Proc. Natl. Acad. Sci. USA
Vol. 78, No. 9, pp. 5543–5548, September 1981
Biochemistry

Construction of a general vector for efficient expression of mammalian proteins in bacteria: Use of a synthetic ribosome binding site*

(molecular cloning/pBR322/simian virus 40 small tumor antigen)

GILBERT JAY†, GEORGE KHOURY†, ARUN K. SETH‡, AND ERNEST JAY‡

†Laboratory of Molecular Virology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205; and ‡Department of Chemistry, University of New Brunswick, Fredericton, New Brunswick, Canada E3B 5A3

Communicated by Igor B. Dawid, June 15, 1981

ABSTRACT **With the premise that mRNAs transcribed in *Escherichia coli* from cloned eukaryotic DNA inserts do not possess the necessary regulatory signals for recognition by prokaryotic ribosomes, we have constructed a general plasmid vector carrying a chemically synthesized prokaryotic ribosome binding site that will ensure the efficient expression of eukaryotic proteins in *E. coli*. In addition to the regulatory signals necessary for ribosome recognition, the synthetic segment contains, at one end, a *Pst* I cleavage site which will direct its insertion to pBR322 DNA and, at the other end, a *Hin*dIII site to facilitate attachment of the passenger eukaryotic gene. Using simian virus 40 (SV40) tumor (t) antigen as a model system, we have ligated the SV40 DNA fragment containing the entire t antigen gene in tandem with the synthetic ribosome binding site to pBR322 DNA at the *Pst* I site, which lies within the coding sequence of the β-lactamase gene. Initiation of transcription at the β-lactamase promoter would produce a chimeric mRNA with the synthetic ribosome binding signals and the SV40 sequence flanked by β-lactamase coding sequences. Utilization of the synthetic regulatory signals for initiation of translation is demonstrated by the efficient synthesis, in bacterial transformants, of authentic SV40 t antigen. Excision of the entire SV40 insert by *Hin*dIII from those clones that have retained intact *Hin*dIII sites at the junction between the ribosome binding site and the SV40 sequence would allow insertion of other heterologous DNAs by using *Hin*dIII linkers. The efficient expression of any DNA insert would require that the entire coding sequence be contiguous and that its termini be randomized by treatment with exonuclease III and nuclease S1 to vary the distance between the translational initiation codon and the synthetic ribosome binding site.**

One of the main goals of recombinant DNA research is to produce significant quantities of medically and agriculturally important proteins whose genes have been cloned on bacterial plasmids. Because previous studies have established that mRNAs are translated with varying efficiencies in prokaryotic cells (for review, see refs. 1 and 2), any effort to ensure the efficient expression of cloned eukaryotic genes in bacteria must consider not only the faithful transcription of the inserted gene but also the specificity and efficiency of translation of its mRNA. Whereas the initiation of transcription presumably is dictated by the existence of a promoter (3), the initiation of protein synthesis appears to be governed, at least in part, by the availability of a ribosome binding site (RBS) (1, 2).

The exact nature of the bacterial RBS is obscure. Nucleotide sequence analysis has revealed a region, about 3–11 bases upstream from the initiation codon, that seems to be conserved among most of the bacterial and phage RNAs (4). This region, referred to as the Shine–Dalgarno (S/D) sequence, is complementary to the 3′ end of the 16S ribosomal RNA (5) and base pairs with it during the process of initiation of protein synthesis (6). The S/D sequence is generally 4–9 bases long, but the reason for this variability is unclear. Thermodynamically, one might speculate that the extent of homology is directly proportional to the efficiency of ribosome binding. In addition to sequence specificity, the location of the S/D sequence relative to the translational initiation codon is also important in determining the efficiency of initiation (7, 8).

In many cases, a second sequence, R-R-U-U-U-R-R (R = purine nucleotide), has been found, in part or in full, either in addition to or in the absence of the S/D sequence. Whereas the S/D sequence is recognized by the 16S ribosomal RNA, this R-R-U-U-U-R-R sequence may exist for recognition by the ribosomal protein S1 which has the *in vitro* property of binding to polyuridylate (9) and has been shown to be indispensable for the translation of certain natural mRNAs (10). The almost exclusive occurrence of this sequence among the RBSs of phage capsid proteins (φX, fd, Qβ, R17, MS2, f2) and ribosomal proteins (L11, L12, S12) suggests a functional role in the initiation of synthesis of those viral and cellular structural proteins that accumulate to high levels in the cell (see ref. 4).

Recently, we synthesized chemically an icosadeoxyribonucleotide containing both of these sequences and showed that this putative RBS was sufficient to induce by itself the highly efficient initiation of protein synthesis in a cell-free system (11).

