

# Ribosome-messenger recognition in the absence of the Shine-Dalgarno interactions

N.V. Tzareva<sup>a</sup>, V.I. Makhno<sup>b</sup>, I.V. Boni<sup>a,\*</sup>

<sup>a</sup>*M.M. Shemyakin and Yu.A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117871 Moscow, Russian Federation*

<sup>b</sup>*Petersburg Nuclear Physics Institute, Russian Academy of Sciences, 188350 Gatchina, Russian Federation*

Received 25 November 1993

## Abstract

In an attempt to understand how *Escherichia coli* ribosomes recognize the initiator codon on mRNAs lacking the Shine-Dalgarno (SD) sequence, we have studied 30S initiation complex formation in extension inhibition (toeprinting) experiments using (–SD)mRNAs which are known to be reliably translated in *E. coli*: the plant viral messenger AIMV RNA 4 and two chimaeric mRNAs coding for  $\beta$ -glucuronidase (GUS) and bearing the 5'-untranslated sequence of TMV RNA ( $\Omega$ ) or the  $\Omega$ -derived sequence (CAA), as 5'-leaders. Ribosomal protein S1 and IF3 have been found to be indispensable for translational initiation. Protein S1 appears to be a key recognition element. S1 binds to sequences within the leaders of (–SD)mRNAs thus providing their affinity to *E. coli* ribosomes.

**Key words:** Translational initiation; RNA–protein interaction; Plant viral RNA leader; Protein S1; Toeprinting; Footprinting

## 1. Introduction

*E. coli* ribosomes are able to translate mRNAs from a wide variety of sources, independently of the length of the Shine-Dalgarno (SD) region [1]. The most intriguing examples of such 'non-specific' translation are plant viral messages such as AIMV RNA 4 and TMV RNA, which can be recognized with high specificity and reliably translated by *E. coli* ribosomes despite the absence of any consensus SD sequences [2,3]. Moreover, 5'-untranslated sequences of these mRNAs or derivatives thereof serve as translational enhancers in the *E. coli* system when placed at the 5'-end of the coding sequence of the reporter gene [4–8].  $\Omega$ , the 68-base leader of TMV RNA and  $\Omega$ -derived sequences as well as the 37-base leader of AIMV RNA 4 are devoid of G-residues and therefore are not able to form a complementary duplex with the 3'-end of 16S RNA. The question is what kind of interaction promotes ribosome-messenger recognition in this unusual case.

A key component which makes *E. coli* ribosomes non-specific in that they are able to translate mRNAs with

weak SD sequences seems to be ribosomal protein S1 [1]. Ribosomes from *B. subtilis* and most other Gram-positive species lack S1 and exhibit specific translation being able to translate only homologous mRNAs, or mRNAs with 'strong' SD regions [1]. Protein S1, the largest ribosomal protein, has bidomain structure [9]. It is attached to the 30S ribosomal subunit via its N-terminal globular domain by means of protein–protein interactions [10,11], whereas its elongated RNA-binding C-terminal domain functions as an 'mRNA catching arm' of the ribosome [12]. Earlier, we have shown that during initiation of translation protein S1 interacts with mRNAs in a site-specific manner [13]. On several natural mRNAs we have identified S1-binding sites and proposed that they serve as additional recognition signals for *E. coli* ribosomes during initiation of protein synthesis [13]. Localization of the S1-binding sites within 5'-untranslated regions of the messengers [13] makes it reasonable to suggest that recognition of the (–SD) plant viral leaders by *E. coli* ribosomes can also be provided by interactions with protein S1.

Here, we present the results of our studies on the role of S1 in initiation complex formation between *E. coli* 30S ribosomal subunits and (–SD)mRNAs containing the plant viral leaders or derivatives thereof. AIMV RNA 4 and two chimaeric  $\beta$ -glucuronidase mRNAs bearing  $\Omega$ -derived sequences upstream of the coding region have been obtained by in vitro transcription and tested for the ability to form 30S initiation complexes by toeprinting technique [14]. We have found that it is protein S1 that

\* Corresponding author. Fax: (7) (095) 330-6538.

