1310. This stretch of 795 nucleotides could code for a protein of 265 amino acids with a molecular weight of 30,000. The amino acid sequence predicted from this region is also shown in Fig. 2.

Ribonuclease T_1-resistant oligonucleotide analysis was carried out by Duesberg et. al [8] on RNA isolated from the defective AMV particles. The RNA from these particles contains 14 unique T_1-oligonucleotides which are unrelated to sequences present in nondefective avian retroviruses and to the transformation-specific sequences of other avian leukemia and sarcoma viruses. Duesberg et al. suggested that these RNA sequences belong to the leukemogenic region of the AMV geneome [8]. We have utilized the computer program devised by Queen [10] and have positively identified 13 of the 14 oligonucleotides in the sequence presented in Fig. 2. Table 1

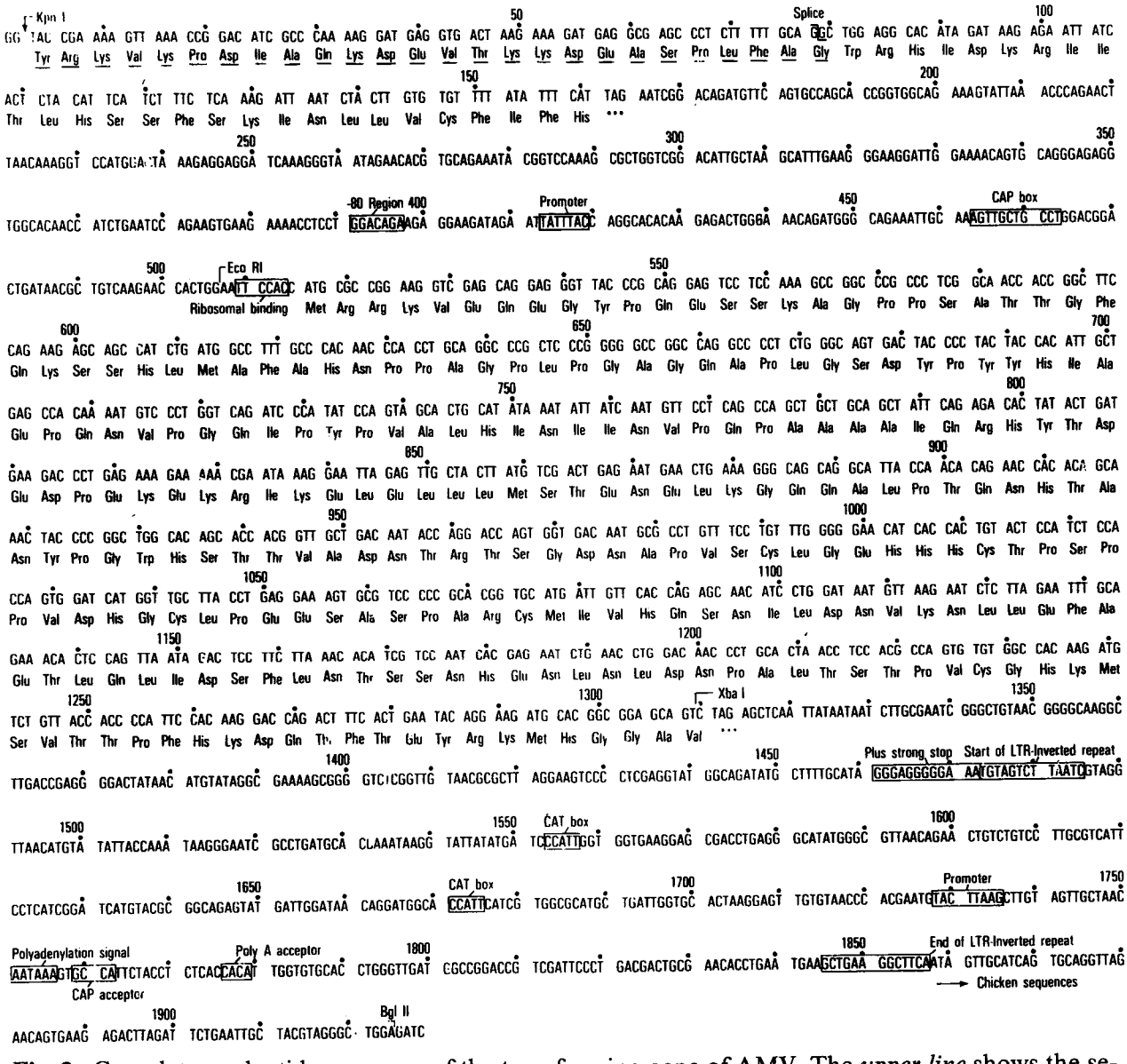

Fig. 2. Complete nucleotide sequence of the transforming gene of AMV. The *upper line* shows the sequence proceeding in the 5′ to 3′ direction and has the same polarity as AMV genomic RNA. The amino acid sequence deduced from the open reading frame is given in the *bottom line.* The major structural features of the genome are indicated

Table 1. The AMV-specific T_1 oligonucleotides and their position in the sequence: 101 (etc.) indicates the number for the nucleotide chromatographic patterns in [10]

T_1 Oligo	Sequence	Position in the sequence	Specificity
101	ATTAATCTACTTG	132 – 144	AMV specific
102	AATTATCACTCTACATTCATCTTTCTCAAAG	101 – 131	AMV specific
103	CACTAACCTCCACG	1207 – 1220	AMV specific
104	AATTATTTACCAG	410 – 422	AMV specific
105	TTTTATATTTCATTAG	149 – 164	AMV specific
106	ACTACCCCTACTACCACATTG	679 – 699	AMV specific
107	CCCACAACCCACCTG	622 – 636	AMV specific
108	CATATAAATATTATCAATG	747 – 766	AMV specific
113	CATTACCAACACAG	892 – 905	AMV specific
110	CAAACTACCCCG	916 – 927	AMV specific
111	ACTCCTTCTTAAACATCG	1153 – 1172	AMV specific
112	TACTCCATCTCCACCAG	1013 – 1029	AMV specific