Eukaryotic mRNAs do not have any common recognition sequence similar to that in prokaryotic mRNAs (12, 13) and probably mediate ribosome binding by a different mechanism (14). We therefore designed a vector for the specific and efficient translation of mRNAs transcribed in bacteria from cloned eukaryotic genes by the insertion of a chemically synthesized strong RBS at the 5′ side of the coding sequence.

As a model system for testing this vector, we chose the gene that codes for the simian virus 40 (SV40) small tumor (t) antigen for the following reasons: (*i*) there is no intervening sequence within the coding region (15) and hence the RNA transcript can be translated without the need for posttranscriptional modification; (*ii*) the gene product can be easily detected and analyzed both quantitatively and qualitatively by using well-defined anti-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: SV40, simian virus 40; RBS, ribosome binding site; S/D, Shine–Dalgarno sequence; R, purine nucleotide; t antigen, small tumor antigen; Tc^R, tetracycline-resistant (resistance); Ap^S, ampicillin sensitive (sensitivity).

* NOTE. This paper runs more than 5 pages because the Editorial Office did not detect the excessive length in time for the authors to have a chance to correct it.

sera (16); and (*iii*) if made in large quantities, the protein would be useful for studies directed at an understanding of its potential role in malignant transformation of cells.

Our results show that the synthesis of authentic SV40 t antigen in bacteria can be induced by the insertion of the synthetic RBS. This approach can be used for the expression of any eukaryotic protein whose RNA transcripts do not require post-transcriptional modifications specific to eukaryotic cells, or whose complementary DNA has been obtained by reverse transcription of its mRNA. Our study also shows that the synthetic RBS contains all the control elements required to specify the initiation of protein synthesis *in vivo* and offers a system for the study of the molecular basis for differences in efficiency of ribosome binding through the use of synthetic RBSs with specific nucleotide substitutions and deletions (11).

EXPERIMENTAL PROCEDURES

Construction of Recombinant Plasmids. Repair of the 12/20 duplex by DNA polymerase I was carried out as described (17), in the presence of [α-^{32}P]TTP, and the labeled product was purified in an 18% polyacrylamide/urea gel. Randomization of the SV40 *Hin*dII/*Hin*dIII fragment A was performed by using exonuclease III and nuclease S1 (18). Conditions for end labeling with ATP and T4 polynucleotide kinase have been described (17). Ligation of the synthetic duplex to DNA fragments was as described (19).

Analysis of t Antigen from Bacterial Clones. Extracts from clones labeled with [^{35}S]methionine were immunoprecipitated with either normal hamster serum or anti-SV40 tumor antiserum (20). Immunoprecipitates were analyzed by electrophoresis on NaDodSO$_4$/polyacrylamide gels (20). Two-dimensional tryptic peptide analysis (21) and isoelectric focusing (22) were carried out as described.

RESULTS

Cloning Strategy. The purpose of this study was to test the assumption that, when correctly positioned upstream from the SV40 *Hin*dII/*Hin*dIII fragment and cloned at the *Pst* I site in pBR322 DNA, a chemically synthesized bacterial ribosome recognition site will direct the efficient synthesis of the SV40 t antigen in *Escherichia coli* carrying this recombinant plasmid. Because insertion is at the *Pst* I site, which lies within the coding sequence of the β-lactamase gene (23), transcription of the insert will utilize the β-lactamase promoter to give a chimeric mRNA, reading from β-lactamase sequences through the synthetic RBS into SV40 sequences. It is predicted that ribosomes that have initiated the synthesis of β-lactamase will terminate translation once they traverse into the synthetic RBS, due to the presence of a termination codon that is strategically positioned in phase with the β-lactamase coding sequence. At the same time, ribosomes can independently initiate in this region, as with a polycistronic mRNA, allowing the translation of SV40-specific sequences. Because the SV40 insert contains the entire coding sequence with the appropriate initiation and termination codons (24, 25), the entire SV40 t antigen should be synthesized.

Construction of Plasmids. A small DNA duplex (Fig. 1) consisting of a 12-unit strand and a 20-unit strand has been synthesized (26) and the nucleotide sequences have been confirmed by established methods (27). The sequence representing the hypothetical RBS is located within the sense strand (12 units) and contains the following features: a 9-base S/D sequence (I), which represents the longest such sequence that has been found in any known prokaryotic mRNA; a heptanucleotide (II), corresponding to the R-R-U-U-U-R-R sequence, which is present in many bacterial and phage RNAs; a TAA termination

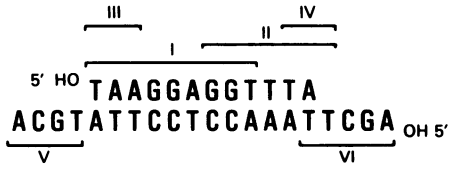

FIG. 1. Nucleotide sequence of the chemically synthesized RBS. The upper sequence represents the sense strand and the lower, the nonsense strand. The various control elements which make up this hypothetical recognition site are discussed in the text.

codon (III) which will be in frame with the β-lactamase coding sequence upon insertion at the *Pst* I site in pBR322 DNA; and a second TAA termination codon (IV) which is in a different reading frame than the first. The complementary strand (the nonsense strand) contains at its protruding 3' terminus a tetranucleotide sequence (V) which will facilitate and direct its ligation at the *Pst* I site in pBR322 DNA; at its protruding 5' terminus there is a *Hin*dIII sequence (VI) which will allow the resection of the SV40 insert to yield a general vector for the expression of other DNA inserts (see *Discussion*).

