Binding of the bacteriophage T4 regA protein to mRNA targets: an initiator AUG is required

Shashikala Unnithan, Louis Green, Louisa Morrissey, Jonathan Binkley, Britta Singer, Jim Karam¹, and Larry Gold*

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Campus Box 347, Boulder, CO 80309 and ¹Department of Biochemistry and Molecular Biology, Medical University of South Carolina, Charleston, SC 29425, USA

Received June 26, 1990; Revised and Accepted November 2, 1990

ABSTRACT

Bacteriophage T4 regA protein translationally represses the synthesis of a subset of early phage-induced proteins. The protein binds to the translation initiation site of at least two mRNAs and prevents formation of the initiation complex (1,2). We show here that the protein binds to the translation initiation sites of other regA-sensitive mRNAs. Analysis of mRNA binding by filtration and nuclease protection assays shows that AUG is necessary but not sufficient for specific binding of regA protein to its mRNA targets. Anticipating the need for large quantities of regA protein for structural studies to further define the regA protein – RNA ligand interaction, we also report cloning the regA gene into a T4 overexpression system. The expression of regA protein in uninfected *E. coli* is lethal, so in our system regA driven by a strong T7 promoter is sequestered in a T4 phage until 'induction' by phage infection is desired. We have replaced the regA sensitive wild-type ribosome binding site with a strong insensitive ribosome binding site at an optimal distance from the regA initiation codon for maximizing expression (3). We have obtained large amounts of regA protein.

INTRODUCTION

In bacteriophage T4-infected *Escherichia coli* control of gene expression occurs at both transcriptional (4) and translational levels (5). Three T4 translational repressors have been identified. The proteins encoded by genes 32 (gp32) and 43 (gp43) repress their own synthesis by binding to the ribosome binding sites of their respective messages, thereby preventing formation of the translation initiation complex (6–8). T4 regA protein also represses its own translation, but unlike gp32 and gp43 which repress only their own synthesis, regA protein represses the synthesis of many additional early T4-induced proteins (5).

Studies on expression of the T4 rIIB gene showed that levels of rIIB protein were lower in regA⁺ relative to regA[–]

infections. Moreover, mutations within and flanking the initiation codon of rIIB mRNA resulted in derepression of the rIIB protein even when the infecting phage was regA⁺ (9). This suggested that regA protein, like gene 32 and 43 proteins, inhibits translation by binding to the translation initiation region of the target mRNA.

The regA gene has been cloned and sequenced (10,11) and the protein product has been purified to homogeneity (12–14). Biochemical studies with the purified repressor have shown that, indeed, regA protein alone can differentially repress the synthesis of several early T4-induced proteins (1,13). An initiation complex of 30S ribosomal subunits and initiator tRNA did not form on rIIB mRNA in the presence of regA protein (1). Further, nuclease protection experiments confirmed that regA protein binds to the nucleotides surrounding the translation initiation region of rIIB mRNA, and mutations within the initiator AUG of rIIB mRNA abolished regA-mediated protection against ribonucleases (1).

In the studies described here, we examined additional mutations within and flanking the T4 rIIB initiation codon and observed that these also abolish or decrease regA protein binding *in vitro*. In addition, we observed that the protein binds to the translation initiation domains of other regA-sensitive, but not to regA-insensitive, T4 mRNAs. Some non-initiator AUG sequences were found that bound regA protein *in vitro* although the mRNAs containing these sequences are known not to be regulated by the regA gene product *in vivo*. Using a filter binding assay, we also examined regA binding to wild type and mutant rIIB mRNAs, and to mRNAs made from synthetic templates. The results of the experiments also suggest that AUG is an important component of regA targets.

To further investigate the nature of regA protein binding to RNA, we need to have regA protein available in sufficient quantity and purity for crystallographic studies. Overexpression of regA protein is problematic for two reasons: a) the protein represses its own synthesis by binding the translation initiation domain of its mRNA, and b) the wild-type protein is toxic to the uninfected cell. In fact, one early attempt to overproduce the regA protein led to the isolation of a non-toxic variant incapable of repression (15). It seemed likely that we could both eliminate

* To whom correspondence should be addressed

self-repression and increase translational yield by replacing the wild-type ribosome binding site (RBS) with a very efficient synthetic one (3).