114	TTACCACCCCATTCCACAAG	1246 – 1265	AMV specific
51	CTCAATTATAATAATCTTG	1316 – 1334	Common C-region
52	TATATTACCAAATAAG	1499 – 1514	LTR (U_3)
53	CACCAAATAAG	1529 – 1539	LTR (U_3)
54	CTAACAATAAAG	1746 – 1757	LTR (U_3)
55	?	?	
56	TCATTCCTCATCG	1616 – 1628	LTR (U_3)
57	CACCATTCATCG	1669 – 1688	LTR (U_3)
1	CCATTCTACCTCTCACCACATTG	1760 – 1782	LTR (U_3)

lists the T_1-oligonucleotides identified by Duesberg [8] and their position within our sequence. Comparison of our sequence with that of the RSV envelope region (Schwartz D, personal communication) reveals that the last 11 amino acids at the carboxy terminus are shared by the two proteins, suggesting that the *amv* gene is incomplete and utilizes the envelope terminator codon. This positions the 3′ terminus of the recombination event at position 1277.

III. Identification of the Transforming Gene (*amv*) Translational Product

Nucleotide sequence analysis of *amv* has revealted the presence of a 795-base open reading frame commencing within the acquired cellular sequences and terminating within the helper viral sequences (Fig. 2). Synthetic peptides prepared on the basis of predicted amino acid sequences of various genes have been utilized in the recent past to prepare antibodies against such proteins. Such antibodies provide a powerful tool for identificaiton and characterization of proteins that could not be previously identified. If the *amv* reading frame were functional, antibodies prepared against the synthetic peptides predicted from this sequence should be capable of precipitating the translational product of this viral *onc* gene. Such an approach not only allows the detection of transforming proteins but also provides additional experimental evidence for the correctness of the open reading frame derived by nucleotide sequence analysis. For this, we chemically synthesized three peptides, each 15 amino acids long. The peptides were coupled to thyroglobulin using 2-ethyl carbondiinide [21] and used for immunization of rabbits.