The 1169-base-pair fragment A obtained by digestion of SV40 DNA with *Hin*dII/*Hin*dIII contains the entire coding sequence for the t antigen. The initiating AUG codon is 12 nucleotides from the 5' end of the sense strand and the terminating UAA codon is 632 nucleotides from the 3' end (24, 25). In order to randomize the distance between the synthetic RBS and the initiating AUG codon,[§] the SV40 *Hin*d fragment A was first treated with exonuclease III followed by nuclease S1, under conditions determined to remove from 0 to about 10 nucleotides from both ends. In addition to randomizing the location of the ATG from the 5' end of the DNA fragment this procedure also converted the staggered ends of the DNA to blunt ends (Fig. 2, step *a*).

In order to provide an efficient and specific route for the insertion of the SV40 sequence containing the synthetic RBS into the *Pst* I site of pBR322 DNA, a novel ligation strategy was devised. The protruding *Hin*dIII sequence present in the synthetic 12/20 duplex was first repaired by DNA polymerase I to give a 16/20 duplex composed of a repaired 16-mer and the original 20-mer (Fig. 2, step *b*). This 16/20 duplex was then ligated to the flush ends of the SV40 fragment (Fig. 2, step *c*). Because only one end of the 5'-hydroxylated 16/20 duplex is blunt, only the 16-mer (the sense strand) can be ligated to the two 5'-phosphorylated ends of the SV40 fragment. The noncovalently associated 20-mer was then removed from the complex by denaturation and purification. It is worth pointing out that, in the presence of a 10-fold molar excess of the synthetic duplex during ligation, self-oligomerization of the SV40 fragment was completely suppressed and quantitative tailing of the SV40 fragment with the 16-mer was obtained as shown by a distinct shift in mobility in polyacrylamide gel (data not shown). Each end of the SV40 fragment contained a single 16-mer.

Tailing of *Pst* I-linearized pBR322 DNA with the synthetic 20-mer was achieved by using the 12/20 duplex (Fig. 2, step *d*). Because *Pst* I cleavage generates receding 5'-phosphates and the 12/20 duplex contains a protruding 3' terminus which is complementary to the *Pst* I cleavage site (Fig. 1, V), only the 20-mer (the nonsense strand) can be ligated to the two 5' ends of the pBR322 DNA. The noncovalently associated 12-mer was then removed from the complex by denaturation and purifi-

[§] Direct ligation of the synthetic RBS to the SV40 *Hin*d fragment A and insertion at the *Pst* I site of pBR322 DNA has failed to produce detectable amounts of t antigen in the transformants. It was assumed that the RBS was too far from the initiation ATG codon.

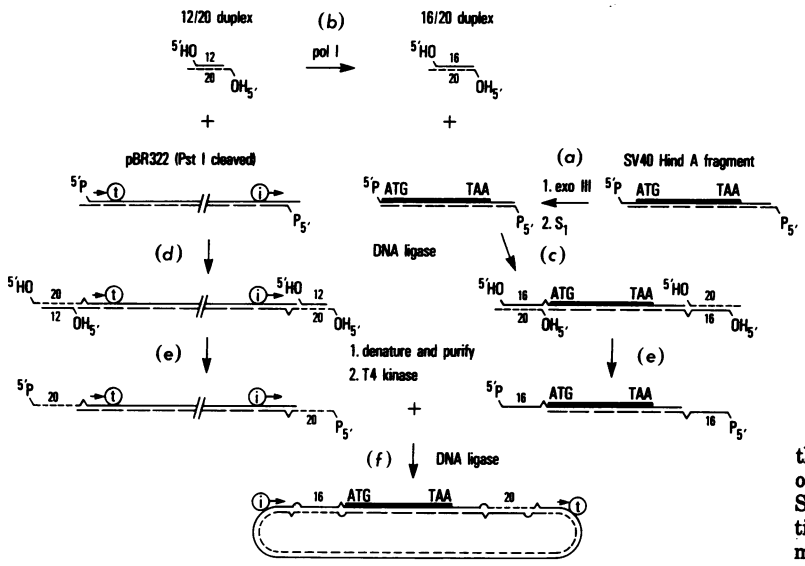

FIG. 2. Schematic representation of the construction of the recombinant plasmids. See text for detailed explanation of each reaction. The location of the coding sequence for the SV40 t antigen is indicated by a thick line. The sites for initiation i and termination t of transcription of the β-lactamase gene on pBR322 DNA are also indicated.

cation. Again, each end of the pBR322 DNA contained not more than one 20-mer. An initial attempt to ligate the 20-mer, in the absence of the complementary 12-mer, to the *Pst* I site of pBR322 DNA was unsuccessful.

