Volume 10 Number 8 1982 Nucleic Acids Research

An adenovirus agnogene

Anders Virtanen, Peter Aleström, Håkan Persson[1], Michael G.Katze[2] and Ulf Pettersson[3]

Department of Microbiology, The Biomedical Center, University of Uppsala, Box 581, S-751 23 Uppsala, Sweden

Received 11 February 1982; Revised and Accepted 25 March 1982

ABSTRACT

The nucleotide sequence of a 550 base pairs long segment, located between map positions 21 and 22.5 in the adenovirus type 2 genome has been determined. S1 nuclease mapping and sequence analysis of cDNA copies of adenovirus mRNA demonstrated that the established sequence includes the i-leader which is spliced to the 5'-end of certain adenovirus mRNAs (1). The i-leader which is 440 nucleotides long contains an open translational reading frame which is preceded by an AUG triplet and which terminates in the third segment of the tripartite leader. A polypeptide with the same molecular weight as predicted from the DNA sequence was identified by in vitro translation of mRNA which had been selected by hybridization to DNA fragments, containing sequences from the i-leader. The results thus suggest that the i-leader, unlike other adenovirus leader segments, is used for translation.

INTRODUCTION

Most late adenovirus mRNAs have tripartite leaders covalently attached to their 5'-ends (2,3). The biogenesis of these mosaic RNAs starts with the synthesis of a very large transcript from the major late adenovirus promoter located 16.3 map units from the left hand end of the adenovirus genome. Subsequently three leader segments are covalently joined to the mRNA body by elimination of three intron sequences. Why the adenoviruses depend on such a complicated mechanism for generation of the 5'-end of their late mRNAs is unclear. We and others (4,5) have shown that the tripartite leader lacks AUG triplets. Hence initiation of translation must occur at an AUG triplet that follows the leader/mRNA-body junction. Sequences have been found in the tripartite leader which could base pair with the 3'-end of the 18S ribosomal RNA (4,6). Since Hagenbüchle et al. (7) have proposed that interactions between the 3' end of the 18S ribosomal RNA and the 5'-end of the mRNA are of importance for initiation of translation, it appears likely that the tripartite leader serves its function at initiation of translation, possibly by giving the adenovirus mRNAs an advantage when competing with cellular mRNAs for ribosome subunits and initiation factors. Besides the most

0305-1048/82/1008-2539$ 2.00/0

frequently occurring tripartite leader several alternatively spliced leaders have been recognized (8). These contain in addition to the three segments of the tripartite leader one or more auxilliary segments derived from other parts of the adenovirus genome. The fiber mRNA, in particular, can be spliced in several ways allowing at least five different combinations of sequences in its 5'-leader (8, Uhlén et al., submitted). Dunn et al. (9) have previously shown that the addition of one extra segment to the leader of the fiber mRNA does not seem to alter the coding properties of the mRNA.

A frequently occurring additional leader segment is the i-leader (1). When present, the i-leader is spliced between the second and the third segments of the normal tripartite leader. It is present in mRNAs both before and after the onset of viral DNA replication and can be attached to several different mRNA bodies.

In the present study we demonstrate that the i-leader contains an open translational reading frame encoding a hypothetical 15.9K polypeptide. Furthermore a 16K polypeptide was translated in vitro from mRNA, selected on DNA fragments which contain i-leader sequences.

MATERIALS AND METHODS

Viral DNA and sequencing: For most experiments a fragment of ad2 DNA, located between map positions 17.0 and 24.5 was used. This fragment was inserted into the HindIII and BamHI cleavage sites of the plasmid pBR322 and propagated in the E.coli strain HB101. For some experiments two cloned Taq I fragments from the same region of the adenovirus type 2 (ad2) genome were used. Plasmid DNA was prepared as described before (10).

The method of Maxam and Gilbert (11) was used for sequencing.

S1 nuclease analysis: The protocol of Favaloro et al. (12) was followed for S1 analysis, using end-labeled or internally labeled (13) fragments as probes. Sequencing ladders were separated on the same gel as the products of the S1 analysis (14).

End-labelling of fragments: The protocol of Maxam and Gilbert (11) was used for labeling of 5'-ends. Recessed 3'-ends were labeled by a so called fill-in reaction using the Klenow-fragment of DNA polymerase I and α-^{32}P-labeled dTTP.

