Journal of Cellular Biochemistry 20:95–103 (1982)
Tumor Viruses and Differentiation 155–163

Nucleotide Sequence and Organization of the Transforming Region and Large Terminal Redundancies (LTR) of Avian Myeloblastosis Virus (AMV)

Takis S. Papas, Keith E. Rushlow, Dennis K. Watson, James A. Lautenberger, Bernard Perbal, Marcel E. Baluda, and E. Premkumar Reddy

Laboratory of Molecular Oncology (T.S.P., K.E.R., D.K.W., J.A.L.), Laboratory of Cellular and Molecular Biology (E.P.R.), National Cancer Institute, Bethesda, Maryland 20205, and UCLA School of Medicine and Johnson Comprehensive Cancer Center, Los Angeles, California 90024 (B.P., M.E.B.)

Avian myeloblastosis virus (AMV) is a replication-defective acute leukemia virus, requiring a helper virus to provide the viral proteins essential for synthesis of new infectious virus. The genome of the AMV has undergone a sequence substitution in which a portion of the region normally coding for the "*env*" protein has been replaced by chicken cellular sequences. These latter sequences are essential for the transforming activity of the virus. 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 within the acquired cellular sequences and terminating with the triplet TAG at a point 33 nucleotides into helper viral sequences to the right of 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.

Key words: acute leukemia virus, transforming gene, DNA sequencing, LTRs, nucleotides

Avian myeloblastosis virus is a replication-defective retrovirus which causes acute myeloblastic leukemia in chickens [1]. In vitro, it transforms a specific class of hematopoietic cells, but does not morphologically transform fibroblasts [2,3]. Thus, it appears that only certain target cells are responsive to the AMV *onc* gene product [1–4]. The genomic structure of AMV and its helper-associated virus (MAV) has been determined by molecular cloning of both viruses followed by restriction enzyme analysis and heteroduplex analysis [5–7]. These studies and earlier studies by Dues-

Received May 19, 1982; revised and accepted August 27, 1982.

0730-2312/82/2002-0095 $03.00 © 1982 Alan R. Liss, Inc.

berg et al [8], who carried out the ribonuclease T_1-resistant oligonucleotide analysis, revealed that both genomes share about a 5-kilobase (kb) region of homology from the 5′-end of the genomes, and the AMV genome has undergone recombination in which the entire helper virus "*env*" gene has been replaced by cellular sequences which appear to code for the transforming function associated with the virus. In an attempt to better understand the structural organization and possible molecular mechanisms involved in transformation by AMV, we have undertaken the sequence analysis of the transforming region of the molecule. Putative regulatory signals for transcription, RNA processing, and translation of the transforming gene (*amv*) have been identified. We have also sequenced the long terminal repeats (LTRs) of the viral DNA molecule and identified the regulatory signals present within this structure.

MATERIALS AND METHODS

Cloning of AMV

The AMV provirus was isolated from a library of chicken recombinant λ-phage constructed with a partial Eco RI digest of DNA from leukemic myeloblasts producing AMV [6]. The clone (λ11A-1) contains the entire AMV provirus and flanking chicken DNA sequences [6]. Portions of the proviral genome were subcloned in the plasmid vector, pBR325, and used for sequencing. One subclone contains AMV sequences located between the Kpn I site and the 3′-proximal Xba I site. The other sublcone contains AMV sequences located between the 3′-proximal Eco RI site and the 3′-viral terminus as well as the adjoining cellular DNA up to an Eco RI site.

DNA Isolation

DNA fragments containing the 3′-terminal sequences of AMV proviral DNA from the λ11A-1 clone were subcloned into the Eco RI site of pBR325. The fragments were isolated from Eco RI digests of the recombinant plasmids, and purified from preparative agarose gels.

End Labelling of DNA

DNA fragments used in restriction endonuclease mapping or chemical sequencing were labeled at their 5′-termini using [γ-^{32}P]ATP (Amersham, 2,000 Ci/mmol) and T4 polynucleotide kinase (P-L Biochemicals; cat no 0734). Prior to labeling, the 5′-ends were dephosphorylated with bacterial alkaline phosphatose (P-L Biochemicals; cat. no. 0976). Treatment of the DNA with phosphatase and labeling of the 5′-termini were carried out according to the procedure of Maxam and Gilbert [9].