The final coupling of the SV40 fragment which has been tailed at both ends with the 16-mer, to *Pst* I-linearalized pBR322 DNA which has been tailed with the complementary 20-mer was performed by using T4 DNA ligase (Fig. 2, step *f*), after phosphorylation of the 5'-hydroxyls by T4 polynucleotide kinase (Fig. 2, step *e*).

Screening of Recombinant Clones. The final ligation mixture was used to transform *E. coli* LE392. Tetracycline-resistant (Tc^R) clones were selected and tested for ampicillin sensitivity (Ap^S) (23). Ninety-three percent of the Tc^R clones were found to be Ap^S. About 250 Tc^RAp^S clones were further tested for the presence of SV40-specific sequences by hybridization with a ^{32}P-labeled SV40 DNA probe generated by nick-translation. Ninety-seven percent of these clones were found to contain a SV40-specific DNA insert. Because the pBR322 DNA could not have acquired SV40 sequences without the appropriate insertion of the synthetic DNA duplex, the finding that >90% of the Tc^R transformants carry an SV40 insert indicates the high efficiency and specificity of the ligation reactions.

Detection of Producer Clones. Twenty-six SV40-containing clones were tested for their ability to express the SV40 t antigen by immunoprecipitation of proteins from extracts of cells, labeled in the presence of [^{35}S]methionine, with a broad-spectrum anti-SV40 tumor antiserum (20). Eleven of these clones showed accumulation of detectable amounts of a M_r 17,000 protein when the immunoprecipitates were analyzed by electrophoresis on a 12.5% polyacrylamide/NaDodSO$_4$ gel (16); this suggests that, depending on the sensitivity of our technique, at least 42% of the clones accumulated a M_r 17,000 protein. Because the SV40 DNA fragment can be inserted into pBR322 DNA in either of two orientations, one would expect 50% of the clones to be in the correct orientation for the expression of t antigen (i.e., the SV40 coding sequence on the same DNA strand as the pBR322 β-lactamase sequence). These findings therefore suggest that about 85% of the expected clones are producing detectable amounts of the M_r 17,000 protein.

The majority of the clones expressed the M_r 17,000 protein, but the level of accumulation of this protein as determined by immunoprecipitation with an excess of antiserum varied dramatically. Of 11 positive clones, 8 were low producers and 3 were high producers; the high producers accumulated >10-fold higher amounts of the M_r 17,000 protein than did the lower producers. A conservative estimate is that the M_r 17,000 protein made up about 0.4% of the total protein in a high-producer cell (see *Discussion*). Examples of a nonproducer (pt22), a low producer (pt23), and a high producer (pt24) are shown in Fig. 3*A*.

Characterization of the Bacterial Product. To demonstrate

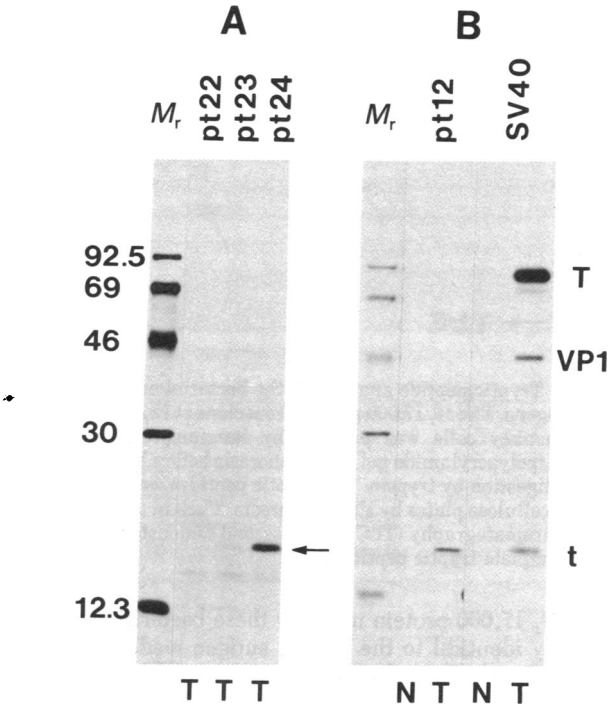

FIG. 3. Detection of SV40 t antigen produced in bacterial clones. [^{35}S]Methionine-labeled cell extracts were treated with either anti-SV40 tumor antiserum (T) or normal serum (N), and the resulting immunoprecipitates were analyzed on a 12.5% polyacrylamide/NaDodSO$_4$ gel. (A) Immunoprecipitates from a nonproducer clone (pt22), a low-producer clone (pt23), and a high-producer clone (pt24). Arrow, M_r 17,000 protein specifically precipitated by the antiserum. (B) Immunoprecipitates from clone pt12 and from SV40-infected African green monkey kidney cells. The SV40-specific proteins precipitated by the antiserum are indicated. The molecular weight markers (M_r) used were phosphorylase b (92,500), bovine serum albumin (69,000), ovalbumin (46,000), carbonic anhydrase (30,000), and cytochrome *c* (12,300).

