Volume 10 Number 20 1982 Nucleic Acids Research

Gene expression: chemical synthesis of *E. coli* ribosome binding sites and their use in directing the expression of mammalian proteins in bacteria

Ernest Jay, Arun K.Seth, Johanne Rommens, Ashwani Sood and Gilbert Jay[+]

Department of Chemistry, University of New Brunswick, Bag Service #45222, Fredericton, New Brunswick, Canada E3B 6E2, and [+]Laboratory of Molecular Virology, National Cancer Institute, National Institutes of Health, Bethesda, MD 20205, USA

Received 14 September 1982

ABSTRACT

Mammalian genes, when inserted into bacterial plasmid or phage DNAs, will not be expressed into the corresponding specific proteins in E. coli unless proper initiation signals required for recognition by E. coli ribosomes are provided. We have studied these signals and chemically synthesized two DNA duplexes each containing different initiation signals. These have been inserted in front of the Simian virus 40 (SV40) small tumor antigen gene (SV40 t gene) at varying distances from the ATG initiation codon prior to its cloning into pBR322 plasmid DNA. Plasmid containing clones carrying either of these two synthetic ribosome binding sites (RBS) at varying distances from the SV40 t gene all produced a 17K protein identical to authentic t antigen by immunologic, electrophoretic and proteolytic digestion analyses. This provides a novel method to ensure the specific expression of any contiguous mammalian gene to be cloned into bacteria, and also a unique in vivo method for studying the structure-function (efficiency) relationship of RBS with specific base changes.

INTRODUCTION

The study of the mechanisms of control in gene expression and especially those in the initiation processes of transcription and translation, has been one of the main thrusts of molecular biology. The recent advances in the technology of molecular cloning and the desire to develop expression vectors allowing efficient production of biologically or clinically important mammalian proteins in bacteria carrying the appropriate mammalian gene insert has further intensified research in this area.

The exact nature of the promoter required for recognition by E. coli RNA polymerase to properly initiate RNA synthesis is thus far not fully defined, apart from the two more or less conserved TATAAT and TTGACA sequences located around the -10 and -35 regions. Foreign DNA when appropriately inserted into bacterial cloning vectors will be faithfully transcribed into mRNA sequences. However early attempts in translation resulted in the production of hybrid protein products and failed to produce specific mammalian proteins. The most likely explanation is that eukaryotic mRNA contains quite different signals (5) and thus lacks the appropriate initiation signals for recognition by prokaryotic ribosomes (6).

In order to ensure the proper initiation of translation of a mammalian gene

0305-1048/82/1020-6319$ 2.00/0

cloned into E. coli we propose to ligate a chemically synthesized DNA duplex in front of the gene. This duplex corresponds to an E. coli ribosome binding site and contains all the necessary initiation signals required to specify protein synthesis. Although nucleotide sequence analysis of bacterial mRNAs and phage mRNAs have provided convincing evidence for the existence of some preferred sequences (recognition elements) immediately to the 5' side of the initiation codon for recognition by prokaryotic ribosomes (for reviews see Ref. 1, 2 and 7), the exact nature of these recognition elements has yet to be fully elucidated. Several more or less conserved features in this region have been recognized in addition to the AUG initiation codon which may play a direct role in the initiation process of translation. These include (1) All or part of the polypurine TAAGGAGGT sequence known as the Shine-Dalgarno (S/D) sequence which is complementary to the 3' end of the 16s ribosomal RNA and can occur at varying distances from AUG (8, 9). (2) One or more termination codons: The absolute requirement of a termination codon at the RBS for initiation has not been clearly demonstrated. Its presence may function to ensure the dissociation of translating ribosomes from the preceding adjacent cistron to allow proper initiation. (3) A RRUUURR sequence: This sequence has been found in part or in full, either in addition to or in the absence of the S/D sequence. The almost exclusive occurrence of this sequence among RBS of phage capsid proteins ($\phi\chi$, fd, Qβ, R17, MS2 and f2) and ribosomal proteins (L11, L12 and S12) suggests a functional role in the initiation of synthesis to high levels in the cell. In addition to the wide variability in the recognition sequences among the different ribosome binding sites, the question of how far the recognition site extends into the 5' region and whether part of the coding sequence adjacent to the AUG codon is required for the initiation process has not as yet been definitively answered.