Restriction Endonuclease Mapping

Restriction endonucleases were purchased from New England Biolabs. Fragments labeled with [^{32}P] at their 5′-ends were digested with appropriate restriction enzymes. The fragments were separated by gel electrophoresis on 5% (wt/vol) acrylamide, located by autoradiography, and purified from the gel using previously described methods [9]. The individual DNA fragments were mapped by the method of Smith and Birnstiel [10].

DNA Sequencing

Sequence determination of DNA fragments labeled with [^{32}P] at a unique 5′-terminus was carried out by the chemical modification method of Maxam and Gilbert

[9]. The five base-specific reactions utilized were G, G+A, A>C, C+T, and C. Bases 1–150 from the 5′-labeled end were identified by fractionation of the chemical cleavage products on 10–20% acrylamide-8 M urea gels. For sequence information beyond 150 bases, the products were fractionated on sequencing gels 80 cm in length and containing 6% acrylamide.

RESULTS

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

The upper portion of Figure 1 shows the restriction map and orientation of the AMV proviral DNA. 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 proviral genome [11]. Within the proviral sequences, we can identify the "*gag*" gene proximal to the 5′-end of the viral DNA, followed by the polymerase gene and the oncogenic sequences. The oncogenic sequences extend to the 3′-end of the viral DNA and beyond the Kpn site of the cloned provirus (Fig. 1).

The lower portion of Figure 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 indicate the length and direction of sequencing.

Nucleotide Sequence of the Transforming Gene of AMV

The nucleotide sequence of the 3′-end of the integrated AMV provirus is shown in Figure 2. Within this sequence we can identify several domains: (1) the terminal

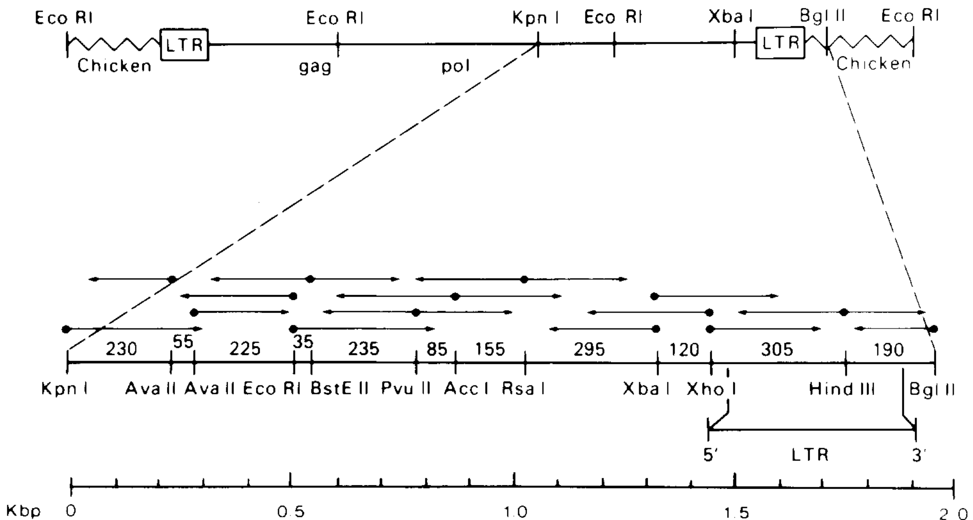

Fig. 1. Restriction enzyme map and strategy of sequencing the transforming gene of AMV. The genome is sequenced using the restriction sites indicated on the diagrammatic map. The 5′-ends were labeled using [γ-^{32}P] ATP and T4 polynucleotide kinase. The labeled end of each fragment is indicated by the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

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 an apparent open reading frame extending between positions 165–515; (3) an open reading frame of 795 bases extending from positions 516–1,310; 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 [5–8]. In order to localize the points of recombination, we have compared the carboxy-terminal sequence of the AMV polymerase gene with that of the myeloblastosis-associated virus (MAV) (our unpublished results) and the nondefective Prague strain of Rous sarcoma virus (PR-RSV) [12]. 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 position 1,277, the sequences of AMV and PR-RSV are entirely different, thus localizing the 5′-end of cellular insertion sequences.