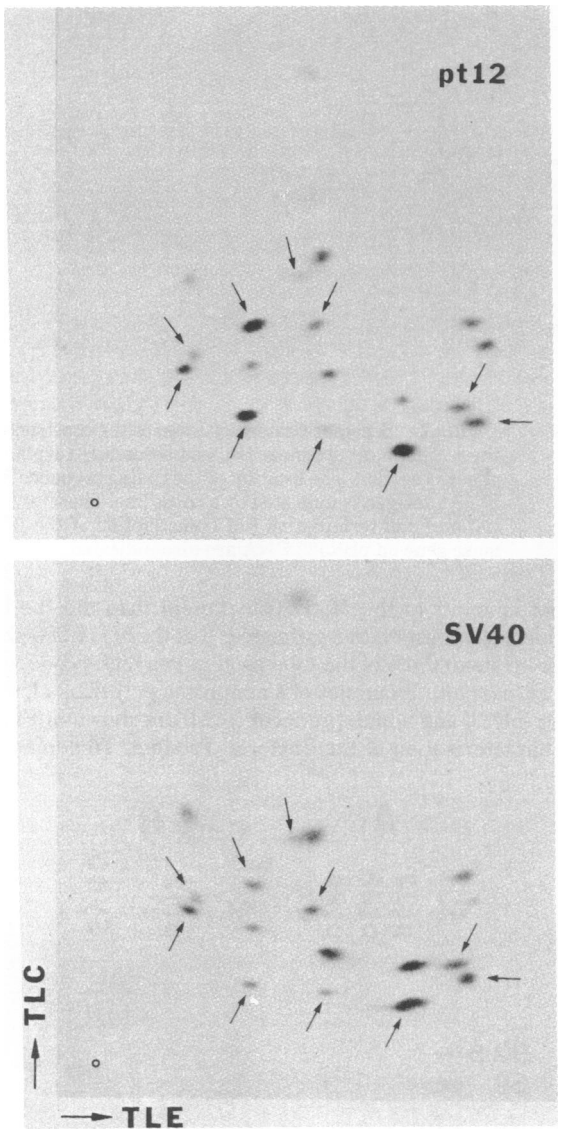

FIG. 4. Tryptic peptide analysis of the bacterial and mammalian SV40 t antigens. The M_r 17,000 protein from clone pt12 and from SV40-infected monkey cells was purified by immunoprecipitation and $NaDodSO_4$/polyacrylamide gel electrophoresis before being subjected to partial digestion by trypsin. The tryptic peptides were analyzed on thin-layer cellulose plates by electrophoresis (TLE) in the first dimension and chromatography (TLC) in the second dimension. The arrows indicate complete tryptic peptides.

that the M_r 17,000 protein made by these bacterial clones are structurally identical to the SV40 t antigen made in infected monkey cells, a high-producer clone (pt12) was selected for further analysis. Immunoprecipitation of [^{35}S]methionine-labeled pt12 extracts with anti-SV40 tumor serum gave a single M_r 17,000 protein band on $NaDodSO_4$/polyacrylamide gels which was indistinguishable in mobility from the t antigen of SV40-infected African green monkey kidney cells (Fig. 3*B*). Neither the bacterial M_r 17,000 protein nor the SV40 t antigen was precipitated by normal serum.

The identity of the bacterial and mammalian proteins was demonstrated by tryptic peptide analysis (Fig. 4). Two-dimensional separation of the digestion products showed no detectable difference between the two proteins. The presence of identical partial tryptic peptides further confirmed the structural identity between the bacterial and mammalian products and contributed to the apparent lack of molar equivalence between certain peptides from the two proteins. In addition to apparent identity in primary structures, both proteins behaved as homogeneous components without charge or size heterogeneity when analyzed in two dimensions on isoelectric focusing and $NaDodSO_4$/polyacrylamide gels (Fig. 5). These data strongly suggest that there is no detectable structural or conformational differences between the authentic t antigen from SV40-infected monkey cells and the M_r 17,000 protein synthesized in bacteria. Because the function of the t antigen remains unknown, it is not possible at present to determine if the bacterial M_r 17,000 protein possesses biological activity.