Preparation of cDNA copies of mRNA: The procedure of Akusjärvi and Pettersson (15) was used with 5'-end labeled fragments as primers.

Hybridization selection of viral RNA and in vitro translation: Early and late cytoplasmic RNA was prepared from ad2 infected HeLa cells as previously

described (16). The RNA was hybridized to plasmid DNA immobilized on nitrocellulose filters as described by McGrogan et al. (17). The eluted RNA was translated in a rabbit reticulocyte cell-free system (18). The proteins were subsequently analyzed on 13% SDS-polyacrylamide gels (19) which were treated as described by Bonner and Laskey (20) for fluorography.

RESULTS

Sequence analysis of a segment of ad2 DNA which encodes the i-leader: Chow et al. (1) have shown that ad2 mRNAs often contain a supplementary leader segment, the i-leader, which has been mapped around position 22 in the ad2 genome. In order to study its structure we decided to determine the nucleotide sequence for a segment which is located around a cleavage site for endonuclease XhoI at map position 22. Akusjärvi and Persson (21) have previously shown that the XhoI cleavage site is located in the sequence for the i-leader. Fig. 1 outlines the sequencing strategy by which we established a 550 base pairs (bp) long sequence (Fig. 2).

The boundaries of the i-leader: In order to map the precise location of the sequences which encode the i-leader we used the modified S1 nuclease mapping procedure (12,14); 3'- and 5'-end labeled fragments from within the i-leader sequence were hybridized to late adenovirus mRNA and the resulting hybrids were trimmed with single strand specific nuclease. The products were subsequently analyzed on sequencing gels in parallel with a sequencing ladder obtained with the same end-labeled fragment as was used in the nuclease assay. To locate the 5'-border of the i-leader a small restriction fragment, 5'-labeled at the HincII cleavage site at map position 21.2, was used (Fig. 1) and in order to map the 3'-border, a fragment, 3'-labeled at

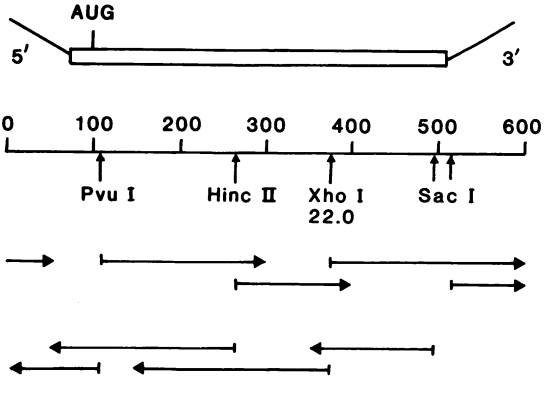

Figure 1 A schematic drawing which illustrates the strategy used to establish the sequence which is depicted in Fig. 2. Selected restriction enzyme cleavage sites are indicated.

```
1           13          25          37          49          61
CTTCATAACCAG CATGAAGGGCAC GAGCTGCTTCCC AAAGGCCCCCAT CCAAGTATAGGT CTCTACATCGTA

acceptor site (i-leader)
↓           85          97          109         121         133
GGTGACAAAGAG ACGCTCGGTGCG AGGATGCGAGCC GATCGGGAAGAA CTGGATCTCCCG CCACCAGTTGGA
                          MetArgAla AspArgGluGlu LeuAspLeuPro ProProValGly

145         157         169         181         193         205
GGAGTGGCTGTT GATGTGGTGAAA GTAGAAGTCCCT GCGACGGGCCGA ACACTCGTGCTG GCTTTTGTAAAA
GlyValAlaVal AspValValLys ValGluValPro AlaThrGlyArg ThrLeuValLeu AlaPheValLys

217         229         241         253         265         277
ACGTGCGCAGTA CTGGCAGCGGTG CACGGGCTGTAC ATCCTGCACGAG GTTGACCTGACG ACCGCGCACAAG
ThrCysAlaVal LeuAlaAlaVal HisGlyLeuTyr IleLeuHisGlu ValAspLeuThr ThrAlaHisLys

289         301         313         325         337         349
GAAGCAGAGTGG GAATTTGAGCCC CTCGCCTGGCGG GTTTGGCTGGTG GTCTTCTACTTC GGCTGCTTGTCC
GluAlaGluTrp GluPheGluPro LeuAlaTrpArg ValTrpLeuVal ValPheTyrPhe GlyCysLeuSer

361         373         385         397         409         421
TTGACCGTCTGG CTGCTCGAGGGG AGTTATGGTGGA TCGGACCACCAC GCCGCGCGAGCC CAAAGTCCAGAT
LeuThrValTrp LeuLeuGluGly SerTyrGlyGly SerAspHisHis AlaAlaArgAla GlnSerProAsp
             XhoI (22.0)
433         445         457         469         481         493
GTCCGCGCGCGG CGGTCGGAGCTT GATGACAACATC GCGCAGATGGGA GCTGTCCATGGT CTGGAGCTCCCG
ValArgAlaArg ArgSerGluLeu AspAspAsnIle AlaGlnMetGly AlaValHisGly LeuGluLeuPro
    donor site (i-leader)
505     ↓   517         529         541
CGGCGACAGGTC AGGCGGGAGCTC CTGCAGGTTTAC CTCGCATAGCCG.....................
ArgArgGln

    acceptor site (third segment of tripartite leader)
     ↓
GTGTAGGTACTC CGCCACCGAGGG ACCTGAGCGAGT.....................................
      ValLeu ArgHisArgGly Thr
```