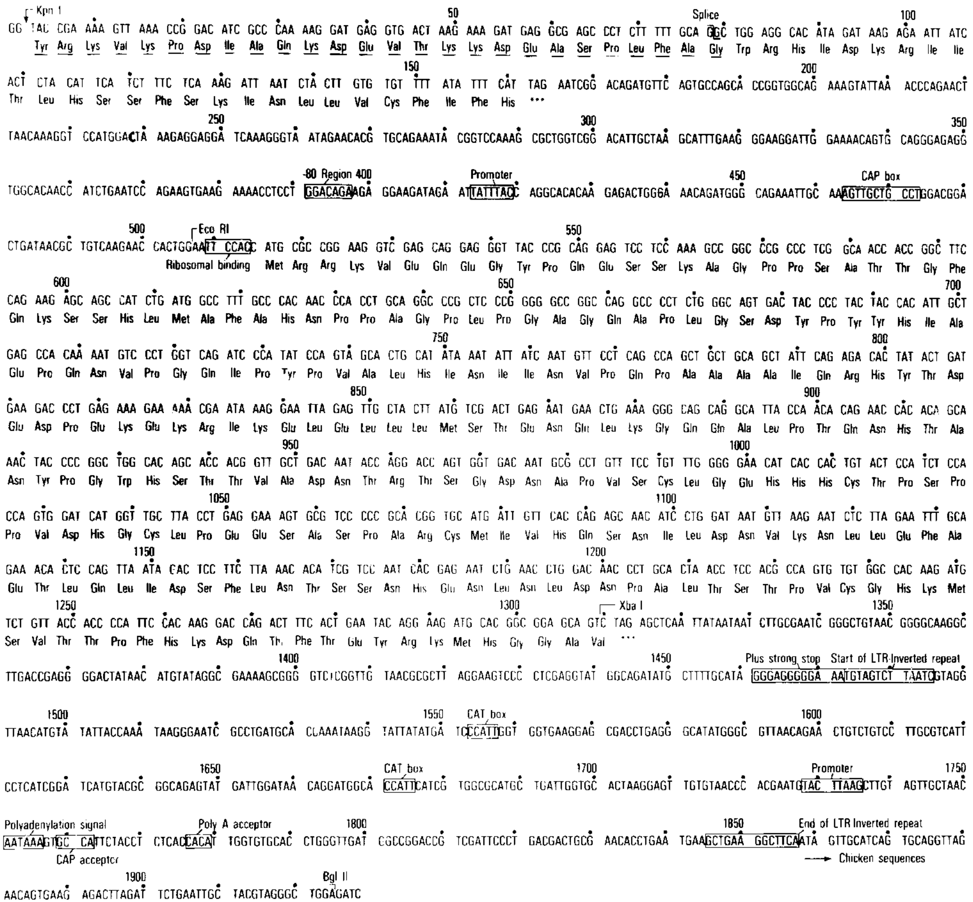

Fig. 2. Complete nucleotide sequence of the transforming gene of AMV. The upper line shows the sequence proceeding in the 5′ to 3′ direction and has the same polarity as AMV genomic RNA. The amino acid sequence deduced from the open reading frame is given in the bottom line. The major structural features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

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 intervening sequence) contain a pyrimidine-rich nucleotide 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 [13].

The product of the AMV transforming gene has yet to be identified. In vitro translation of AMV RNA revealed the occurrence of three proteins of molecular weights 56K, 49K, and 32K daltons, one of which could constitute the putative transforming proteins [14]. 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 1,310. 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 Figure 2. Upstream to the first ATG, the open reading frame extended for an additional stretch of 306 bases but without an initiator codon. The putative transforming protein shown in Figure 2 would contain five methionine residues in addition to starting N-formyl methionine, and cyanogenbromide cleavage would produce six peptides.

