Volume 15 Number 8 1987 Nucleic Acids Research

All foot and mouth disease virus serotypes initiate protein synthesis at two separate AUGs

D.V.Sangar, S.E.Newton, D.J.Rowlands and B.E.Clarke

Wellcome Biotechnology Ltd, Ash Road, Pirbright, Surrey, UK

Received January 21, 1987; Revised and Accepted March 17, 1987

ABSTRACT

Translation of the foot and mouth disease virus genome in vitro and in vivo indicated that all seven serotypes initiate protein synthesis at two separate AUGs.
Sequence analysis of the region surrounding these AUGs has shown that the efficiency with which the initiating AUG is recognised is dependant on the flanking nucleotides. However, in vitro, the major factor determining which AUG is used is the concentration of Mg^{2+}.

INTRODUCTION

Foot and mouth disease viruses (FMDV) comprise the aphthovirus genus of the family Picornaviridae. The viruses are spherical, icosahedral particles, approximately 28nm in diameter and exist as seven distinct serotypes. The viruses contain a single segment of single stranded positive sense RNA of about 8.5kb which can programme in vitro translation systems to make virus polypeptides identical to those detected in infected cells (1).
In common with most other eukaryotic mRNAs, picornavirus RNA is polyadenylated at the 3' end; however the 5' end does not contain the characteristic "cap" structure but is blocked by a covalently linked protein, VPg. This protein can be removed from the RNA by the action of a host enzyme and plays no role in the translation of the RNA (2).
The FMDV and cardioviruses contain a polycytidylic acid tract of unknown function located, for the aphthoviruses, approximately 400 bases from the 5' end of the genome (3). This feature allows the dissection of the RNA into two regions by digestion with RNase H in the presence of oligo dG(3). The shorter fragment formed by this treatment, that is the region between VPg and the poly C tract, is termed the S fragment whereas the region of the RNA between the poly C tract and the 3' poly A tail is called the L fragment. Even though FMDV is not capped the RNA is an efficient message for protein synthesis both in vivo and in vitro but the factors which contribute to this efficiency are not clear. It is unlikely that the S

fragment or poly C tract have any role in translation since both full length and L fragment RNA make the same polypeptides *in vitro* and are equally efficient as messages (4). The first AUG which is followed by a long open reading frame (ORF) is located at nucleotide 677 from the poly C tract and before this AUG there are several other potential initiating AUGs. There is evidence with one strain of FMDV that one of these AUGs does initiate protein synthesis *in vitro*, producing a protein with a mol wt of approximately 10k (5). However, this feature is not conserved between different serotypes (6) and thus its significance is not clear.

We have shown previously (7) that, with FMDV type A, protein synthesis initiates at either of two AUGs, which are in the same reading frame, resulting in the production after proteolytic processing of proteins of mol wts 20k (Lab) or 16k (Lb). Using two different type A viruses we were able to demonstrate that the ratio of Lab:Lb differed with the strain of virus used and was dependent on whether the proteins were synthesised *in vivo* or *in vitro*. There appeared to be a mechanism operating *in vivo* to ensure the production of both Lab and Lb. We were therefore interested to determine whether this was a conserved feature of all FMDV serotypes. For this purpose we have sequenced representatives of each of the seven serotypes of the virus around the initiating region. We show in this paper that the possession of two initiating AUGs is a common feature of all the FMDVs examined and further confirm that the ratio of Lab:Lb varies both with the virus strain examined and also the conditions used to synthesise these proteins. The relevance of these results to proposed models for initiating eukaryotic protein synthesis is discussed.

MATERIALS AND METHODS

Preparation of Virus and Virus RNA

Representatives of each of the seven serotypes of FMDV were purified and the RNA extracted as previously described (3). The viruses chosen for examination were: A10 (61), A12 (119), A24 Cruzeiro; O1 BFS and O1 Kaufbeuren; C1 Oberbayern; Asia 1 Pak 1/54; SAT1 Bot 1/68; SAT2 3/57 and 2/67; SAT3 Bec 1/65.

Preparation of infected cell extracts

Monolayers of BHK 21 cells were infected with a high multiplicity of virus (approximate MOI 100:1). After 30 min the cells were washed with methionine free Eagle's medium and incubated in this medium until 2.5h

after infection, when they were pulsed for 5 min with 100µCi of (^{35}S) methionine. The monolayers were then washed with methionine free Eagle's medium, removed from the glass by adding 1 ml of disruption buffer (0.25M Tris pH6.8, 2% SDS, 2%, β- mercaptoethanol, 20% glycerol, 0.01% phenol red) and boiled for 5 min. This extract was used for polyacrylamide gel electrophoresis.