Sequence Analysis of the Recombinant DNA. To confirm the retention of an appropriately inserted RBS immediately adja-

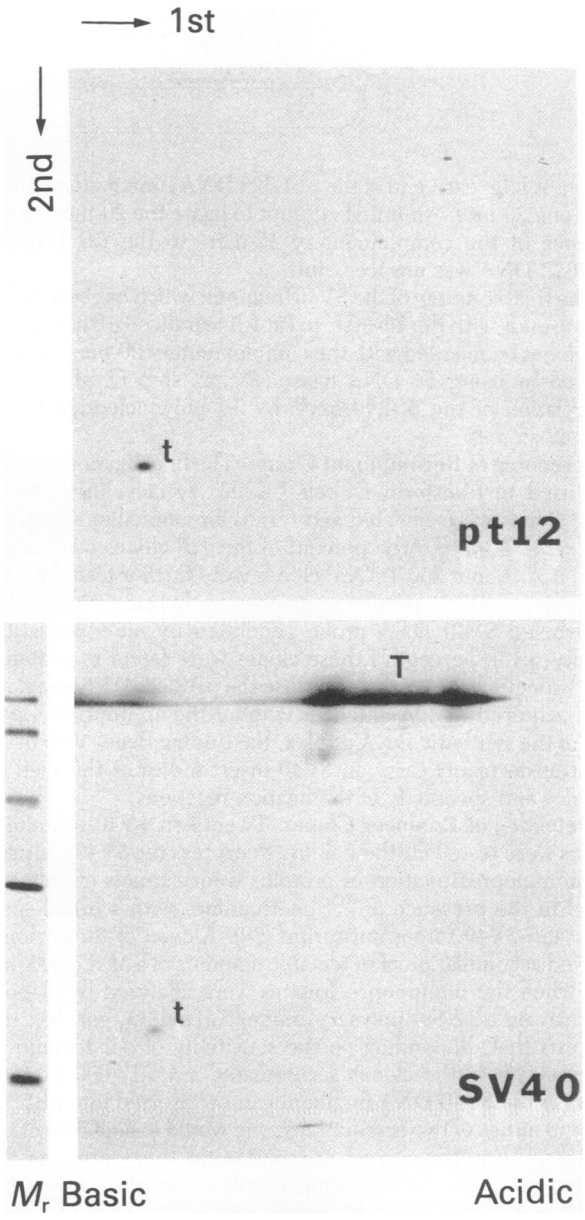

FIG. 5. Isoelectric focusing of the bacterial and mammalian SV40 t antigens. Immunoprecipitates from clone pt12 and from SV40-infected monkey cells, obtained with anti-SV40 tumor antiserum, were analyzed by isoelectric focusing in the first dimension and $NaDodSO_4$/polyacrylamide gel electrophoresis in the second dimension. The location of t antigen is indicated. Molecular weight markers (M_r) used in the second dimension were the same as in Fig. 3.

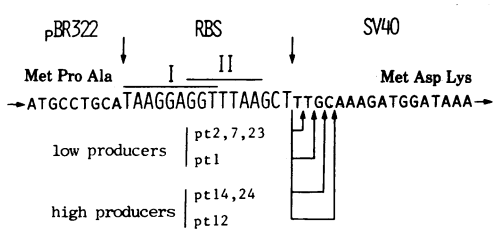

FIG. 6. Nucleotide sequences around the RBS of producer clones. Digestion of DNA from each of the clones with *Hpa* II yielded a 1300-base-pair fragment that contained the entire SV40 insert as well as the synthetic RBS located 53 nucleotides from one end. This fragment was labeled at its 5′ terminus by using [γ-^{32}P]ATP and T4 polynucleotide kinase and cleaved with *Tag* I to generate two fragments. The smaller fragment (496 base pairs) was isolated and its sequence was determined by using the method of Maxam and Gilbert (28). The regions indicated by arrows represent deletions in the various clones.

cent to the SV40 initiation codon and to obtain an insight into the molecular basis for the variability in the amount of t antigen made by different clones, the DNA sequence at the proximal recombinant junction of seven producer clones was obtained (Fig. 6). The data show clearly that all seven producer clones have retained the inserted RBS correctly positioned between the β-lactamase coding sequence located upstream and the SV40 t antigen coding sequence downstream. In all cases, the first termination codon within the inserted RBS is in phase with the β-lactamase gene.

Although the junction between the β-lactamase sequence and the RBS is identical in all of the clones, the junction between the RBS and the SV40 sequence differs. This observation is expected because ligation of the β-lactamase sequence on the left was determined specifically by the *Pst* I-generated sticky ends whereas coupling at the right was by blunt-end ligation using the SV40 *Hin*dII/*Hin*dIII fragment A whose ends had previously been randomized by combined exonuclease III and nuclease S1 treatment.

The sequences also indicate a direct correlation between the level of t antigen production by a clone and the number of deleted nucleotides at the RBS–SV40 junction. None of the producer clones retains the intact sequence. The low producers have lost one or two nucleotides and the high producers have lost three or four nucleotides.