RNase-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 genome [8]. We have utilized the computer program devised by Queen and Korn [15] and have positively identified 13 of the 14 oligonucleotides in the sequence presented in Figure 2. Table I lists the T_1-oligonucleotides identified by Duesberg [8] and their position within our sequence. All but one of the AMV transformation-specific oligonucleotides fall within the cellular sequence inserted in AMV. Also, all but one of the helper-related oligonucleotides are found beyond the terminator signal TAG of the large open reading frame in the region shared with the MAV helper. T_1-oligonucleotide 51 located between position 1,316 and 1,334 is present in both AMV and its helper virus. Comparison of our sequence with those of MAV and RSV envelope region [12] reveals that 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 1,277.

Putative promoter and splicing signals in the transforming region. Analysis of viral DNA sequences to the right of Kpn I site revealed the occurrence of a putative acceptor splice point at position 78. In general, splicing-acceptor sites (at the 3′-end of intervening sequences) contain a pyrimidine-rich nucleotide track followed by the nucleotide AGG. Indeed, the sequence AGG at positions 78–81 was preceded by a track of ten pyrimidines. It is interesting to note this putative splice-acceptor site also constitutes the point of recombination between helper viral sequences and the call-derived *amv* sequences.

In addition to the splice-acceptor signal, the cell-derived *amv* sequences revealed the occurrence of an AT-rich transcription initiation signal at positions 413–419. This promoterlike signal present in AMV is identical to that reported for conalbumin (Table II). In addition, other putative translation control signals such as the mRNA

capsite (position 403), the −80 region (positions 391–397), and the ribosome-binding sites (positions 509–514) could be identified within the *amv* genome. The similarities between the putative regulatory signals found in the *amv* oncogene and those known to be present in other eukaryotic genes are summarized in Table II.

Sequence of the LTR

One of the LTRs of AMV extended 385 bases between positions 1473 and 1857. We identified the extent of the repeat by additional sequence determination of the second LTR at the 5′-end of the proviral genome. Examination of the LTR sequence revealed a number of latent features.

Inverted terminal repeats of 13 nucleotides. The sequence TGTAGTCT-TAATC appears at the termini of the LTR at positions 1473–1485 and 1845–1857.

Direct repeats. The sequence ACCAAATAAGG occurs both at positions 1505–1515 and positions 1530–1540. The function of these direct duplications remains to be determined.

Transcription initiation signal. A promoterlike sequence [17–19], TACT-TAAG, was found at positions 1728–1735. This AT-rich sequence precedes by 23 nucleotides the GCCA sequence most likely to be the RNA capping site. Moreover, the sequence CCATT was detected at a position 88 bases upstream from the 5′-cap structure. An analogous sequence has been shown to occur 77 ± 10 base pairs

TABLE I. The AMV-specific T_1 Oligonucleotides and Their Position in the Sequence

T_1 Oligo	Sequence	Position in the sequence	Specificity
101[a]	ATTAATCTACTTG	132–144	AMV specific
102	AATTATCACTCTACATTCATCTTTCTCAAAG	101–131	AMV specific
103	CACTAACCTCCACG	1,207–1,220	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	ACTCCTTCTTAAACACATCG	1,153–1,172	AMV specific
112	TACTCCATCTCCACCAG	1,013–1,029	AMV specific
114	TTACCACCCCATTCCACAAG	1,246–1,265	AMV specific
51	CTCAATTATAATAATCTTG	1,316–1,334	Common C-region
52	TATATTACCAAATAAG	1,499–1,514	LTR (U_3)
53	CACCAAATAAG	1,529–1,539	LTR (U_3)
54	CTAACAATAAAG	1,746–1,757	LTR (U_3)
55	?	?	
56	TCATTCCTCATCG	1,616–1,628	LTR (U_3)
57	CACCATTCATCG	1,669–1,688	LTR (U_3)
1	CCATTCTACCTCTCACCACATTG	1,760–1,782	LTR (U_3)

[a]Indicates the number for the nucleotide chromotographic patterns in [8].

upstream from the 5′-end of the mRNA capping site of most eukaryotic structural genes [25].