In vitro translation

All translations were done in the rabbit reticulocyte lysate system (8) (purchased from Amersham International) using 10µl of lysate and 2µg of RNA in each reaction. The lysate contained 10µCi of (^{35}S) methionine in each reaction.

Incubations were at 30°C and translation was stopped by the addition of 20µl of RNase A (2mg/ml) and incubation for 5 min at 30°C. Samples were prepared for electrophoresis by the addition of 50µl of disruption buffer and boiling for 1 min. Incubations requiring increased Mg^{2+} were carried out by the addition of 1/5 vol of 0.5MKCl, 1mMDTT, 10% glycerol, 50mMgCl_2 to the solution of RNA prior to the addition of this to the lysate. It was found that the use of this mixture gave more consistent results compared with the addition of Mg^{2+}alone.

Polyacrylamide gel electrophoresis

The gel system of Laemmli (9) was used in all experiments with a 10% resolving gel. In some experiments the gels were then treated with 'Amplify' (Amersham International).

Sequencing procedures

Two different sequencing procedures were used. For Asia 1, A12 and SAT2 3/57 cDNA was synthesised using oligo (dT) primer for the first strand synthesis. Double stranded cDNA was digested with Sau 3A or Taq 1 and cloned into Bam H1 or Acc 1 digested M13mp11. These M13 libraries were screened using a clone from the A10 serotype representing the Lab/Lb region (7). DNA sequencing of positive clones was by the chain termination method (10).

For all the other viruses used in these experiments the sequence of the Lab/Lb region was determined by primer extension dideoxy sequencing directly from virus RNA (11, 12) using 2µg viral RNA and a synthetic deoxynucleotide primer ('5 AAAGTTTTCTTCTCAC 3') complementary to the sequence of A10 located approximately 25 bases to the 3' side of the Lb initiation site, a region highly conserved in different serotypes of FMDV.

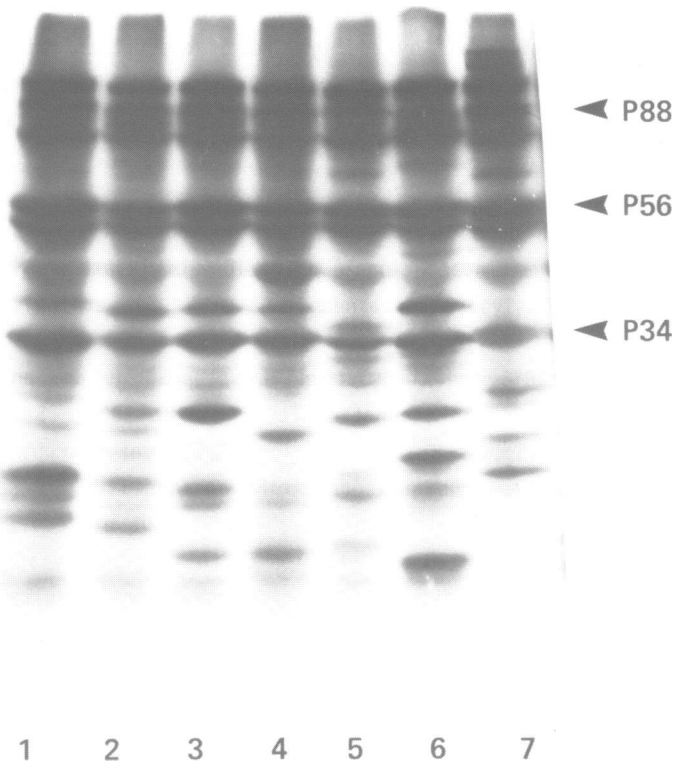

Figure 1 The polypeptides synthesised in infected BHK21 cells during a 5 min pulse with (^{35}S) methionine. Lane: (1) A10, (2) O_1, (3) C_1, (4) Asia 1, (5) SAT1, (6) SAT2, (7) SAT3. Three virus induced polypeptides are marked to indicate molecular weights (P88 = P1, P56 = P3d and P34 = P2c).