**Abbreviations:** Omega ( $\Omega$ ), 5'-untranslated sequence of tobacco mosaic virus (TMV) RNA; AIMV RNA 4, the subgenomic mRNA for alfalfa mosaic virus coat protein; GUS, the sequence coding for  $\beta$ -glucuronidase of *Escherichia coli*; RT, reverse transcription; AMV RTase, reverse transcriptase from avian myoblastosis virus.

provides ribosome-messenger recognition in the absence of SD interactions, and identified target sites for S1 within the 5'-untranslated plant viral sequences.

## 2. Materials and methods

### 2.1 30S subunits and protein S1

Several types of 30S subunits isolated from *E. coli* MRE-600 strain have been used in the experiments. Initiation factor-free 30S subunits were isolated from so-called 'tight 70S ribosomes'. 30S particles with high content of IF3 (0.5–1 molecules per subunits) were separated from cell extract as free 30S subunits during the first step of modified isolation procedure for 70S 'tight couples' [15]. The detailed protocol will be published elsewhere. In fact, these particles are so-called 'native 30S subunits' with the suggested initiation function in the cell [16]. 30S particles lacking protein S1 were prepared using poly(U)-Sepharose [9]. Protein S1 was isolated and purified as described previously [11].

### 2.2 Preparation of mRNAs

RNA synthesis in vitro and purification were performed with the reagents and according to protocols of the Riboprobe Gemini System II (Promega). The plasmid pT7-2-42 bearing the sequence for AIMV RNA 4 under the control of the T7 promoter was provided by K. Langeris (University of Leiden). The plasmid was linearized at the *EcoRI* site and used for mRNA preparation.

The plasmids pJII1001 and pJII1031, containing chimaeric genes coding for  $\beta$ -glucuronidase (GUS) [17] with the  $\Omega$ -derived 5'-untranslated regions, were a gift of Dr. T.M.A. Wilson. The plasmid pJII1001 linearized at the *SalI* site downstream from the GUS sequence did not produce workable quantities of the  $\Omega$ -GUS mRNA from the T7 promoter, possibly because of the inefficiency of incorporation of U-residues in the first nucleotides transcribed (see also [8]). To obtain sufficient amount of the  $\Omega$ -GUS mRNA, the *EcoRI-SalI* fragment of the pJII1001 was inserted into the corresponding sites of pGEM-4 (Promega) under the control of the SP6 promoter. The resulting plasmid pG4J-01 linearized at the *SalI* site gave sufficiently high yield of the  $\Omega$ -GUS transcript.

The second GUS encoding plasmid pJII1031 contains behind the T7 promoter one of the two main sequence motifs of the original  $\Omega$ , (CAA)<sub>n</sub> as a 5'-untranslated sequence (see [8]). To obtain the (CAA)<sub>n</sub>-GUS transcript for toeprinting assay, pJII1031 was linearized at the *EcoRV* site that led to shortening of the GUS sequence up to the position +558 without any influence on translational initiation efficiency of the resulting mRNA.

### 2.3 Extension inhibition analysis (toeprinting) of 30S initiation complex formation with (–SD)mRNAs

To form initiation complexes, 30S ribosomal subunits, the prepared in vitro and purified (–SD)mRNAs, and uncharged initiator tRNA (Boehringer-Mannheim) were incubated in binding buffer (20 mM Tris-HCl, pH 7.6/10 mM MgCl<sub>2</sub>/100 mM NH<sub>4</sub>Cl/6 mM  $\beta$ -mercaptoethanol) for 10 min at 37°C. Usually, a 10  $\mu$ l incubation probe contained about 0.4 pmol of the mRNA annealed with the 5'-labeled primer, 4 pmol of 30S subunits and 40 pmol of tRNA<sup>Met</sup>. In experiments on S1 dependence of initiation complex formation, free S1 in binding buffer or an equal volume of the buffer was added. Primers used in primer extension reactions had complementarity to the regions within the known coding sequences of AIMV RNA 4 [18] and the GUS mRNA [17] (corresponding positions indicated in brackets):

AIMV RNA 4 primer  
5'-AGGCTTCGGCAGTTGAGC-3' (+81 to +98)  
GUS primer  
5'-GCGATCCAGACTGAATGC-3' (+57 to +74)

Efficiency of initiation complex formation was tested by toeprinting according to Hartz et al. [14] with minor modifications.