Polyadenylation signal. The sequence AATAAA is usually found 10–30 nucleotides upstream from the dinucleotide CA, the preferred site for polyadenylation [26]. For the AMV LTR, the sequence AATAAA (1751–1756) occurs 20 nucleotides upstream of the CA dinucleotide.

The positioning of LTRs at both ends of the integrated provirus strongly suggests that the viral genomic RNA is initiated at the promoter of the left LTR and terminated at the polyadenylation signal of the right LTR. There is evidence from studies with avian retroviruses that transcripts can be initiated by "downstream" promotion of cellular sequences from right LTR [27]. Thus, it appears that promoters present in both LTRs could be functionally active in an integrated provirus.

DISCUSSION

Nucleotide sequence analysis of the transforming gene of avian myeloblastosis virus has revealed several important features of its molecular organization. By

TABLE II. Putative Regulatory Signals Present in the Cellular Insertion Sequence of AMV

−80 Region				
Sequence	Source	Homology	Distance from $\overset{-1}{\text{AC}}\rightarrow$	Reference
GGACAGA	AMV	6/7	−79	
GGACAAA	Conalbumin		−78	[16]
Promoter region				
Sequence	Source	Homology	Distance from $\overset{-1}{\text{AC}}\rightarrow$	Reference
TATAAAT	General		−20 to −30	[17]
TATTTAC	AMV	4/7	−56	
TATTTAT	Ad 2 early	5/7	−23	[18]
TATATAT	Ovalbumin	6/7	−24	[19]
"CAP" box (AC→)				
Sequence	Source	Homology	Distance from ATG	Reference
GTTGCTCCT • • • $\overset{+1}{\text{AC}}$	General		Variable	[20]
AGTTGCTGCCT • • $\overset{+1}{\text{AC}}$	AMV	9/10	−39	
AGTTGCT•CCT • • $\overset{+1}{\text{AC}}$	β-Globin^{maj}	10/11	−43	[21]
Initiator ATG region				
Sequence	Source	Homology		Reference
C/AAAPy*ATG*	General			[22]
C AC C*ATG*	AMV	7/7		
C AA C*ATG*	Conalbumin	6/7		[16]
A AC C*ATG*	Mouse α-globin	6/7		[23]
Ribosome binding				
Sequence	Source	Homology	Distance from ATG	Reference
TTCCGC	General		Variable	[24]
TTCCAC	AMV	5/6	−7	

comparison of the nucleotide sequence with information available from sequence studies of myeloblastosis-associated virus (MAV) and RSV, it was possible to localize the junction points between *amv* and helper viral sequences. The AMV transforming region revealed a single open reading frame commencing within the cell-derived *amv* sequences and terminating at a point 33 nucleotides into the helper viral sequences to the right of *amv*. Our findings imply that, if this open reading frame were the coding sequence for the AMV transforming protein, the last 36 bases including the terminator codon are contributed by the helper viral genome that constitutes the carboxy terminus of the "*env*" gene. This sequence could code for a protein of 265 amino acids with a molecular weight of approximately 30,000 daltons.

A messenger RNA generated from the AMV transforming region should direct the synthesis of the transforming protein with the predicted amino acid sequence shown in Figure 2. This messenger RNA could be generated either by splicing with the leader sequence derived from the 5′-terminus of genomic RNA or by independent promotion. Splicing is generally used in the synthesis of viral subgenomic messages. Leader sequences identified in MC29 [28] and RSV [12] cloned proviruses contain the 5′-LTR, a noncoding region, and 18 nucleotides coding for six amino acids of the N-terminal portion of the viral protein p19 [12,28]. The splice donor portion of these sequences agrees with the consensus splice sequence of eukaryotic genes [13]. Earlier studies on the analysis of subgenomic mRNAs present in AMV nonproducer cells have revealed the occurrence of such spliced mRNAs.