RESULTS

Protein synthesis in vivo and in vitro

The polypeptides synthesised in infected BHK21 cells are shown in Fig 1. Although the migration of the large polypeptides is similar in all seven serotypes, the migration of the smaller polypeptides is variable. It is therefore not possible from the above results to establish which polypeptides represent Lab and Lb. In order to determine the position of Lab and Lb for serotype A10 we took advantage of the finding that, in vitro, elevation of the Mg^{2+} concentration from 2mM to 4mM stimulates production of Lab with a concomitant decrease in Lb (Fig. 2).

A similar result was obtained with the other six serotypes (Fig.3). The

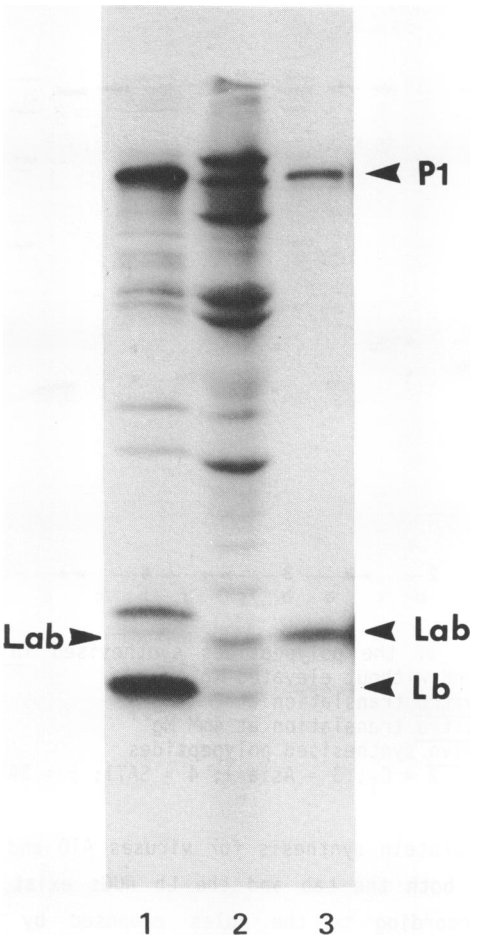

Figure 2 The effect of Mg^{2+} concentration on the polypeptides synthesised in vitro using serotype A10.
1 = Polypeptides synthesised in vitro at 2mM Mg^{2+}
2 = Polypeptides synthesised in vivo during a 7min pulse with (^{35}S) methionine
3 = Polypeptides synthesised in vitro at 4mM Mg^{2+}

identity of these proteins, for serotypes O1, A12 and C1, was confirmed by partial proteolytic digestion (results not shown). These results show that all seven serotypes of FMDV use two alternative initiation sites in the same reading frame for protein synthesis.

Sequences surrounding the initiation sites for protein synthesis

We have previously observed that the flanking nucleotides surrounding the

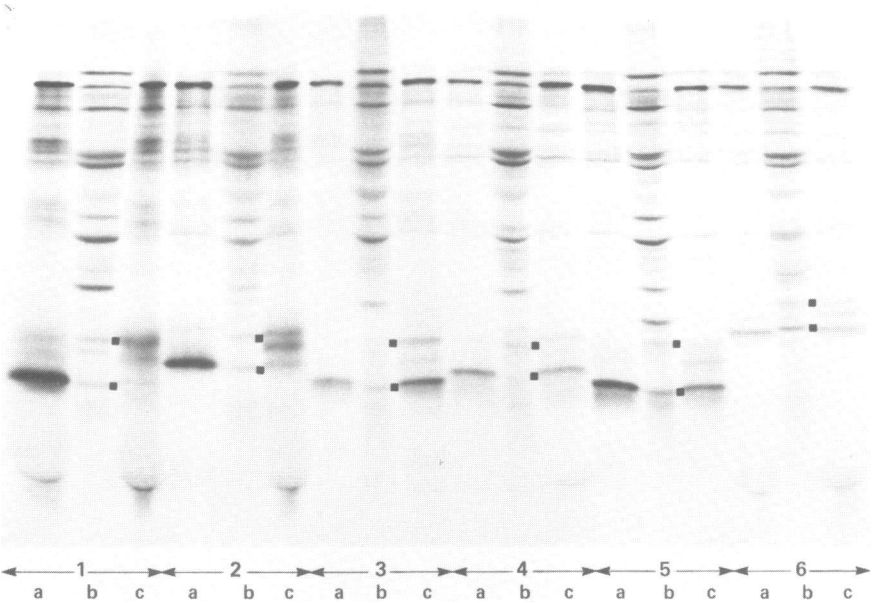

Figure 3 Comparison of the polypeptides synthesised in vivo with those made in vitro with and without elevated Mg^{2+}.
Lane a = In vitro translation at 2mM Mg^{2+}
Lane c = In vitro translation at 4mM Mg^{2+}
Lane b = In vivo synthesised polypeptides
1 = Serotype O_1; 2 = C_1; 3 = Asia 1; 4 = SAT1; 5 = SAT2; 6 = SAT3.

initiation site for protein synthesis for viruses A10 and A12 are not the same (7). For A10 both the Lab and the Lb AUGs exist in a favourable sequence context according to the rules proposed by Kozak (13) for initiation of translation. However, for A12 the Lab AUG exists in an unfavourable sequence whilst the Lb site is favourable.