Fig. 2. The established sequence which is located around map position 22 in the ad2 genome. The XhoI cleavage site at map position 22 is indicated as well as the donor- and acceptor sites which are used for splicing of the i-leader. The open reading frame is translated into a hypothetical amino acid sequence. The acceptor site which is used to splice the i-leader to the third segment of the tripartite leader is also shown.

the XhoI cleavage site at position 22.0 was used. By this approach it was only possible to locate the boundaries for the i-leader within 2-5 base pairs since the nuclease removes nucleotides from the ends of the hybrids. To locate the boundaries more precisely we searched in the established sequence for nucleotide combinations which are known to occur at donor and acceptor sites for splicing (22). At the 5'-boundary of the i-leader, the sequence 5'-TCGTAGGT-3' occurs. This sequence is a prototype sequence of an acceptor site for splicing and if we assume that the GT/AG rule (23) is followed the splice would occur between the two G's in the sequence (Fig. 2). The 3'-boundary of the i-leader was mapped at the sequence 5'-AGGTCAG-3' which is a characteristic donor site (22) and the splice would be predicted

Table 1. The predicted amino acid composition of the 15.9 K polypeptide[1)]

Lys	3	Glu	11	Ile	2
His	7	Gln	3	Leu	18
Arg	12	Pro	7	Tyr	3
Asp	8	Gly	11	Phe	4
Asn	1	Ala	16	Trp	4
Thr	7	Val	19	Cys	2
Ser	5	Met	2		

1) The initiator methionine and the seven amino acids which are encoded in the third leader segment are included.

to occur between the two G's (Fig. 2).

A second method was used to independently prove that the predicted positions were used for splicing of the i-leader (Fig. 3); two primers were prepared in order to make cDNA copies which cover the two boundaries of the i-leader. One primer, the XhoI-RsaI primer which is located in the third segment of the tripartite leader, was 5'-labeled at the XhoI-site and hybridized to late ad2 mRNA before extension with reverse transcriptase. Analysis of the extension products on a 6 % polyacrylamide gel revealed three bands: one corresponds to the unextended primer, one to a copy of the tripartite leader and one to a copy of a leader which includes the i-segment. Part of the sequence of the latter band was determined and the results confirmed the proposed location of the 3'-border. The second primer, the HincII/HhaI primer, which is located within the i-leader, was labelled at the HincII-end. The labelled primer was hybridized to late ad2 mRNA and extended by reverse transcriptase. Analysis of the products by gel electrophoresis revealed a 305 nucleotide long band. This is the expected size of a fragment that extends between the HincII site and the 5'-end of a leader which includes the i-segment. The band was eluted and part of its sequence was determined. The results confirmed the predicted location of the 5'-border.

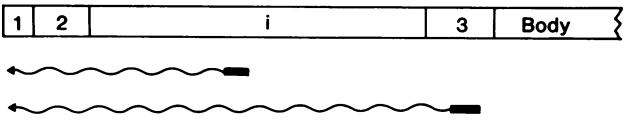

Fig. 3. A schematic drawing which illustrates how the boundaries of the i-leader were determined by extension of end-labeled primers.