Two DNA duplexes containing different, specific recognition elements have been synthesized. These have been ligated in front of the Simian virus (40) small tumor antigen (SV40 t antigen) at varying distances from the ATG initiation codon prior to its insertion into the single Pst site in pBR322 plasmid DNA. Thus initiation of transcription at the β-lactamase promoter will produce a chimeric mRNA containing the SV40 t antigen mRNA coding sequence preceded by the synthetic RBS site and flanked by β-lactamase coding sequences. E. coli transformants containing either of the synthetic RBS together with the SV40 t antigen gene produced authentic SV40 t antigen. The choice of the SV40 t gene for our model studies and the results from the use of one of the RBS has previously been reported in a preliminary communication (10).

MATERIALS AND METHODS

The four deoxyribonucleosides were purchased from Calbiochem and checked by

HPLC before use. The four fully protected deoxyribonucleotides were prepared according to published procedures (11, 12). Restriction enzymes, Exonuclease III, S1 nuclease, DNA ligase and polynucleotide kinase were purchased from Bethesda Research Laboratories. SV40 DNA and pBR322 plasmid DNA were prepared according to published procedures (13, 14).

Synthesis of oligodeoxyribonucleotides - The synthesis of the oligodeoxyribonucleotides was carried out using the 'Triester Method' (15, 16) with a mixture of triisopropylbenzenesulfonyl chloride (TPSCl) and 1H-tetrazole as condensing reagent (16). The isolated yields of each step ranged between 60-90% and the products were purified by preparative TLC using 10% methanol in chloroform as solvent.

Purification of the final product - The final products were purified on silica gel TLC. The material was then treated with concentrated ammonium hydroxide in a sealed tube for 4 hours at room temperature followed by another 4 hours treatment at 50°. The tritylated material was purified on silica gel TLC using a mixture of isopropanol:NH_4OH:H_2O in the ratio of 55:10:35 as the developing solvent. The tritylated compound, present as the major band, was extracted using 2M ammonium hydroxide. After lyophilization, detritylation was performed by treatment with 80% acetic acid (0.5 ml) at 4° for 20 minutes. The acetic acid was removed by evaporation under reduced pressure. The fully deblocked product was then applied to a G50 sephadex (0.6 x 90 cm) column and eluted with a 10 mM triethylammonium bicarbonate solution. The first peak which usually contained more than 80% of the total applied material contained pure product. Purity was demonstrated by gel electrophoresis analysis after 5'-labeling using [^{32}P]γATP and T4 polynucleotide kinase.

Characterization of the synthetic oligomers - The oligomers were characterized using the 'wandering spot' technique (17, 18) according to published procedures (18, 19) and the 'Maxam and Gilbert' base-specific chemical cleavage method (20) which was modified to allow suitable cleavage of small oligomers. G cleavage: 20 min at 20° with dimethylsulfate; A/G cleavage: 80 min at 45° with pyridine-formic acid; T/C cleavage: 30 min at 45° with hydrazine; C cleavage: 30 min at 45° with hydrazine in NaCl. In the hydrazine cleavage reaction, after the first ethanol precipitation, the 0.3 M sodium acetate (NaOAc) solution was treated with 20 µl of acetylacetone for 5 min. Ethanol was then added for the second precipitation. This procedure using acetylacetone to remove the hydrazine prior to heating with piperidine prevents the possibility of non-specific cleavages.

Controlled ExoIII/S1 hydrolysis of duplex DNA - The entire SV40 t gene lies within the Hind(II + III)-A fragment. The ATG initiation codon lies only 12 nucleotides from the terminus at the A-C junction of the fragment (21, 22). In order to attach the synthetic RBS at varying distances from the ATG codon, the termini of the SV40-Hind A fragment were randomized by treatment with Exonuclease