CONCLUSIONS

One procedure that will allow expression in *E. coli* of complete eukaryotic gene products free of bacterial sequences makes use of an *E. coli* DNA fragment which bears the promoter and S/D sequence of the *lac* operon (8). When appropriately ligated to the eukaryotic insert, the *lac* promoter will ensure the transcription of the insert and the *lac*Z S/D sequence presumably will induce the binding of ribosomes (29).

Several authentic eukaryotic proteins have been synthesized by using this method (8, 30–33), but it is not necessarily an efficient system at the translational level. Bacterial and phage RNAs have been shown to be translated at varying efficiencies *in vivo* (1, 2). In order to maximize the synthesis of eukaryotic proteins in *E. coli*, one must consider not only the rate of transcription of the DNA insert but also the efficiency of translation of the RNA transcripts. Although the *lac* promoter is efficient at the level of transcription, its RBS contains only a 4-base S/D sequence for base pairing with the 3′ end of 16S ribosomal RNA. We consider this to be a weak binding site for ribosomes.

It is reasonable to postulate that, thermodynamically, the more extensive the S/D sequence the higher the efficiency of ribosome binding and the initiation of protein synthesis. This assumption apparently is supported by experimental observations (29). Furthermore, it seems likely, as we have pointed out (see Introduction), that a second conserved sequence R-R-U-U-U-R-R, also may be intimately involved in ribosome binding.

Based on these assumptions, we have chemically synthesized the DNA analog of a hypothetical strong RBS, which contains a 9-base S/D sequence together with the complete 7-base R-R-T-T-T-R-R sequence. Its insertion, together with the coding sequence for the SV40 t antigen, within the β-lactamase gene has led to the efficient expression of authentic t antigen in *E. coli*. This observation offers an alternate and novel procedure for the expression of cloned eukaryotic genes in bacteria and confirms the hypothesis that, when correctly positioned upstream from the appropriate initiation codon a synthetic RBS is necessary and sufficient for the initiation of protein synthesis. In addition, this finding offers an *in vivo* system to test the relative importance of a 4-base and a 9-base S/D sequence as well as the role of the R-R-U-U-U-R-R sequence. Ultimately, one can deduce from such studies the ideal RBS that will maximize the expression of eukaryotic proteins whose genes have been cloned by this procedure.

This study further confirms the importance of the location of the RBS relative to the translational initiation codon (7, 8). The molecular basis for the difference between high and low producers can be explained by comparison of their DNA sequences. Of seven producer clones analyzed, none has a sequence that represents the flush ends of the *Hin*d fragment A. The four low producers were found to have either one or two nucleotides deleted beyond the flush ends and the three high producers have either three or four nucleotides deleted. There is no major difference in the amount of antigen accumulated by various high producers; clone pt12, which has four nucleotides deleted, accumulated about 1 1/2 times as much t antigen as clone pt24, which has three nucleotides deleted. However, the difference in accumulation between the low and high producers is >10-fold. The deletion of one additional nucleotide, therefore, resulted in >10-fold increase in efficiency. Whether a more extensive deletion can further increase the efficiency awaits analysis of more producer clones.

It can be estimated that no less than 0.4% of the [^{35}S]methionine incorporated over a period of 5 min at 37°C by clone pt12 is associated with t antigen. This is a conservative estimate because the t antigen synthesized in these clones is highly unstable. (Such a conclusion is based on the observation that, although the extent of incorporation of [^{35}S]methionine is directly proportional to the time of labeling over a period of 30 min, the extent of accumulation of t antigen as determined by immunoprecipitation in the presence of an excess antiserum reaches a plateau in less than 5 min.) Nevertheless, in view of this instability, estimation of steady-state accumulation of t antigen in these bacterial clones is of little significance.

In order to derive a general vector for the efficient expression of other eukaryotic proteins, we selected clones that have retained *Hin*dIII sites at both ends of the SV40 insert. These recombinant clones were derived from ligation of the 16/20 duplex at its filled-in *Hin*dIII terminus to the flush ends of the SV40 *Hin*dII/*Hin*dIII fragment A, thus generating complete *Hin*dIII cleavage sites. This would allow the resection of the entire SV40 sequence upon treatment with *Hin*dIII, leaving the synthetic RBSs covalently associated with both ends of the original *Pst* I site in pBR322 DNA. The only *Hin*dIII site within the pBR322 DNA lies in the promoter region of the Tc^R gene and therefore can be protected against *Hin*dIII digestion during the removal of the SV40 insert, by *E. coli* RNA polymerase (34). Proper insertion of any coding sequence can be achieved by first

randomizing the ends of the DNA fragment to be cloned with exonuclease III/*BAL* 31 and nuclease S1 and subsequently ligating to this vector through *Hin*dIII linkers. A method has been developed for the proper attachment of linkers without the necessity for recleavage with the restriction enzyme to avoid cutting putative internal sites (unpublished data). Expression of the insert would then depend on the conservation of the entire coding sequence devoid of intervening sequences (e.g., cDNAs derived by reverse transcription of mRNAs) and the optimal distance from the RBS to the initiation codon. Such a vector should have general applications because of the ease of proper insertion of heterologous DNAs to achieve efficient expression of proteins.