In vitro translation of viral RNA, selected by hybridization to specific fragments:

To confirm that mRNAs with the i-leader code for a specific polypeptide, the following experiments were performed; cytoplasmic RNA from ad2 infected HeLa cells was selected by hybridization on cloned fragments of viral DNA. Both early RNA, prepared from cycloheximide treated HeLa cells, and late RNA, prepared 18 hours after infection were used. The selected RNA was translated in a reticulocyte cell free system and the results (Fig. 5) demonstrate that late RNA, selected on clones pT111 and pT3, both of which contain viral sequences from the i-leader region (Fig. 4), direct the synthesis of a polypeptide with an approximate molecular weight of 16K. RNA, obtained from ad2 infected HeLa cells, treated with cytosine arabinoside and selected by hybridization similarly programmed the synthesis of this 16K polypeptide (data not shown). In contrast early RNA prepared from cycloheximide treated cells failed to synthesize detectable levels of the 16K protein. As expected, cytoplasmic RNA selected on clone pT4 (Fig. 4) which exclusively contains intervening sequences between the i- and third leader segments, did not direct the sythesis of the 16K polypeptide (Fig. 5).

Detection of i-leader sequences by S1 analysis

The S1 nuclease mapping procedure of Berk and Sharp (24) was used to investigate the presence of i-leader in three different cytoplasmic RNA preparations: late RNA prepared 18 hours post infection, RNA from cytosine arabinoside treated cells extracted 16 hours post infection and early RNA prepared 8 hours post infection from cells treated with 25 µg/ml cyclo-

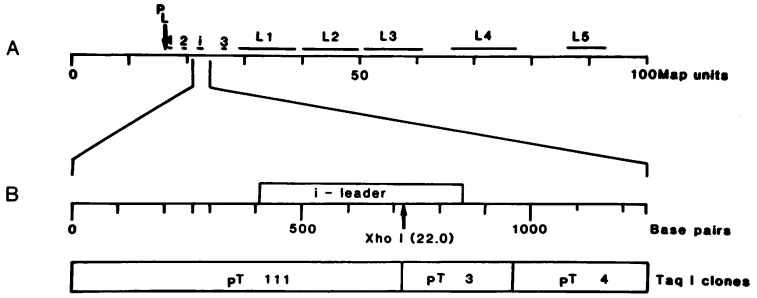

Fig. 4. A. A schematic map of the ad2 genome which shows the location of the major late promoter (P_L) the leader segments (1,2,i, and 3) and the five cotermination families of late mRNA (L1-L5).

B. An enlargement of the i-leader region showing the precise location of the i-leader as well as the cloned Taq-fragments which were used for hybridization selection.

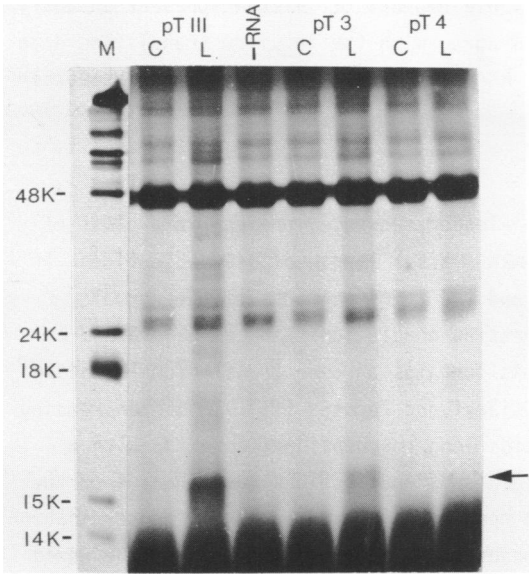

Fig. 5. In vitro translation of mRNA which had been selected on cloned restriction enzyme fragments. The location in the ad2 genome of the cloned fragments is shown in Fig. 4. Both late (L) and early mRNA (C), prepared from cells maintained in 25 µg/ml cycloheximide were used for selection. S^{35}-methionine labeled ad2 virions were used as size markers (M). The 16K polypeptide is indicated by an arrow.

heximide.