We decided to circumvent the second problem, that of regA protein toxicity, by using a T7 expression system (16,17) resident, when desired, within a bacteriophage T4 vector. The T7 ϕ 10 late promoter is not recognized at all in the absence of T7 gene 1 RNA polymerase; when gene 1 protein is provided, the T7 promoter becomes the most active promoter in the cell. We describe here our cloning of regA, the conditions used to overexpress it, and the purification of the protein.

MATERIALS AND METHODS

Bacteria and phage

E. coli strains NAPIV, NAPIV^{su+}, CR63, CR63(lambda); C600; and BB were used as hosts for phage infections (18–21). The isolation and properties of T4 rIIB mutations P53, z37, P53rev1 and FC301 have been described (15,22–24). Phage T4D and T4(rIIB Δ 26, Sa Δ 9, 42⁻, 56⁻, 46⁻, 33⁻, 55⁻) are from our collection. Phage lambda DE3 (16,17) was obtained from Bill Studier.

Plasmids

Recombinant plasmids pEM104 (carrying T4 gene regA and a 5' terminal segment of gene 43), pJM1168 (carrying T4 genes rpbA, 45, 44, 62, regA, and adjoining sequences), pTH4556 (carrying genes rpbA and 45), pTH4553 (carrying gene 45.2), and pPG1303 (carrying genes dexA and ORF1) were used to amplify *in vivo* the mRNAs of genes which are not abundantly expressed during a wild type T4 infection. The constructs and the methods for overproducing and purifying plasmid encoded mRNAs have been described in detail elsewhere (10,15,25,26). The HaeIII-HindIII fragment from the plasmid pLG400.HH870 (23) containing the wild type rIIB sequence or the HD263 mutation was inserted into the SmaI-HindIII site of the plasmid pSP65. The recombinant pSP65 plasmids were then digested with BglI and used as templates for *in vitro* transcription using SP6 RNA polymerase. Oligonucleotides were synthesized and purified as described (25). The construction of synthetic DNA templates is shown in Fig. 1.

In vitro transcription, RNA purification and RNA sequencing

In vitro transcription with SP6 RNA polymerase was performed as described (27) except that the concentrations of ATP, CTP, and GTP were 1 mM each and the UTP concentration was 0.2 mM to achieve highly labeled RNA. The specific activity of alpha-³²P UTP was about 7.5 Ci/mmol. The wild type labeled rIIB mRNA was separated from unincorporated alpha-³²P UTP by centrifugation through a 1 ml Sephadex G-50 column (28). *In vitro* transcription from synthetic DNA templates with T7 RNA polymerase and RNA gel purification were performed as described (29), except that the ATP, CTP, and GTP concentrations in the labeling reactions were 0.5 mM each and the UTP concentration was 0.05 mM. The specific activity of alpha-³²P UTP was approximately 20 Ci/mmol. Labeling of oligos, preparation of crude mRNA from T4 infections, and primer extension with AMV reverse transcriptase have been described earlier (25); primer extension reactions for footprinting were carried out at 48°C instead of 37°C.

Enzymes and reagents

Transfer RNAs were purchased from Subriden. T4 polynucleotide kinase was purchased from United States Biochemical Corporation and was used in accordance with the manufacturer's instructions. T7 RNA polymerase and the trinucleotide AUG were gifts from Olke Uhlenbeck. SP6 polymerase and the plasmid pSP65 were purchased from Promega. Restriction enzymes and T4 DNA ligase were from New England Biolabs. The Ribonucleases A and T1 were purchased from Boehringer Mannheim Biochemicals. RNases were diluted into a buffer containing 10mM Tris acetate, pH 7.4, 60 mM NH₄Cl, 10 mM Mg(OAc)₂, 6 mM β -mercaptoethanol prior to use in footprinting reactions. AMV reverse transcriptase was purchased from Life Sciences and diluted into either reverse transcriptase buffer (50 mM Tris-Cl pH 8.3, 60 mM NaCl, 6 mM Mg(OAc)₂, 10 mM dithiothreitol) for RNA sequencing reactions or into the footprinting buffer when used in primer extension RNase-protection reactions. RegA protein was purified as described below, and diluted into regA storage buffer (50 mM Tris-Cl, pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol and 50% glycerol) before use. RegA storage buffer was added to reactions lacking regA protein. IPTG, poly(U) agarose, Mono S column HR5/5 and DEAE sephacel were from Pharmacia. Taq DNA polymerase was from Promega.