III followed by S1 nuclease (23, 24) to remove 1 to 12 nucleotides from each end. 1 μg of SV40 DNA, previously digested with a mixture of Hind(II + III), was incubated at 25° in 15 μl containing 50 mM Tris-HCl pH 7.8, 90 mM NaCl, 4 mM DTT, 4 mM MgCl and 2 units of Exonuclease III. Aliquots of 2 μl samples were withdrawn up to 20 min and were heated at 70° for 2 min to inactivate the enzyme. The aliquots were recombined and the volume was readjusted to 15 μl. The reaction mixture was heated to 70° for 1 minute and allowed to slowly cool to 10° in a sealed eppendorf tube. To this was added 1.8 μl of a 10 times S1 buffer containing 500 mM sodium acetate pH 4.5, 60 mM $ZnSO_4$, 500 mM NaCl followed by 3 units of S1 nuclease in 1 μl. The mixture was incubated at 10° for 30 min. The reaction was terminated by addition of EDTA to 20 mM. The volume was brought up to 200 μl containing 0.3 M NaOAc and extracted with phenol and precipitated with ethanol. The partially hydrolyzed SV40-Hind A band was isolated after separation on a 4% polyacrylamide gel. The controlled activities of the ExoIII and S1 hydrolysis were assayed by a simple method (unpublished).

The ligation of DNA fragments and the transformations of the *E. coli* LE 392 using the recombinant plasmids were performed as previously described (10). The characterization of the SV40 t antigen produced by the transformants and the sequence analysis of the producer clones around the RBS and the SV40 t gene insert was carried out as previously described (11).

RESULTS

Strategy and results for the synthesis of the 16/20 RBS duplex - Since the entire SV40 t gene lies within the Hind(II + III) A fragment with the initiation codon only 12 nucleotides away from the terminus (Hind III cleavage site), a simple method for the specific attachment of a synthetic RBS would be to synthesize a RBS containing a Hind III cleavage site adjacent to the initation signals. A Pst I cleavage site on the opposite end would allow insertion into the Pst I site in pBR322 DNA. The following 12/20 duplexed RBS was synthesized using the Triester Method.

```
                        12
met pro ala                             met asp lys
-atgcctgcaTAAGGAGGTTTAagcttgcaaagatggataaa-
-tacggACGTATTCCTCCAAATTCGAacgtttctacctattt-
pBR322          20                SV40
```

In addition to these restriction sites, the duplex contains the following features: (1) a 7-base AAGGAGG S/D sequence that corresponds to the longest such polypurine sequence found in known prokaryotic mRNA. Furthermore, the two flanking T residues are also complementary to corresponding 16s rRNA sequence. (2) A GGTTTAA sequence corresponding to the RRUUURR sequence and (3) A TAA termination codon which will be in frame with the β-lactamase coding sequence upon insertion into the Pst I site in pBR322 DNA. This will ensure dissociation of the 70S

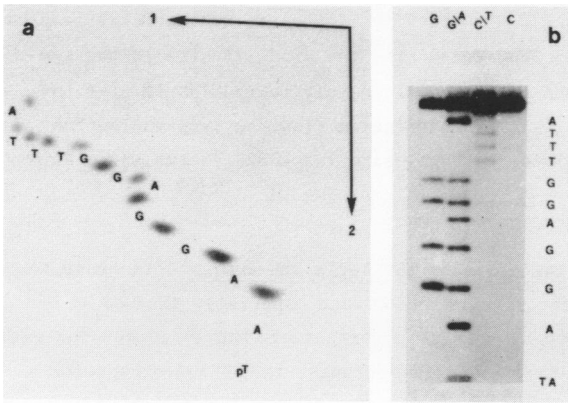

Fig. 1. Sequence analysis of the TAAGGAGGTTTA dodecamer for the upper strand of the 12/20 RBS duplex using (a) the 2D 'wandering spot' and (b) the 'Maxam and Gilbert' chemical cleavage techniques.

ribosome translating the β-lactamase sequence and allow initiation at the synthetic RBS for the SV40 t gene.

The syntheses of the two oligomers were accomplished by consecutive condensations of suitably protected trinucleotide and tetranucleotide blocks with isolated yields ranging from 60-75%. Their sequences were verified by sequence analyses shown in Figures 1 and 2.