### 2.4 RNase protection experiments (footprinting of the S1-binding sites)

For the  $\Omega$ -GUS mRNA, 0.3–0.4 pmol of the mRNA was incubated in binding buffer with increasing amounts of S1 or equal volumes of the buffer for 10 min at 37°C. Subsequently, 3  $\mu$ g of carrier tRNA and

1  $\mu$ l (0.25 U) of RNase PhyM (Pharmacia) were added. The reaction volume was 10  $\mu$ l. Digestion was allowed to proceed for 15 min at 37°C and stopped by the addition of 20  $\mu$ l of 0.4 M Na-acetate (pH 5.2)/20 mM EDTA containing 0.2 mg/ml of carrier tRNA, followed by phenol extraction and precipitation with ethanol.

AIMV RNA 4 was digested with RNase T2 (10<sup>–4</sup> U, Sankyo) for 10 min at 37°C in the presence of or without S1. All procedures were essentially the same as described above for the  $\Omega$ -GUS transcript. The digested mRNAs were analysed by primer extension.

In the case of the (CAA)<sub>n</sub>-GUS mRNA, a special protocol was developed to minimize a number of operations in order to prevent the leader from an undesirable decay. 0.3 pmol of the (CAA)<sub>n</sub>-GUS transcript in the presence of 0.7 pmol of the 5'-labeled GUS primer was incubated without or with increasing amounts of S1 in 7  $\mu$ l of binding buffer for 15 min at 37°C. Subsequently, each reaction probe was supplied with 1  $\mu$ l (3 mg/ml) of carrier tRNA, 1  $\mu$ l (0.1 U) of RNase PhyM and 1  $\mu$ l of RT-mix (0.8 mM dNTP, 1.8 U of AMV reverse transcriptase (Omutninsk) in binding buffer). Reaction was allowed to proceed for 15 min at 37°C and stopped by addition of 40  $\mu$ l of 0.5% SDS/5 mM EDTA, followed by phenol extraction and ethanol precipitation. Primer extension products were analysed on a 8% sequencing gel.

## 3. Results

### 3.1 Initiation complex formation between *E. coli* 30S ribosomal subunits and (–SD)mRNAs: requirements for protein S1 and IF3

The mRNAs chosen for the present work contain the naturally occurring 5'-untranslated region (AIMV RNA 4) or the derivatives of the original TMV leader sequence ( $\Omega^x$ -GUS and (CAA)<sub>n</sub>-GUS mRNAs). The nucleotide sequences of the translational initiation regions of these (–SD)mRNAs are shown in Fig. 1. Additional nucleotides at the 5'-ends of the transcripts appearing as a result of gene engineering manipulations are omitted. Toeprinting data show (see below) that they have no effect on specificity of translational initiation. The results of sequencing revealed that the transcript obtained from pG4J-01 ( $\Omega^x$ -GUS mRNA) had two point deletions in the original  $\Omega$  sequence (deleted nucleotides indicated in parentheses; Fig. 1).

To test the (–SD)mRNAs for the ability to form translational initiation complexes with *E. coli* ribosomes, toeprinting technique developed by Hartz et al. [14] has been used. In general, toeprinting efficiency is a good indication of mRNA translational activity. Initiation factor-free 30S subunits plus uncharged initiator tRNA form detectable by toeprinting initiation complexes with the majority of phage or bacterial messages ([13,14,19] and our unpublished results). We found that this was not the case with the (–SD)mRNAs. Among the several preparations of 30S ribosomal subunits used in toeprinting experiments, only those of them which contained initiation factor 3 (IF3) were able to form ternary initiation complexes generating revertase stop signals (toeprints). Another component that was found to be absolutely indispensable was protein S1. Fig. 2 illustrates these results obtained with AIMV RNA 4, but the same requirements were shown for the both GUS mRNAs as well.