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 (Table I). This type of independent promotion would not utilize the transcription controls of the viral 5′-LTR. Within the 350 base pairs (bp) region in front of the putative leukemogenic sequence we have identified transcriptional signals similar to those present in other eukaryotic genes [16–24]. A six-base AT-rich sequence characteristic of eukaryotic promoters was identified at positions 413–417, −56 bp from capping site. Similarly, signals such as −80 bp region and ribosomal binding sites have also been identified within this region (Fig. 2). If these signals were to be utilized for the transcription of *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. The presence of these regulatory signals for transcription implies that the *amv* insert was generated by recombination directly with host DNA and not with a reverse transcript (complementary DNA) of the mRNA coded by chicken sequences homologous to *amv*.

This is the first report of the nucleotide sequence of a viral component responsible for acute myeloblastic leukemia. In this sequence there is an open reading frame encoding for a protein of approximately 30,000 daltons which has yet to be identified. Based on our predicted amino acid sequence, oligopeptides have been synthesized and are being used to produce antisera which may recognize the AMV transforming gene produced in leukemic cells.

REFERENCES

1. Duesberg PM: Cold Spring Harbor Symp Quant Biol 44:13, 1979.
2. Baluda MA, Goetz IE: Virology 15:185, 1961.
3. Moscovici C: Curr Top Microbiol Immunol 71:79, 1975.

4. Bishop JM: Annu Rev Biochem 47:35, 1978.
5. Souza LM, Baluda MA: J Virol 36:317, 1980.
6. Souza LM, Briskin MJ, Hillyard RL, Baluda MA: J Virol 36:325, 1980.
7. Souza LM, Komaromy MC, Baluda MA: Proc Natl Acad Sci USA 77:3004, 1980.
8. Duesberg PM, Bister K, Moscovici C: Proc Natl Acad Sci USA 77:5120, 1980.
9. Maxam AM, Gilbert W: Methods Enzymol 65:499, 1980.
10. Smith HO, Birnstiel ML: Nucleic Acids Res 3:2387, 1976.
11. Rushlow KE, Lautenberger JA, Reddy EP, Souza LM, Baluda MA, Chirikjian JG, Papas TS: J Virol 42:840, 1982.
12. Schwartz D: personal communication.
13. Sharp PA: Cell 23:643, 1981.
14. Andersen SM, Chen JH: J Virol 40:107, 1981.
15. Queen CG, Korn LJ: Methods Enzymol 65:595, 1980.
16. Cochet M, Gammon F, Hen R, Maroteaux L, Perrin F, Chambon P: Nature 282:428, 1979.
17. Gammon F, O'Hare K, Perrin F, LePennec JP, Benoist C, Cochet M, Breathnach R, Royal A, Garapin A, Cami B, Chambon P: Nature 278:428, 1979.
18. Maat J, van Ormondt H: Gene 6:75, 1979.
19. Benoist C, O'Hare K, Breathnach R, Chambon P: Nucleic Acids Res 8:127, 1980.
20. Lewin B: In Lewin B (ed): "Gene Expression," Vol. 2. New York: John Wiley and Sons, 1980, p822.
21. Konkel D, Tilghman S, Leder P: Cell 15:1125, 1978.
22. Busslinger M, Portmann R, Irminger JC, Birnstiel ML: Nucleic Acids Res 8:957, 1980.
23. Nishioka Y, Leder P: Cell 18:875, 1979.
24. Maat J, van Beveren CP, van Ormondt H: Gene 10:27, 1980.
25. Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, DeRiel JK, Forget BG, Weissmann SM, Slighton JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot NJ: Cell 21:653, 1980.
26. Proudfoot NJ, Brownlee GG: Nature 263:211, 1976.
27. Hayward WS, Neel BG, Astrin SM: Nature 290:475, 1981.
28. Papas TS, Schulz RA, Lautenberger JA, Pry TW, Samuel KP, Chirikjian JG: In Yohn DS, Blakeslee JR (eds): "Advances in Comparative Leukemia Research." Amsterdam: Elsevier/North Holland, 1981, pp 405-407.