When the 12/20 RBS duplex was attached to the front of the SV40-Hind A fragment and the resulting 'translational unit' was inserted into pBR322 DNA and cloned into E. coli LE 392, none of the ampicillin sensitive clones analyzed produced detectable amounts of t antigen. It was assumed that this was because the initiation signals in the synthetic RBS were too far away from the AUG initiation codon. The SV40-Hind A fragment was thus first treated with ExoIII nuclease to remove 1-8 nucleotides from the 3' ends of both strands followed by removal of the single stranded regions by S1 nuclease to give a mixture of even ended fragments. This resulted in randomizing as well as shortening the distance between the ATG initia-

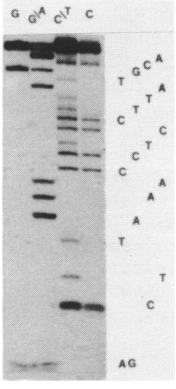

Fig. 2. Sequence analysis of the AGCTTAAACCTCCTTATGCA icosamer for the lower strand of the 12/20 RBS duplex.

tion codon and the initiation signals. In order to ligate the synthetic RBS duplex to these randomized, and even ended SV40 fragments, the Hind III site in the synthetic duplex was repaired using DNA polymerase I and all four dNTPs to give the 16/20 RBS duplex. It will contain the Pst restriction cleavage tetranucleotide sticky end sequence for insertion into the Pst I site in pBR322 DNA as shown below.

TAAGGAGGTTTAagct
ACGTATTCCTCCAAATTCGA

Strategy and results for the synthesis of the 11/15 RBS duplex - In order to study the functions of the different signals in RBS and especially that of the RRUUURR sequence, we have decided to synthesize a variant of the 16/20 RBS duplex. This variant is also identical to the RBS for β-lactamase in pBR322 except for one base change shown by the arrow.

β-lactamase	AAAAGGAAGAGTatg
	↓ met
11/15 RBS duplex	TAAGGAGGAGT
	ACGTATTCCTCCTCA

In comparison with the 16/20 RBS duplex, both contain (1) the Pst I cleavage site sticky end sequence, (2) the identical TAAGGAGG S/D sequence and (3) the TAA termination codon which is similarly strategically placed to be in phase with the β-lactamase reading frame. The only major difference is the omission of the GGTTTAA sequence in the 11/15 RBS duplex. This 11/15 RBS duplex can be directly ligated to the SV40-Hind A fragment which has been treated with ExoIII and S1 nucleases.

Construction of recombinant plasmids - The construction of the recombinant plasmids was carried out using a novel strategy to give highly specific ligation reactions. The top 11-mer strand of the 11/15 RBS duplex was first ligated to the even ended SV40 fragment (step a) in Figure 4 using the 11/15 duplex which was hydroxylated on the 5' ends of both strands. Using a 50 fold excess of the 11/15

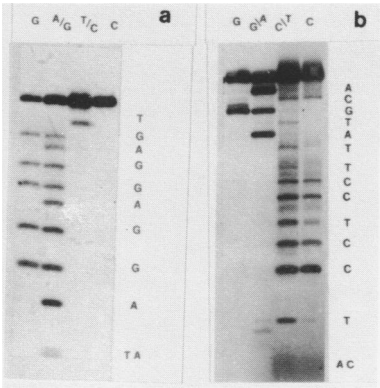

Fig. 3. Sequence analysis of the TAAGGAGGAGT undecamer for the upper strand in (a) and the ACTCCTCCTTATGCA pentadecamer for the lower strand in (b) of the 11/15 RBS duplex. It should be noted that the hydrazine cleavages in (b) gave significant amounts of non-specific cleavage products. These cleavages were carried out without the addition of acetylacetone after the first precipitation step as described in the Methods section. It thus appears that the complete removal of trace amounts of hydrazine by reaction with acetylacetone prior to heating at 85° with piperidine gives improved results.

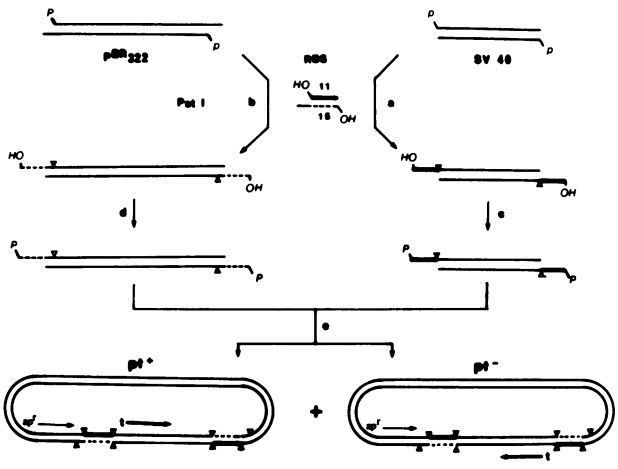

Fig. 4. Strategy for the construction of the recombinant plasmid containing the 11/15 RBS duplex. ap^r represents the direction and start point of the ampicillin resistance (β-lactamase) gene and the start point of the SV40 t antigen gene.