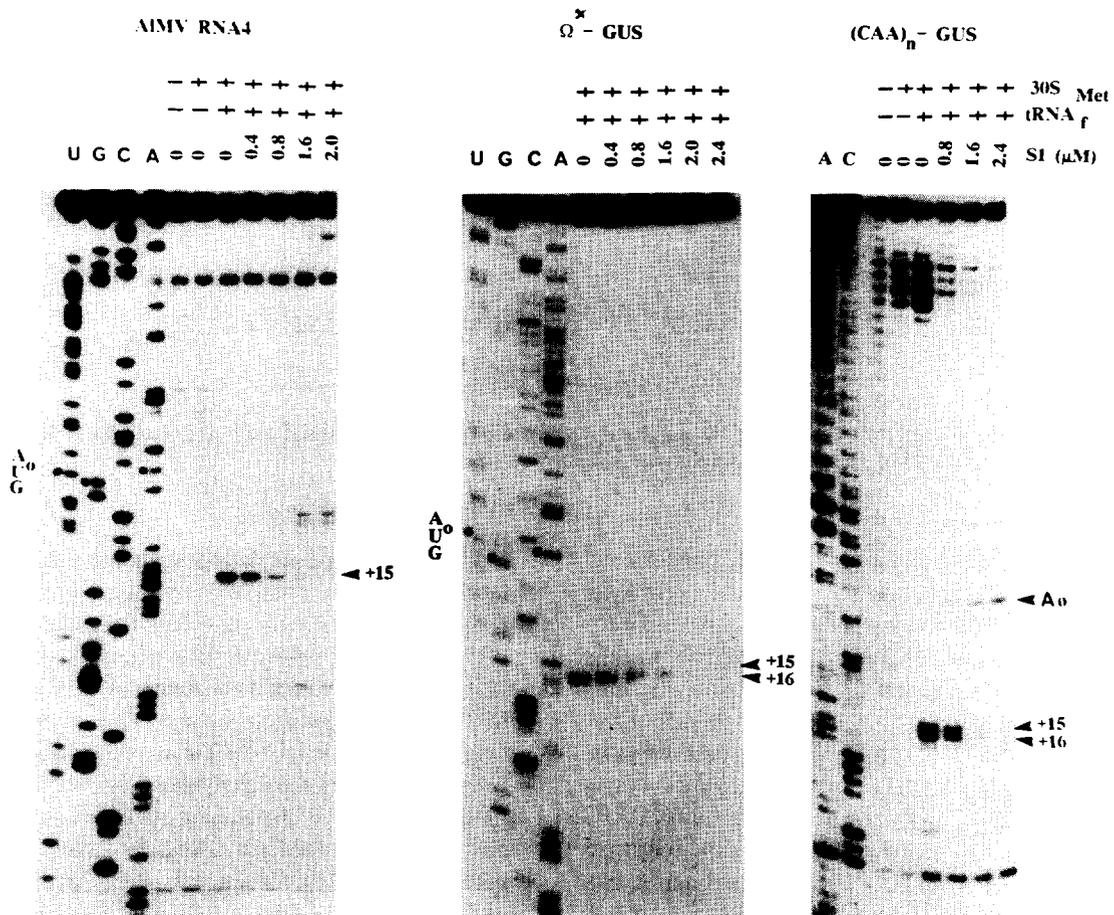


Fig. 3. Effect of the addition of free S1 on the ternary complex formation with the  $(-SD)$ mRNAs. The toeprint positions are indicated by arrows. Toeprinting experiments were performed as described in section 2; the concentrations of free S1 are indicated on the top of the lanes.

element (Fig. 4B). A high affinity to polypyrimidines is a characteristic feature of S1 [9]. Earlier we have found U-rich sequences to be the targets for S1 on the *ssb* mRNA and phage RNAs  $Q_{\beta}$  and fr [13]. The length of the S1 protected region on AIMV RNA 4 (about 22 nucleotides) indicates that two molecules of S1 can sit side by side on one RNA chain (an optimal binding site for one molecule of S1 is 10–12 nucleotides [9]).

It seems to be interesting, that  $\Omega^x$  and the  $(CAA)_n$  leader expose different targets for S1 binding. The digestion patterns (Fig. 4A) show that the  $(CAA)_n$  leader is easy cleavable that points to the opened single-stranded conformation, whereas the central C,A-rich portion of  $\Omega^x$  is not accessible for ribonuclease attacks, perhaps because it is involved in secondary structure formation. In contrast, the A,U-rich 3'-portion of  $\Omega^x$  is both well-digested and strongly protected by S1. Most likely, the higher efficiency of the  $(CAA)_n$ -GUS mRNA is explained by its single-stranded opened conformation because it is well known that the lack of secondary structure in a translational initiation region favours the initiation complex formation [23].

As in the case of AIMV RNA 4, the size of the S1-

protected regions in the both GUS mRNAs is large enough to bind two ( $\Omega^x$ ) or three [ $(CAA)_n$ -leader] S1 molecules (Fig. 4B). This points to the possible cooperativity of S1 binding.