Cytoplasmic RNA was hybridized to ad2 DNA, the DNA-RNA hybrids were digested with S1 nuclease and the resulting DNA fragments were separated on a 1.4% alkaline agarose gel. The fragments were subsequently transferred to a nitrocellulose sheet and the i-leader sequences were visualized by hybridization with nick-translated ^{32}P-labeled clone pT 111 DNA (Fig. 4). From the autoradiogram it can be concluded that there are i-leader sequences in late RNA (Fig. 6 lane L), and RNA from cytosine arabinoside treated cells (Fig. 6

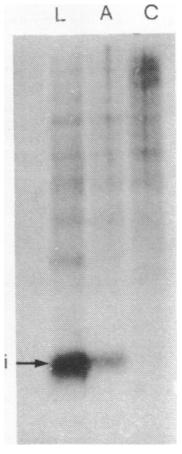

Fig. 6. S1 analysis of mRNA with the i-leader. RNA, prepared from cells either late (L) after ad2 infection or from cells maintained in cytosine arabinoside (A) or in cycloheximide (C) was hybridized to unlabeled ad2 DNA under the conditions of Berk and Sharp (24). The hybrids fragments were separated in an alkaline agarose gel followed by transfer to nitrocellulose (25). The i-leader was revealed by hybridization, using ^{32}P-labelled clone pTlll as a probe.

lane A) whereas there is no detectable amounts of i-leader present in early RNA (Fig. 6 lane C). These results agree with the *in vitro* translation data (Fig. 5) in showing that i-leader RNA is predominantly expressed late after adenovirus infection.

DISCUSSION

The present study shows that the i-leader sequence, present in a class of intermediate and late adenovirus mRNAs has a length of 440 nucleotides. It is spliced in between the second and the third segments of the normal tripartite leader. The most interesting conclusion that can be drawn from the established sequence is that it contains an open translational reading frame as recently proposed by Akusjärvi and Persson (21). This open reading frame could encode a 145 amino acids long polypeptide (Table I), with a potential AUG initiator triplet located 27 nucleotides downstream from the 5'-boundary of the i-leader. This open reading frame contains no termination codon within the i-leader which means that the predicted polypeptide would have to terminate at an UGA termination codon, located in the third leader segment (Fig. 2). All leader segments, which have been sequenced thus far and which are attached to late adenovirus mRNA, lack AUG triplets (4,5). The i-leader is the first known exception in the adenovirus system and raises the question as to whether the open reading frame is used for translation. In the present communication, we have used several cloned restriction enzyme fragments to show that only fragments which include i-leader sequences can select mRNA which is translatable into a 16K polypeptide (Fig. 5). The conclusion of these results is that mRNAs with i-leaders encode a 16K polypeptide. The 16K polypeptide is most likely encoded within the i-leader although the results do not exclude the remote possibility that it is translated from the body of a specific class of i-leader containing mRNAs. Amino acid radiosequence analysis is required to establish this point beyond doubt. The small amounts of 16K polypeptide available makes, however, this kind of analysis impossible.

Adenovirus mRNAs with the i-leader resemble in some respects the mRNA for the VP1 protein of SV40. The first initiator AUG which follows the capped 5' end of the VP1 mRNA is not used for initiation of VP1 synthesis. This mRNA is thus one of the few exceptions from the rule that the first AUG that follows the cap is used for initiation of translation (26). Recently results have, however, been provided which suggest that the first AUG in the VP1 mRNA is used for translation of a short polypeptide, the so called agno-

protein (27). The 16S mRNA for VP1 of SV40 is consequently a polycistronic mRNA, encoding both VP1 and the agnoprotein. The adenovirus i-leader could encode a protein with a similar function as the agnoprotein of SV40. The 16K polypeptide differs, however, from the agnoprotein of SV40 in several respects; it is considerably larger and less basic than the agnoprotein and no apparent sequence similarities can be detected between the amino acid sequences of the agnoprotein and the 16K polypeptide. The SV40 agnoprotein is moreover encoded by leaders attached to all late mRNAs whereas the 16K polypeptide is encoded by leaders present on a small fraction of intermediate and late adenovirus mRNA.