duplex, ligation of the 11-mer to the SV40 fragment was quantitative and suppressed all self joinings or oligomerizations. Similarly, the lower 15-mer strand was specifically joined to the Pst-linearized pBR322 (step b). Ligation of the 15-mer to the pBR322 results in an incomplete Pst I site and the product is not cleavable by Pst I. Thus the inclusion of Pst I enzyme during this ligation step would convert any pBR322 DNA that had oligomerized or circularized back into the linear form for ligation to the RBS oligomer resulting in a quantitative specific ligation. After removal of the excess oligomers, by purification on polyacrylamide gel, and phosphorylation of the 5' termini using T4 polynucleotide kinase, the two fragments were annealed and ligated (step e) to give the two recombinant plasmids, pt^+ and pt^-. pt^+, in which the SV40 t gene is oriented colinearly with the β-lactamase gene, is expected to express t antigen. pt^-, in which the t gene is oppositely oriented to the β-lactamase gene, will not give the correct coding mRNA sequence when transcribed from the β-lactamase promoter. Consequently it is not expected to express t antigen. Since the two strands of the RBS are not palindromic, no intra-ligation occurred resulting in very specific ligation between the pBR322 containing and SV40 containing fragments. The 16/20 duplex was similarly ligated as was previously described (10).

Screening of recombinant clones and detection of producer clones - After transformation of E. coli LE 392, the tetracycline resistant and ampicillin sensitive clones (Tc^rAp^s) were selected for further screening for the presence of SV40 specific sequences by hybridization with ^{32}P-labeled SV40 DNA. More than 90% of

the $Tc^{r}Ap^{s}$ clones contained SV40 sequences and, as expected, about half (42% for the 16/20 duplex and 45% for the 11/15 duplex) expressed detectable amounts of t antigen. The characterization of the t antigen product has been previously described (10). There was at least a 10 fold difference between the high producer and the low producer clones.

Sequence analysis of the recombinant DNA - The RBS-SV40 insert lies within the J fragment of the pBR322-Hpa II cleavage map and contains a single Taq I site. Digestion of the DNA from each of the clones with Hpa II enzyme produced a 1300 base pair fragment that contained the entire RBS-SV40 insert with the RBS located 53 nucleotides from one terminus. The fragment from each 16/20 producer clone was terminally labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase and cleaved with Taq I enzyme to generate two fragments. The smaller fragment (about 496 base pairs) was isolated from a polyacrylamide gel and sequenced using the 'Maxam and Gilbert' technique. The sequence at the region of insertion for one of these clones is shown in Figure 5. Those clones that have from one to two nucleotides removed from the SV40 sequence produced much lower quantities of t antigen as compared to those that have three or four nucleotides removed. The t antigen producers identified thus far for the 16/20 duplex containing clones are summarized in Figure 6.

Since the expected sequences except the few nucleotides between the RBS and the ATG initiation codon for the t gene in the SV40 fragment of the DNA clones are

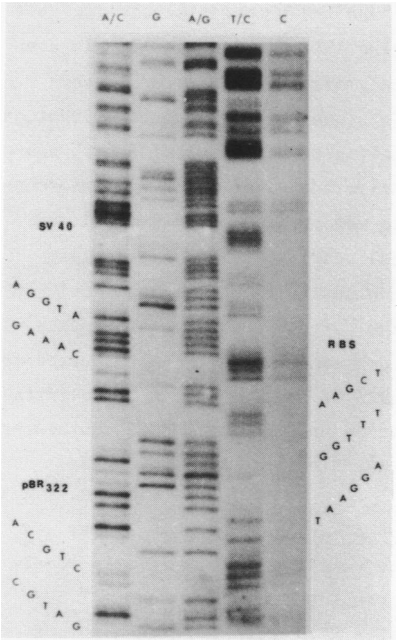

Fig. 5. Sequence analysis of a 16/20 RBS duplex containing SV40 t antigen producer clone using the 'Maxam and Gilbert' chemical cleavage technique. The region of insertion including the RBS flanked by part of the β-lactamase coding sequence and part of the SV40 sequence is shown. Three nucleotides (TTG) have been removed from the SV40 sequence in this clone.