The results suggest that not only U-rich sequences (AIMV RNA 4, RNAs  $Q_{\beta}$  and fr, the *ssb* mRNA) but other base motifs can serve as targets for S1 within 5'-untranslated regions of mRNAs. The fact that regions found to be crucial for recognition of  $(-SD)$ mRNAs by ribosomes have different sequences without any homology, makes it unlikely that some complementarity between 16S RNA and leaders underlies the recognition mechanism. Base pairing between  $\Omega$ -derived leaders and 16S RNA have been proposed earlier as a possible explanation of their efficiency as translational enhancers in *E. coli* [7,8]. It is more reasonable to suppose that RNA-protein interactions work in these case and protein S1 as a part of ribosomal mRNA binding centre recognizes and binds  $(-G)$  plant viral RNA leaders thus providing their affinity to the ribosome. The repressor function of S1 might be easily explained in this case as a competition between free S1 and S1 anchored on the 30S ribosome for the same target on the mRNA.

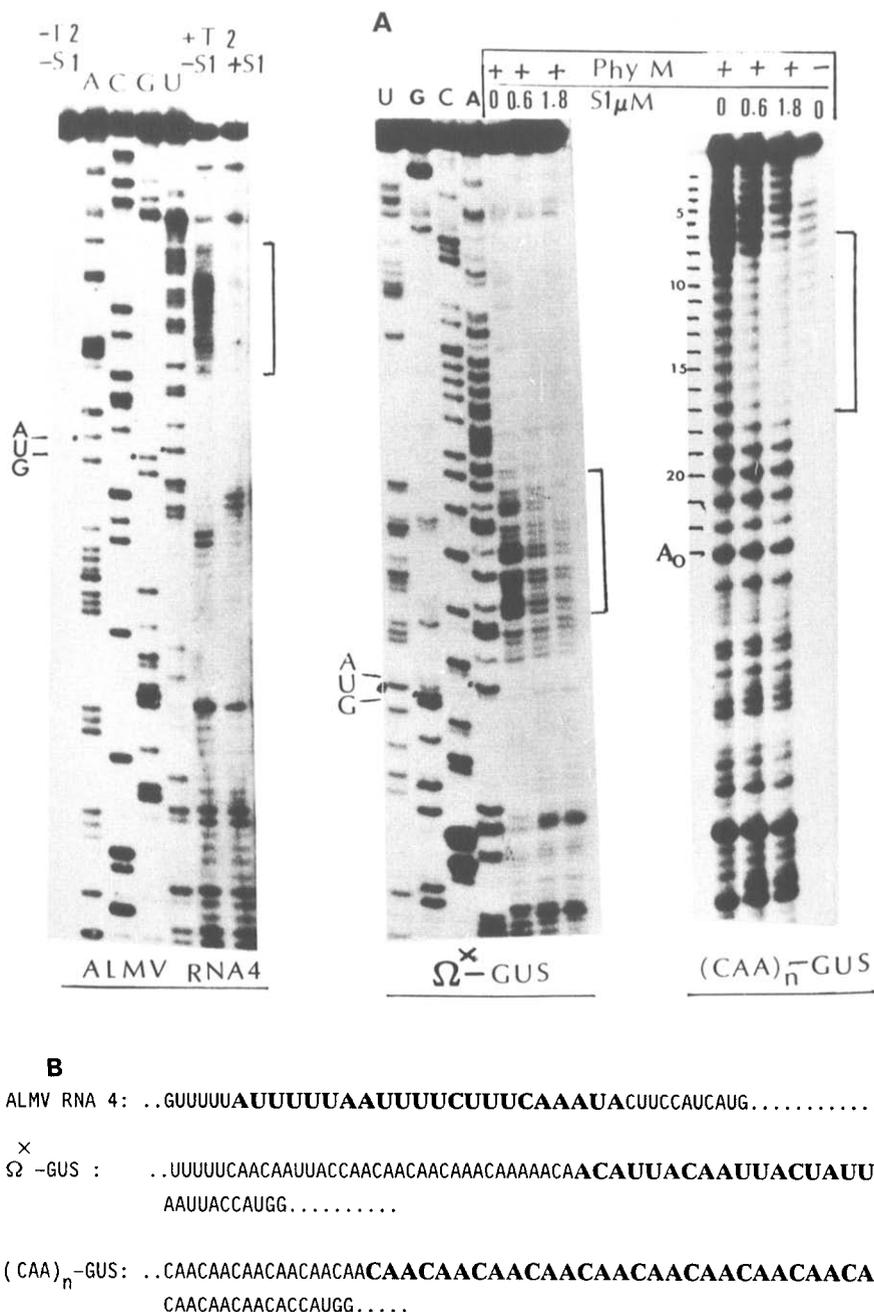


Fig. 4. (A) Primer extension RNA footprinting of protein S1 bound to the (-SD)mRNAs. RNase digestions were performed as described in section 2. The S1 concentration in a partial RNase T2 digest of ALMV RNA 4 was 2  $\mu$ M. Increasing concentrations of S1 in partial RNase PhyM digests of the  $\Omega^x$ -GUS and (CAA)<sub>n</sub>-GUS mRNAs are given on the top of the lanes. The bands corresponding to the CAA repeats in partial digests of the (CAA)<sub>n</sub>-GUS mRNA are indicated by numbers on the left of the corresponding gel autoradiogram. (B) The nucleotide sequences of the (-SD)mRNA leaders with indication of the regions protected by S1 against RNases (boldfaced).

#### 4. Discussion

In this article we present evidence for that *E. coli* 30S ribosomal subunits are able to recognize the translational starts on (-SD)mRNAs bearing 5' untranslated regions of plant viral RNAs or derivatives thereof, and to form stable ternary initiation complexes only in the presence of ribosomal protein S1 and IF3. As was shown

earlier, IF3 does not influence the affinity of the 30S subunit for mRNAs with or without SD sequence [24,25], its role in the initiation process (besides subunit anti-association function) concerns mainly the selection of the initiator tRNA [19]. IF3 can be regarded as kinetic effector and fidelity factor of 30S initiation complex formation [26]. Hartz et al. observed stimulative effect of IF3 on ternary complex formation detectable by