In an attempt to identify the *amv* translational product, the myeloblasts nonproductively infected with AMV were labeled for 6 h with [^{35}S]-methionine and the cell lysates prepared by detergent lysis. The cytoplasmic extracts were then immunoprecipitated with preimmune sera or anti *amv*-1, anti *amv*-2, and anti *amv*-3. As

negative controls, we used uninfected cells derived from chicken embryo fibroblasts. As shown in Fig. 3, two proteins with an apparent molecular weight of 30,000 daltons and 57,000 daltons were precipitated with anti *amv*-2. These proteins were not precipitable with preimmune sera nor were they detectable by any of the sera used in control cells that were not infected by AMV. Also, these two proteins were not precipitable from cells infected with helper virus alone. These observations strongly suggest that p30 and p57 are encoded by the transforming region of *amv*.

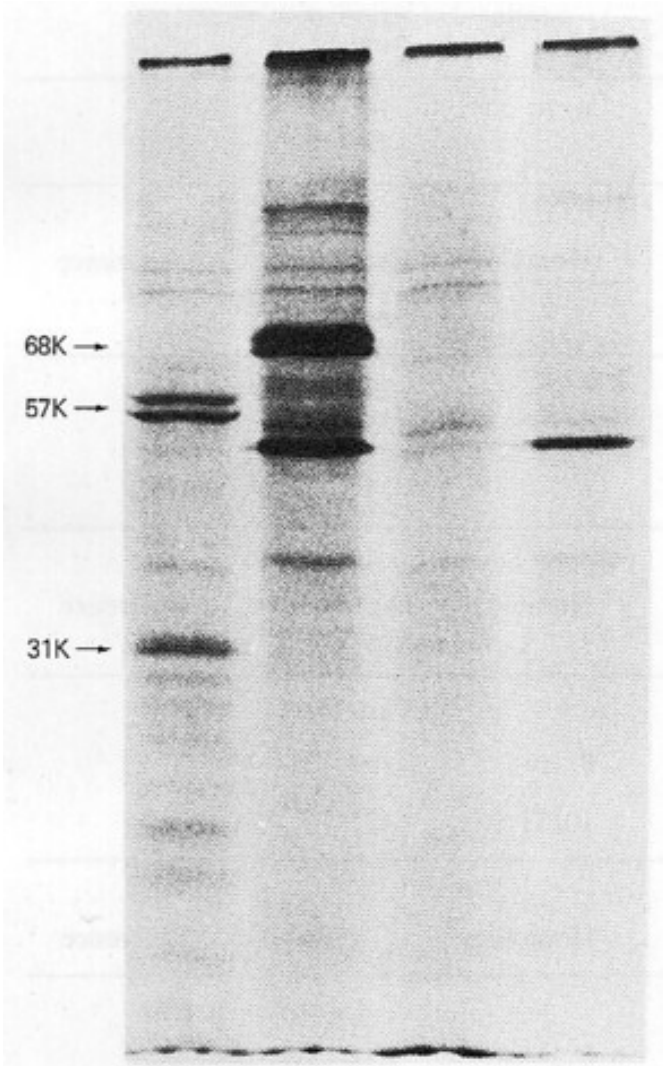

Fig. 3. SDS-PAGE analysis of immunoprecipitated cell lysates. Panel numbering is from *left* to *right*. *Panel 1,* AMV-transformed nonproducer cells + anti *amv*-2. *Panel 2,* chicken embryo fibroblasts + anti *amv*-2. *Panel 3,* AMV-transformed nonproducer cells + prebleed of anti *amv*-2. *Panel 4,* chicken embryo firbroblasts + prebleed anti-*amv*-2. Pr68 in *panels 2 and 4* has been identified to be actin

D. Discussion

A message generated from the AMV-transforming region should direct the synthesis of the transforming protein with the predicted amino acid sequence shown in Fig. 2. This messenger RNA could be generated either by splicing with the leader sequence derived from the 5′ terminus of genomic RNA or by splicing with the leader sequence derived from the 5′ terminus of genomic RNA or by independent promotion.