Fig. 6. Nucleotide sequence around the RBS of the 16/20 RBS producer clones. The region between the arrows represents the deletions in the various classes of clones identified.

known, their complete sequence determination around the RBS can be accomplished using only the (A + G) and (T + C) reactions. Figure 7 shows the sequence analysis of two of the 11/15 RBS containing clones. The t antigen producers identified thus far for the 11/15 RBS duplex containing clones are shown in Figure 8.

While all of these classes of the 11/15 RBS containing clones produced significant amounts of authentic t antigen, preliminary results show none of these produced as much t antigen as the high producers in the 16/20 RBS clones despite the smaller distance between the S/D sequence and the ATG codon in these clones. More extensive studies to compare the efficiency of each class for both RBS containing clones in producing t antigen is being carried out in both *in vivo* and in cell-free transcription-translation coupled systems.

DISCUSSION

Using the gene that codes for the SV40 small tumor (t) antigen as a model, we have developed a novel method for the expression of mammalian genes in bacteria. One other procedure that will allow specific expression of eukaryotic gene products in bacteria free of bacterial sequences makes use of an *E. coli* DNA fragment which bears the promoter and the S/D sequence of the lac operon (24). In our approach, short DNA duplexes which contain certain initiation signals which are believed to be required for recognition by *E. coli* ribosomes to initiate translation (6) are chemically synthesized and appropriately attached in front of the eukaryotic gene

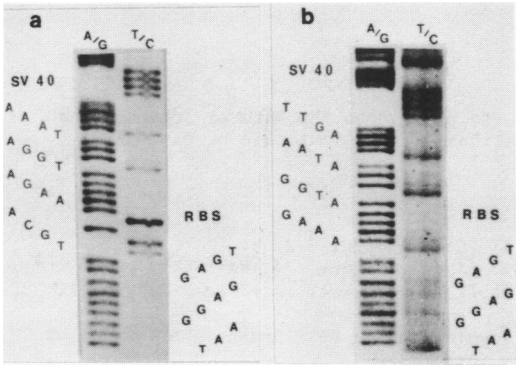

Fig. 7. Sequence analysis of two of the 11/15 RBS duplex containing SV40 t antigen producer clones. In the clone shown in (a) the entire Hind III sticky end sequence plus a single T nucleotide has been removed from the SV40 fragment by the Exo III/S1 nuclease treatment. In (b) the entire Hind III sticky end plus the TTGC tetranucleotide has been removed. A significant difference in the hydrazine reaction is observed when performed with and without the addition of acetylacetone as shown in (a) and (b), respectively.

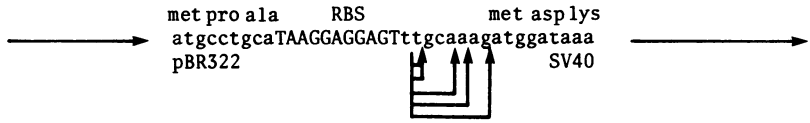

Fig. 8. Nucleotide sequence around the RBS of the 11/15 RBS producer clones. The region between the arrows represent the deletions in the various classes of clones identified.