toeprinting [19]. Our results can be interpreted as an extreme case of such stimulation when detectable amounts of 30S initiation complexes can be formed only in the presence of IF3. In the absence of the SD duplex which enhances the local concentration of the initiation codon in the vicinity of ribosomal P site, IF3 seems to be a crucial factor facilitating codon–anticodon interactions.

As mentioned above, a key recognition element providing the affinity of mRNA to the ribosome is protein S1 [1,9,12,13]. In the absence of S1 *E. coli* ribosomes are able to translate only messengers with strong SD domains [27]. Earlier we suggested, that 30S-mRNA recognition can be provided by two types of interactions: RNA–RNA (SD and other possible base pairing between 16S RNA and mRNA), and RNA–protein interactions between mRNA and protein S1 [13]. Different mRNAs can use both or only one of them. We believe that mRNAs bearing plant viral leaders represent the case when mRNA–protein S1 interactions play the most important role in initiation complex formation. According to our observations, efficiency of initiation inhibition by an excess of free S1 correlates well with the specific contribution of mRNA–protein S1 interactions to mRNA binding by ribosomes. Among the mRNAs tested by toeprinting for S1-mediated inhibition of ternary complex formation (–SD)mRNAs with plant viral leaders reveal the most pronounced effect (our unpublished results).

In principle, one might argue that recognition of plant viral 5' untranslated regions by *E. coli* ribosomes is an artificial event that bears no relationship to the ribosome–messenger interactions in a bacterial cell. However, in *E. coli*, besides the mRNAs with strong or moderate SD domains, there exist messengers with minor SD or even without any similarity to the consensus SD sequence. At the same time, these 'minor SD' mRNAs are translated with significant efficiency, perhaps because the requirement for SD interactions can be compensated by the alternative recognition mechanism including interactions of mRNAs with protein S1 (see [13]). As we have shown here, protein S1 is able to bind RNA sequences with different base motifs that allows it to be involved in binding of wide mRNA spectrum. Structural basis underlying such wide sequence specificity of S1 is not yet clear. Most likely, it is connected with the complex design of the RNA-binding domain of S1 which comprises four highly homologous repeats with possible different recognition potentials.

*Acknowledgements:* This work was supported by Russian State Program 'Cell-Free Protein Synthesis'. We thank T.M.A. Wilson for the

providing of the pJII1001 and pJII1031 plasmids and helpful discussion on the problem. We also thank K. Langeris for providing pT7-2-42 and I. Chernov for synthesis of deoxyoligonucleotide primers.