We have now determined the nucleotide sequences in the region of the two initiating AUGs for representatives from the other six serotypes of FMDV (Fig 4). It is evident that the strains analysed could be separated into three groups. Firstly, the A10 virus possesses favourable initiation sites both for Lab and Lb and under standard conditions of translation in vitro RNA from this virus makes similar amounts of both proteins. The second group comprises most of the other serotypes examined A12, C1, Asia 1, SAT1, SAT2 (3/57 and 2/67) and SAT3. These viruses have a poor site for initiation of protein synthesis at Lab, (PyPyPyPyAUGA) but a favourable site for initiation at Lb. Under standard in vitro conditions, little or

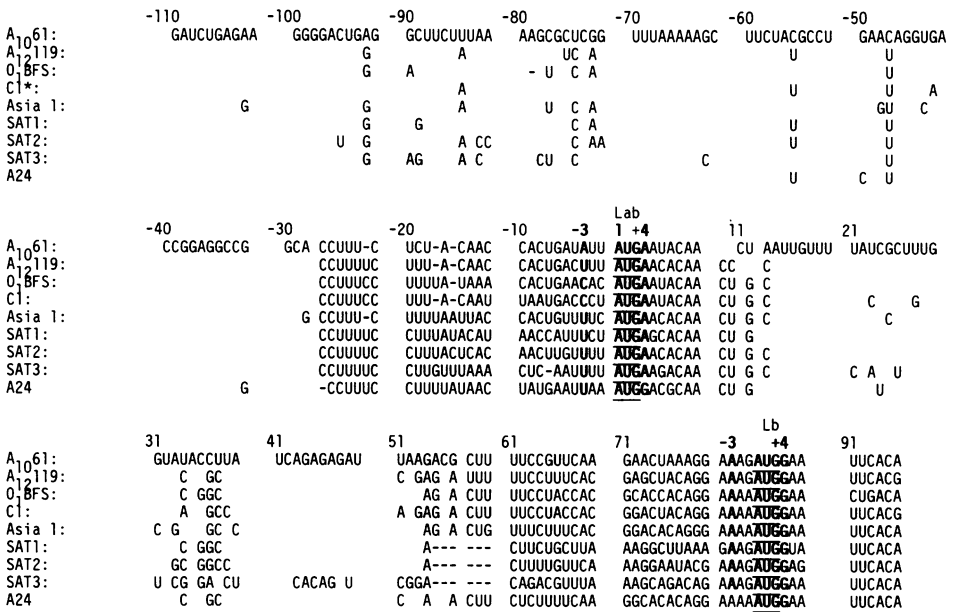

Figure 4 Comparison of the nucleotide sequences of the seven serotypes of FMDV from nucleolute -110 to +96. For the majority of the sequence only differences are noted; blank spaces denote identity. The symbol |-| denotes a deletion which is required to align the sequences. For both Lab and Lb, a complete sequence comparison is given from the start of the polypyrimidine tract to position +12 from the A of each AUG. This region is boxed in. The sequence for C_1 is from previously published data (5) and has only been determined up to 87 bases upstream of the AUG for Lab. The sequence for A24 only extends to -82.

no Lab was made and initiation was mainly at Lb. The third group comprises the viruses of the O serotype (O1 BFS and O1 Kaufbeuren) and A24. This group possesses an unfavourable site for initiation at Lab (PyxxAUGA) but a favourable Lb site. However, in vitro both Lab and Lb were detected. Although these viruses have a pyrimidine at position -3 from the Lab AUG this was surrounded by purines; it is possible that these may compensate in some way for the lack of a purine at -3.

In all the examples studied, with the exception of the Lb initiation region of SAT3, both the Lab and Lb AUGs are preceded by a polypyrimidine tract commencing 15-20 bases upstream of the AUG. This feature has been observed in several eukaryotic mRNAs (14, 15).