before its insertion into bacterial cloning vectors. This provides proper signals which the eukaryotic gene lacks to guarantee specific translation at the proper initiation AUG codon. The two synthetic RBS duplexes which have been synthesized and attached in front of the SV40 t gene directed the efficient expression of the SV40 t antigen when cloned into pBR322 in E. coli. This provides a general method for the expression of any contiguous mammalian gene in E. coli and an in vivo system for studying the structure-function relationship of RBS containing different combination of initiation signals. In the results of our studies there is strong support that only the initiation codon and the region of 1-15 nucleotides upstream from it are essential to initiate protein synthesis in natural bacterial messenger RNA. The importance of the location of each set of the RBS sequence relative to the translation initiation codon is also confirmed (24, 25). Further, our results using the two RBS suggest that the presence of a RRUUURR sequence enhances the efficiency of the site to initiate protein synthesis. None of the 11/15 RBS containing clones that lack the equivalent RRUUURR signal when compared to the 16/20 RBS containing clones produced as much SV40 t antigen in preliminary studies. More vigorous studies are presently being carried out in both in vivo and in cell-free systems to more critically evaluate the efficiency of these two RBS to initiate translation. It is hoped that by these studies a better understanding of the molecular mechanism for the control of translation in bacteria may be gained and the most efficient system for the expression of biochemically and clinically important proteins in E. coli is achieved.

ACKNOWLEDGEMENTS

This work was supported by grants (to E.J.) from the Natural Sciences and Engineering Research Council and the National Cancer Institute of Canada.

REFERENCES

1. Kozak, M. and Shatkin, A. J. (1978) Cell 13, 201-212.
2. Barralle, F. E. and Brownlee, G. G. (1978) Nature 274, 84-87.
3. Grunberg-Manago, M. and Gros, F. (1977) Prog. Nucl. Acid Res. Mol. Biol. 20, 209-284.
4. Steitz, J. A. (1979) Biological Regulation and Development (Goldberger, R. F., ed.) pp. 349-399, Plenum Press, New York.

5. Kazak, M. (1978) Cell 15, 1109-1123.
6. Jay, E., Seth, A. K. and Jay, G. (1980) J. Biol. Chem. 255, 3809-3812.
7. Strerer, G. F. E., Walkinshaw, M. D., Arnott, S. and Morre, D. J. (1980) Nucleic Acids Res. 8, 3895-3907.
8. Shine, J. and Dalgarno, L. (1975) Nature 254, 34-38.
9. Steitz, J. A. and Jakes, K. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 4734-4738.
10. Jay, G., Khoury, G., Seth, A. and Jay, E. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 5543-5548.
11. Shaller, H., Weimann, G., Lerch, B. and Khorana, H. G. (1963) J. Am. Chem. Soc. 85, 3821.
12. Narang, S. A., Brousseau, R., Hsiung, H. M. and Michniewicz, J. J. (1980) Methods in Enzymology, Vol. 65, pp. 610-619, Academic Press, New York.
13. Wu, R., Jay, E. and Roychoudhury, R. (1976) Methods Cancer Res., Vol. 12, pp. 87-176, Academic Press, New York.
14. Bolwar, F., Rodriquez, R. L., Green, P. J., Betlach, M. C., Heyneker, H. L. and Boyer, H. W. (1977) Gene 2, 95-113.
15. Itakura, K., Katagiri, N., Narang, S. A., Bahl, C. P., Mariano, K. J. and Wu, R. (1975) J. Biol. Chem. 250, 4592-4600.
16. Seth, A. and Jay, E. (1980) Nucleic Acids Res. 8, 5445-5459.
17. Sanger, F., Donelson, J. E., Coulson, A. R., Kossell, H. and Fisher, D. (1974) J. Mol. Biol. 90, 315-333.
18. Jay, E., Bambara, R., Padmanabhan, R. and Wu, R. (1974) Nucleic Acids Res. 1, 331-353.
19. Tu, C. D., Jay, E., Bahl, C. P. and Wu, R. (1976) Anal. Biochem. 74, 73-93.
20. Maxam, A. and Gilbert, W. (1980) Methods in Enzymology, Vol. 65, pp. 499-559.
21. Reddy, V. B., Thummappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. and Weissman, S. M. (1978) Science 200, 494-502.
22. Fiers, W., Contreras, R., Haegeman, G., Rogier, R., Van de Voorde, A., Van Heuverswyn, H., Van Herrweghe, J., Volckaert, G. and Ysebaert, M. (1978) Nature (London) 273, 113-120.
23. Wu, R., Ruben, G., Siegel, B., Jay, E., Spielman, P. and Tu, C. D. (1976) Biochemistry 15, 134-140.
24. Taniguchi, T., Guarenta, L., Roberts, T. M., Kimelan, D., Douhan, J. and Ptashne, M. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5230-5233.