This work was supported in part by grants (to E.J.) from the National Cancer Institute of Canada and the Natural Science and Engineering Research Council of Canada (no. 9987).

1. Grunberg-Manago, M. & Gros, F. (1977) *Prog. Nucleic Acid Res. Mol. Biol.* **20**, 209–284.
2. Steitz, J. A. (1979) in *Biological Regulations and Development*, ed. Goldberger, R. F. (Plenum, New York), pp. 349–399.
3. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Biochem.* **13**, 319–353.
4. Scherer, G. F. E., Walkinshaw, M. D., Arnott, S. & Morre, D. J. (1980) *Nucleic Acids Res.* **8**, 3895–3907.
5. Shine, J. & Dalgarno, L. (1975) *Nature (London)* **254**, 34–38.
6. Steitz, J. A. & Jakes, K. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4734–4738.
7. Dunn, J. J., Buzash-Pollert, E. & Studier, F. W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2741–2745.
8. Roberts, T. M., Bikel, I., Yocum, R. R., Livingston, D. M. & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5596–5600.
9. Tal, M., Aviram, M., Kanarek, A. & Weiss, A. (1972) *Biochim. Biophys. Acta* **281**, 381–392.
10. Szer, W., Hermoso, J. M. & Leffler, S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 2325–2329.
11. Jay, E., Seth, A. K. & Jay, G. (1980) *J. Biol. Chem.* **255**, 3809–3812.
12. Kozak, M. & Shatkin, A. J. (1978) *Cell* **13**, 201–212.
13. Baralle, F. E. & Brownlee, G. G. (1978) *Nature (London)* **274**, 84–87.
14. Kozak, M. (1978) *Cell* **15**, 1109–1123.
15. Crawford, L. V., Cole, C. N., Smith, A. E., Paucha, E., Tegtmeyer, P., Rundell, K. & Berg, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 117–121.
16. Prives, C., Gilboa, E. Revel, M. & Winocour, E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 457–461.
17. Wu, R., Jay, E. & Roychoudhury, R. (1976) *Methods Cancer Res.* **12**, 87–176.
18. Wu, R., Ruben, G., Siegel, B., Jay, E., Spielman, P. & Tu, C. D. (1976) *Biochemistry* **15**, 134–140.
19. Sugino, A., Goodman, H. M., Heyneker, H. L., Shine, J., Boyer, H. W. & Cozzarelli, N. R. (1977) *J. Biol. Chem.* **252**, 3987–3994.
20. Jay, G., Jay, F. T., Chang, C., Friedman, R. M. & Levine, A. S. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3055–3059.
21. Jay, G., Ferrini, U., Robinson, E. A., Khoury, G. & Appella, E. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6562–6566.
22. Jay, G., Shiu, R. P. C., Jay, F. T., Levine, A. S. & Pastan, I. (1978) *Cell* **13**, 527–534.
23. Bolivar, F., Rodriquez, R. L., Green, P. J., Betlach, M.-C., Heyneker, H. L. & Boyer, H. W. (1977) *Gene* **2**, 95–113.
24. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. & Weissman, S. M. (1978) *Science* **200**, 494–502.
25. Fiers, W., Contreras, R., Haegeman, G., Rogier, R., Van De Voorde, A., Van Heuverswyn, H., Van Herrweghe, J., Volckaert, G. & Ysebaert, M. (1978) *Nature (London)* **273**, 113–120.
26. Seth, A. & Jay, E. (1980) *Nucleic Acids Res.* **8**, 5445–5459.
27. Jay, E., Bambara, R., Padmanabhan, R. & Wu, R., (1974) *Nucleic Acids Res.* **1**, 73–93.
28. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
29. Roberts, T. M., Kacich, R. & Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 760–764.
30. Goeddel, D. B., Heyneker, H. L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D. G., Ross, M. J., Miozzari, G. Crea, R. & Seeburg, P. H. (1977) *Nature (London)* **281**, 544–548.
31. Guarente, L., Lauer, G., Roberts, T. M. & Ptashne, M. (1980) *Cell* **20**, 543–553.
32. Taniguchi, T., Guarente, L., Roberts, T. M., Kimelan, D., Douhan, J. & Ptashne, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5230–5233.
33. Thummel, C. S., Burges, T. L. & Tjian, R. (1981) *J. Virol.* **37**, 683–697.
34. Rodriquez, R. L., West, R. W., Heyneker, H. L., Bolivar, F. & Boyer, H. W. (1979) *Nucleic Acids Res.* **6**, 3267–3287.