With regard to its regulation the i-leader differs from most other late adenovirus mRNAs. The data presented here suggest that the i-leader is made both in infected cells treated with cytosine arabinoside and in cells, harvested late after infection. In contrast cells which have been treated with cycloheximide early after infection contain reduced amounts of mRNA with i-leader sequences (Fig. 6). These results suggest that the enhanced synthesis of i-leader mRNA occurs independent of DNA replication, but may require the synthesis of specific proteins.

ACKNOWLEDGEMENTS

We are indebted to Monica Pettersson for competent technical assistance, to Marianne Gustafson and Ingegärd Schiller for excellent secreterial aid. This investigation was supported by grants from the Swedish Cancer Society, the Swedish MRC and the Swedish National Board for Technical Development. M.G.K. was supported by a long term fellowship from the European Molecular Biology Organization.

[1]Present address: Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, Bethesda, MD 20205, USA.

[2]Present address: Molecular Biology and Genetics Unit of the Graduate School, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA

[3]Present address: Department of Medical Genetics, Biomedical Center Box 589, S-751 23 Uppsala, Sweden.

NOTE ADDED IN PROOF

Lewis and Mathews (Cell, Volume 21, 303-313, 1980) have identified a 13.6 K polypeptide by in vitro translation which maps between coordinates 21.5 and 22.0. This polypeptide is most likely identical with the 16 K polypeptide described in the present studies.

REFERENCES

1. Chow, L.T., Broker, T. and Lewis, J.B. (1979). J.Mol. Biol. 134, 265-303.
2. Chow, L.T., Gelinas, R., Broker, T. and Roberts, R.J. (1977). Cell 12, 1-8.
3. Berget, S.M., Moore, C. and Sharp, P.A. (1977). Proc.Natl.Acad.Sci. USA 75, 4853-4857.
4. Akusjärvi, G. and Pettersson, U. (1979). Cell 16, 841-850.
5. Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. (1979). Cell 16, 851-861.
6. Ziff, E.B. and Evans, R.M. (1978). Cell 15, 1463-1476.
7. Hagenbüchle, O., Santer, M. and Argetsinger-Steitz, J. (1978). Cell 13, 551-563.
8. Chow, L.T. and Broker, T. (1978) Cell 15, 497-510.
9. Dunn, A.R., Mathews, M.B., Chow, L.T. and Sambrook, J. (1978). Cell 15, 511-526.
10. Aleström, P., Akusjärvi, G., Perricaudet, M., Mathews, M.B., Klessig, D. and Pettersson, U. (1980). Cell 19, 671-681.
11. Maxam, A. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Natl.Acad.Sci. U.S.A. 74, 560-564.
12. Favaloro, J., Treisman, R. and Kamen, R. (1980). Methods of enzymology 65, 718-749.
13. Rigby, P.W., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
14. Weaver, R.S. and Weissmann, C.W. (1979) Nucl. Acid. Res. 7, 1175-1193.
15. Akusjärvi, G. and Pettersson, U. (1978). Proc. Natl. Acad. Sci. USA 75, 5822-5826.
16. Persson, H., Pettersson, U. and Mathews, M.B. (1978) Virology 90, 67-79.
17. Mc Grogan, M., Spector, D.J., Goldenberg, C.J., Halbert, P. and Raskas, H.J. (1979) Nucl. Acid. Res. 6, 593-607.
18. Pelham, H.R.B. and Jackson, R.J. (1976) Eur. J. Biochem. 67, 247-256.
19. Maizel, J.V. (1971) Methods in virology, p. 180-244. Academic Press Inc. New York.
20. Bonner, W.M. and Laskey, R.A. (1974) Eur. J. Biochem. 46, 83-88.
21. Akusjärvi, G. and Persson, H. (1981). Nature 292, 420-426.
22. Seif, I., Khoury, G. and Dhar, R. (1979). Nucleic Acid. Res. 6, 3387-3398.
23. Breathnach, R., Benoist, C., O'Hare, K., Gannon, G. and Chambon, P. (1978). Proc.Natl.Acad. Sci. USA 75, 4853-4857.
24. Berk, A.J. and Sharp, P.A. (1978). Proc. Natl. Acad. Sci. 75, 1274-1278.
25. Southern, E.M. (1975) J. Mol. Biol. 98, 503-518.
26. Kozak, M. (1978) Cell 15, 1109-1123.
27. Jay, G., Nomura, S., Anderson, C.W. and Khoury, G. (1981) Nature 290, 113-118.