## References

- [1] Roberts, M.W. and Rabinowitz, J.C. (1989) *J. Biol. Chem.* 264, 2228–2235.
- [2] Castel, A., Kraal, B., Konieczny, A. and Bosch, L. (1979) *Eur. J. Biochem.* 101, 123–133.
- [3] Wilson, T.M.A. (1986) *Virology* 152, 277–279.
- [4] Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, Ph.C. and Wilson, T.M.A. (1987) *Nucleic Acids Res.* 15, 3257–3273.
- [5] Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, Ph.C. and Wilson, T.M.A. (1987) *Nucleic Acids Res.* 15, 8693–8711.
- [6] Gallie, D.R., Sleat, D.E., Watts, J.W., Turner, Ph.C. and Wilson, T.M.A. (1988) *Nucleic Acids Res.* 16, 883–893.
- [7] Gallie, D.R. and Kado, C.I. (1989) *Proc. Natl. Acad. Sci. USA* 86, 129–132.
- [8] Wilson, T.M.A., Saunders, K., Dowson-Day, M.J., Sleat, D.E., Trachsel, H. and Mundry, K.W. (1990) in: *Post-Transcriptional Control of Gene Expression* (McCarthy, J.E.G. and Tuite, M.F. eds.) NATO ASI Series, vol. H49, pp. 261–275, Springer-Verlag, Berlin/Heidelberg.
- [9] Subramanian, A.R. (1983) *Prog. Nucleic Acid Res. Mol. Biol.* 28, 101–142.
- [10] Georginis, S. and Subramanian, A.R. (1980) *J. Mol. Biol.* 141, 393–408.
- [11] Boni, I.V., Zlatkin, I.V. and Budowsky, E.I. (1982) *Eur. J. Biochem.* 121, 371–376.
- [12] Subramanian, A.R. (1984) *Trends Biochem. Sci.* 9, 491–494.
- [13] Boni, I.V., Isaeva, D.M., Musychenko, M.L. and Tzareva, N.V. (1991) *Nucleic Acids Res.* 19, 155–162.
- [14] Hartz, D., McPheeters, D.S., Traut, R. and Gold, L. (1988) *Methods Enzymol.* 164, 419–425.
- [15] Makhno, V.I., Peshin, N., Semenov, Yu.P. and Kirillov, S.V. (1988) *Mol. Biologiya* 22, 670–679 (Russ.).
- [16] Schlessinger, D., Mangiarotti, G. and Apirion, D. (1967) *Proc. Natl. Acad. Sci. USA* 58, 1782–1789.
- [17] Jefferson, R.A., Burgess, S.M. and Hirsh, D. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8447–8451.
- [18] Brederode, F.T., Koper-Zwarthoff, E.C. and Bol, J.F. (1980) *Nucleic Acids Res.* 8, 2213–2223.
- [19] Hartz, D., McPheeters, D.S. and Gold, L. (1989) *Genes Dev.* 3, 1899–1912.
- [20] Van Dieijen, G., Van der Laken, C.J., Van Knippenberg, P.H. and Van Duin, J. (1975) *J. Mol. Biol.* 93, 351–366.
- [21] Winter, R.B., Morrissey, L., Gauss, P., Gold, L., Hsu, T. and Karam, J. (1987) *Proc. Natl. Acad. Sci. USA* 84, 7822–7826.
- [22] McPheeters, D.S., Stormo, G.D. and Gold, L. (1988) *J. Mol. Biol.* 201, 517–535.
- [23] de Smit, M.H. and Van Duin, J. (1990) *Prog. Nucleic Acids Res. Mol. Biol.* 38, 1–35.
- [24] Calogero, R.A., Pon, C.L., Canonaco, M.A. and Gualerzi, C.O. (1988) *Proc. Natl. Acad. Sci. USA* 85, 6427–6431.
- [25] Canonaco, M.A., Gualerzi, C.O. and Pon, C.L. (1989) *Eur. J. Biochem.* 182, 501–506.
- [26] Gualerzi, C.O. and Pon, C.L. (1990) *Biochemistry* 29, 5881–5889.
- [27] Farwell, M.A., Roberts, M.W. and Rabinowitz, J.C. (1992) *Mol. Microbiol.* 6, 3375–3383.