The sequence of the untranslated regions of the virus RNAs was highly conserved to the 5' side of the polypyrimidine tract preceding Lab. This

	Lab	Lb
$A_{10}61$:	MNTTNCFIALVYLIREIKTLFRSRTKGKMEFT	
$A_{12}119$:	MNTTNCFIALVHAIREIRAFFLSRATGKMEFT	
O_1BFS:	MNTTDCFIALVQAIREIKALFLPRTTGKMELT	
C1*:	MNTTDCFTAVVNAIREIRALFLPRTTGKMEFT	
Asia 1:	MNTTDCFIPLLYALREIKALFLSRTQGKMEFT	
SAT 1:	MSTTDCFIALVQAIREIK--LLLKGLKKMVFT	
SAT 2 3/57:	MNTTDCFIALVQAIREIK--LLFKGIRKMEFT	
SAT 3:	MKTTDCFNVLFEIFHRFG--QTFKADRKMEFT	
A_{24}:	MDATDCFIALVHAIREIKTLLFSRHTGKMEFT	
O_1Kauf*:	MNTTDCFIALVQAIREIKALFLSRTTGKMELT	
SAT 2 2/67:	MNTTDCFIALVQAIREIK--LLFKGIRKMKLT	

Figure 5 The amino acid sequence of the seven serotypes of FMDV for the La region. The sequences are aligned by the use of deletions (-) where required. The sequences for O_1K and C_1 are from previously published data (5).

conservation contrasts with the variation observed in the coding region for the unique portion of Lab, in which only four amino acids are conserved across the serotypes (Fig 5). The region of the RNA coding for the Lb polypeptide shows a much greater degree of conservation (data not shown).

DISCUSSION

The most generally accepted mechanism by which eukaryotic ribosomes recognise the initiating AUG for protein synthesis is the "scanning model" proposed by Kozak (16). In this model ribosomes enter at the free 5' end of the RNA and scan until encountering the first AUG, at which stage protein synthesis begins. This simple model has since been modified to take into account the effect of flanking nucleotides around the AUG, which have an extensive modulating effect, and also that internal AUGs can act as initiating sites provided that these upstreams are followed by terminating codons (13).

Aphthovirus RNA has a number of features which distinguish it from the majority of other eukaryotic RNAs and which make it probably the most extreme RNA for the scanning model to explain. The RNA is not capped and

secondary structure analysis of the S fragment predicts a highly stable hair-pin structure extending over the entire 360-370 nucleotides (12). The Δ G of this structure has been calculated to be approx. -150K cals. This compares with a value of -50K cals which, when presented as a complete base paired hairpin, blocks the passage of ribosomes (17). It should be noted, however, that the S fragment does not present any complete base paired region which has a Δ G lower than this critical figure (A.R. Carroll, personal communication). Furthermore, initiation of protein synthesis occurs approximately 1190 nucleotides from the 5' end of the genome, and if ribosomes scan this region they have to bypass a total of 10 AUGs before encountering the correct AUG (6). Nonetheless, a number of features of FMDV RNA sequence are consistent with the scanning model (16), eg. with the A10 serotype all of the upstream AUGs are followed by in phase termination codons and many of them are in an unfavourable context (18). However, the observed relative frequency of initiation at the Lab and Lb AUGs with different FMDVs described in this paper are not easily explained by Kozak's model for ribosome recognition of initiating AUGs during the scanning process.

The efficiency of initiation at the Lab initiation site does depend on the nature of the flanking nucleotides. With A10 virus, where the AUG is located in a favourable context, Lab is synthesised efficiently in infected cells and under defined salt conditions it is the main protein made in vitro. However there are significant amounts of Lb synthesised, which would not be predicted by the scanning model. It is also difficult to correlate the results obtained with the O serotypes and A24 with the scanning model. With these examples the Lab AUG is not in a favourable environment yet Lab is made with a similar relative efficiency to that found with A10. It is possible that the purines present around the -3 pyrimidine have a modifying effect but this has not been observed with other eukaryotic messages (18). The third group of viruses (C1, SAT1, SAT2, SAT3, Asia 1 and A12) have a poor context for initiation at the Lab AUG and indeed in vivo and under standard conditons in vitro these viruses make little Lab. In all the viruses we have examined the sequence around the Lb AUG is favourable for initiation and it appears that if ribosomes bypass the Lab they all initiate at this position.