There are at least four transcription and translation regulatory sequences representing a minimum of 31 properly arranged nucleotides within a region 124 nucleotides immediately upstream from the putative leukogenic sequence. The arrangement of these nucleotides cannot be due to chance and indicates that the *amv* insert was probably not acquired by recombination between viral DNA and a cDNA copy of cellular mRNA transcribed from the *c-amv* sequences. The creation of the AMV genome may be explained by a deletion-recombination mechanism first postulated for the formation of the Abelson virus genome by David Baltimore (personal communication). According to the model, a MAV provirus with a large deletion in the 3′ half of the viral genome starting at position 78 beyond the *Kpn*I site (Fig. 2) would have been integrated in the vicinity of the chicken *amv* sequences. Transcriptions initiated by the MAV 5′-LTR generated a hybrid MAV-chicken RNA extending from the remainder of the viral genome to the 3′ terminus of the cellular c-*amv* sequences. Subsequent splicing within the *c-amv* sequences which contain three introns not present within AMV then generated a hybrid viral chicken mRNA terminating at the *myb* 3′ terminus. This hybrid mRNA could have been packaged in a MAV virion and subsequently copied in cDNA by the viral reverse transcriptase. This was followed by recombination with MAV DNA to create AMV by addition of the MAV 3′ end sequences starting at position 1316 beyond the *Kpn*I site. Splicing is generally used in the synthesis of viral subgenomic message. Leader sequences identified in MC29 [11] and RSV (Schwartz D, personal communication) cloned proviruses contain the 5′-

LRT, a noncoding region and 18 nucleotides coding for six amino acids of the N-terminal portion of the viral protein p19 ([11]; Schwarz D, personal communication). The splice donor portion of these sequences agrees with the consensus splice sequence of eukaryotic genes [9].

The alternate model for controlling the expression of the transforming gene would utilize the transcriptional signals found within the cellular insertion sequences in the region which lies between the polymerase gene and the open reading frame (Fig. 2). This type of independent promotion would not utilize the transcription controls of the viral 5′-LTR. Within the 350 base pair region in front of the putative leukemogenic sequence we have identified transcriptional signals similar to those present in other eukaryotic genes [12–20]. A six-base AT-rich sequence characteristic of eukaryotic promotors was identified at position 413–417, −56 bp from the capping site. Similarly, signals such a −80 bp region

Table 2. Landmarks of AMV as suggested by DNA sequences

−80 region				
Sequence	Source	Homology	Distance from $\overset{-1}{A}C\rightarrow$	Reference
GGACAGA	AMV	6/7	−79	
GGACAAA	Conalbumin		−78	[12]
Promoter region				
Sequence	Source	Homology	Distance from $\overset{-1}{A}C\rightarrow$	Reference
TATAAAT	General		−20 to −30	[13]
TATTTAC	AMV	4/7	−56	
TATTTAT	Ad 2 early	5/7	−23	[14]
TATATAT	Ovalbumin	6/7	−24	[15]
"CAP" box (AC→)				
Sequence	Source	Homology	Distance from ATG	Reference
GTTGCTCCT . . . . . $\overset{+1}{A}C$	General		Variable	[16]
AGTTGCTGCCT . . $\overset{+1}{A}C$	AMV	9/10	−39	
AGTTGCT · CCT . . $\overset{+1}{A}C$	β-globin^{maj}	10/11	−43	[17]
Initiator ATG region				
Sequence	Source	Homology		Reference
C/AAAPy*ATG*	General			[18]
C AC C*ATG*	AMV	7/7		
C AA C*ATG*	Conalbumin	6/7		[12]
A AC C*ATG*	Mouse α-globin	6/7		[19]
Ribosome binding				
Sequence	Source	Homology	Distance from ATG	Reference
TTCCGC	General		Variable	[20]
TTCCAC	AMV	5/6	−7	