An important finding that must be considered in all the above discussions is the effect of Mg^{2+} concentration on the specificity of initiation. Elevation of the Mg^{2+} concentration increased the amount of Lab produced

with a concomitant decrease in Lb. This switch to Lab production occurred independent of the sequence around the Lab AUG. The effect of Mg^{2+} was specific for the Lab initiating AUG and there was no evidence of any false initation at other AUGs, either in the untranslated region or within the coding region of the genome. Changes in the specificity of initiation has been reported for poliovirus where elevation of the Mg^{2+} concentration alters the AUG selected; however, in that example it has been suggested that the alternative AUG is located in the 5' untranslated region (19).

The functions of either Lab or Lb are not fully understood, it has been reported that Lb is a protease (20) which cleaves itself off the growing nascent strand, although it appears unlikely that this is its sole function. It is interesting that the common region of Lab/Lb is far more conserved than the unique region of Lab and it is this conserved region which has proteolytic activity (20). It is possible that the Lab unique region modifies the activity of Lb in some way. Whatever the function of Lab may be it is presumably important since it is conserved in all seven serotypes.

In conclusion it is not clear at present whether translation of FMDV RNA occurs by a true scanning mechanism or whether some of the untranslated region is bypassed. It may be relevant that the S fragment forms a hairpin structure up to the poly C tract and it is possible that ribosomes could feed directly on to the poly C tract before scanning starts. Alternatively, ribosomes may enter, perhaps by recognition of the polypyrimidine tracts closer to the initiating AUG. If this is correct it could be predicted that initiation at Lab could be inhibited independently of Lb except for SAT3 which does not have a polypyrimidine tract prior to the Lb AUG. Further studies are in progress using hybrid arrest translation with synthetic oligonucleotides complementary to sequences in the untranslated region to determine where ribosomes start scanning the aphthovirus genome. The most important factor controlling which of the two possible AUGs is recognised *in vitro* is the level of Mg^{2+} and experiments are in progress to examine whether this is also important *in vivo*.

ACKNOWLEDGEMENTS

We are grateful to A. Brown and R.P. Clark for skilled technical assistance, A. Syred for synthesising the oligonucleotide primers, R.O. Campbell for supplying some of the purified RNA and F. Brown and A.R. Carroll for advice and for critically reading the manuscript.

REFERENCES

(1) Sangar, D.V. (1979). J.Gen.Virol 45 1-13.
(2) Ambros, V., Petterson, R.F. and Baltimore, D. (1978). Cell 15 1439-1461.
(3) Rowlands, D.J., Harris, T.J.R. and Brown. F. (1978). J.Virol 16 335-343.
(4) Sangar, D.V., Black, D.N., Rowlands, D.J., Harris, T.J.R. and Brown F. (1980). J. Virol 33 59-68.
(5) Forss, S., Strebel, K., Beck, E. and Schaller, H., (1984). Nucl. Acids Res. 12 6587-6601.
(6) Clarke, B.E., Carroll, A.R. and Rowlands, D.J. Manuscript in prep.
(7) Clarke, B.E., Sangar, D.V., Burroughs, J.N., Newton, S.E., Carroll, A.R. and Rowlands, D.J. (1985). J.Gen.Virol 66 2615-2626.
(8) Pelham, H.R.B. and Jackson, R.J. (1976). Biochemistry 67 247-256.
(9) Laemmli, U.K. (1970). Nature (London) 227 680-685.
(10) Sanger F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA 77 6468-6472.
(11) Caton, A.J., Brownlee, G.G., Yewdell, J.W. and Gerhard, W. (1982). Cell 31 417-427.
(12) Newton, S.E., Carroll, A.R., Campbell, R.O., Clarke, B.E. and Rowlands D.J. (1985). Gene 40 331-336.
(13) Kozak, M. (1984). Nucleic Acids Research 12 3873-3893.
(14) Beck, E. Forss, S., Strebel, K., Cattaneo, R. and Feil, G. (1983). Nucleic Acids Research 11 7873-7885.
(15) Lewin, B. (1980). In Gene expression 2 679-683.
(16) Kozak, M. (1980). Cell 22 7-8.
(17) Kozak, M. (1986). Proc.Natl.Acad.Sci. USA 83 2850-54
(18) Kozak, M. (1986). Cell 44 283-292.
(19) Dorner, A.J. Dorner, F.L., Larsen, G.R., Wimmer, E., and Anderson, C.W. (1982). J. Virol 42, 1017-1028.
(20) Strebel, K. and Beck, W. (1986). J. Virol 56 893-899.