and ribosomal binding sites have also been identified within this region. If these signals were to be utilized for the transcription of the *v-amv* gene, this would be the first example of a case in which the virus has incorporated the cellular regulatory signals for the transcription of its *onc* gene.

It would be very difficult at this point to conclude that these proteins are different or related to each other. It is possible that the p57 is a modified version of p30 (glycosylation, phosphorylation, etc.). Alternatively, it is possible that the two different proteins are generated from the same reading frame by two different mechanisms. The p30 protein could be the translational product of a mRNA derived by independent promotional signals (Table 2) identified in the transforming region. The p57 protein could be derived from a spliced mRNA generated from leader sequences provided by the helper virus and spliced to a region of the *amv* sequences 438 bases upstream from the ATG of the open reading frame. If this latter possiblility exists, a suppressor tRNA should be available in order to suppress translational terminator signal (TAA) at position 308 (Fig. 2). Alternatively, the p57 could be a translational product of *c-amv* encoded mRNA which would be expected to be much larger at the carboxy terminus.

References

1. Bishop JM (1978) Annu Rev Biochem 47:35
2. Duesberg PM (1979) Cold Spring Harbor Symp on Quant Biol 44:13
3. Baluda MA, Goeth IE (1961) Virology 15:185
4. Moscovice C (1975) Curr Top Microbiol Immunol 71:79
5. Souza LM, Komaromy MC, Baluda MA (1980) Proc Natl Acad Sci USA 77:3004
6. Rushlow KE, Lautenberger JA, Reddy EP, Souza LM, Baluda MA, Chirikjian JG, Papas TS (1982) J Virol 42:840
7. Souza LM, Baluda MA (1980) J Virol 36:317
8. Duesberg PM, Bister K, Moscovici C (1980) Proc Natl Acad Sci USA 77:5120
9. Sharp PA (1981) Cell 23:643
10. Queen CG, Korn LJ (1980) Methods in Enzymol 65:595
11. Papas TS, Schulz RA, Lautenberger JA, Pry TW, Samuel KP, Chirikjian JG (1981) In: Yohn DS, Blakeslee JR (eds) Advances in Comparative Leukemia Research. Elsevier/North Holland, Amsterdam pp 405–407
12. Cochet M, Gammon F, Hen R, Maroteaux L, Perrin F, Chambon P (1979) Nature 282:428
13. Gammon F, O'Hare K, Perrin F, LePennec JP, Benoist C, Cochet M, Breathnach R, Royal A, Garapin A, Cami B, Chambon P (1979) Nature 278:428
14. Maat J, van Ormondt H (1979) Gene 6:75
15. Benoist C, O'Hare K, Breathnach R, Chambon P (1980) Nucleic Acids Res 8:127
16. Gene Expression, Vol 2: Eucaryotic Chromosomes. Lewin, 822
17. Konkel D, Tilghman S, Leder P (1978) Cell 15:1125
18. Busslinger M, Portmann R, Irminger JC, Birnstiel ML (1980) Nucleic Acids Res 8:957
19. Nishioka Y, Leder P (1979) Cell 18:875
20. Maat J, van Beveren CP, van Ormondt H (1980) Gene 10:27
21. Kagan A, Glick M (1979) In: Jaffe BM, Behrman HR (eds) Methods of Hormone Radioimmunoassay. Academic Press, New York pp 328–329