Toeprint inhibition analysis

Primer extension inhibition analysis of translation initiation complexes has been described (1,6,30,31). An annealing mix containing crude nucleic acid from phage infected cells (or cells in which a T4 RNA was transcribed from a plasmid) and a 5' end labeled oligonucleotide primer in standard buffer without magnesium was heated to 60°C for 3 min and then immersed in a dry ice-ethanol bath. At this point, Mg(OAc)₂ was added to a final concentration of 10 mM. The mixture was allowed to thaw on ice and aliquots were distributed into reaction tubes containing dNTPs at a final concentration of 0.4 mM. Purified regA protein was added and the reactions were incubated on ice for 5 min. Next, 10 pmoles of preactivated 30S ribosomal

Figure 1. Construction of three synthetic regA targets. (A) Synthetic single-stranded oligonucleotides were annealed to yield the double stranded T7 promoter sequences shown. (B) Oligos 1, 2) or 3) were ligated to the T7 promoter DNA to give three DNA templates, each containing a different sequence around the AUG. (C) Pictured here are the RNAs transcribed from the templates shown in (B) by *in vitro* transcription with T7 RNA polymerase.

subunits and 100 pmoles of tRNA were added and the reactions were shifted to 37°C for 10 min. Finally, AMV reverse transcriptase was added and the reaction mixture was incubated for 15 min at 37°C. An equal volume of formamide loading buffer was added and the samples heated at 95°C for 3 min before they were analyzed on an 8% sequencing gel.

Nuclease protection analysis (footprinting)

The 5' end labeled primers were annealed to crude nucleic acid as described above. An annealing mix containing the primer and RNA in magnesium-free standard buffer was heated at 60°C for 3 min, and then immersed in a dry ice-ethanol bath. Mg(OAc)₂ was added to a final concentration of 10 mM and the annealing mix was allowed to thaw on ice. The mixture was aliquoted into reaction tubes containing dNTPs at a final concentration of 0.4 mM. Purified regA protein was added to each reaction and allowed to incubate with the RNA for 10 min on ice. RNase T1 or RNase A was then added in a range of concentrations that resulted in only partial digestion of the RNA. After continued incubation for 5 min on ice, AMV reverse transcriptase was added and the reactions were shifted to 48°C for 15 min. An equal volume of formamide loading buffer was added to each reaction. After incubation at 95°C for 3 min, the samples were analyzed on an 8% sequencing gel.

Nitrocellulose binding assays

The nitrocellulose binding assays were performed as described (8), with the exception that the reactions were incubated for 3 minutes at 37°C immediately after mixing the RNA and regA protein. The binding reaction buffer contained, and the nitrocellulose filters were washed with, 50 mM potassium acetate, 50 mM Tris-HCl (pH 7.5), 10 mM dithiothreitol. Internally labeled RNAs transcribed by T7 RNA polymerase from synthetic DNA templates were used in these experiments. The concentration of the RNA was approximately 4×10^{-10} M, and the regA protein concentration ranged from 5×10^{-8} to 5×10^{-6} M. After filtration of the binding reactions, the filters were dried, counted, and analyzed as described (8), except that the data points from one to three separate experiments were fit to a theoretical bimolecular binding curve using the Pascal program *BestKd* (D. Irvine, unpublished). This program generates a curve described by the following solution for the quadratic equation:

$$B = \frac{2 \times A \times P}{[R + K_d + P] + \sqrt{[R + K_d + P]^2 - 4 \times P \times R}}$$

Quadratic Solution for Receptor-Ligand Binding

where B is the percent of total RNA bound to the filter, A is the percent of RNA at which binding saturates (70 to 80% for the RNAs used here), P is the total protein concentration, R is the total RNA concentration, and K_d is the equilibrium dissociation constant for the bimolecular reaction. The quadratic solution gives accurate results for reversible, bimolecular receptor-ligand binding even when conditions required for accuracy of a Michaelis-Menten type of approximation (i.e., $[Receptor] < K_d < [Ligand]$) are not satisfied. This particular form chosen for solution of the quadratic equation also reduces numerical roundoff during computations (32).

Competition experiments

In the competition experiments, regA protein (6×10^{-6} M) was added to reaction mixtures containing internally labeled wild type

rIIB RNA (2×10^{-8} M) and cold competitor RNA (0 to 90×10^{-6} M) in 100 mM Tris-HCl (pH 8.0), 80 mM KCl. The tubes were incubated on ice for 15 minutes, then at 37°C for 10 minutes. The reaction mixtures were then applied to pre-wet nitrocellulose filters on a Millipore filtration apparatus. Each filter was washed with 2 ml of 100 mM Tris-HCl (pH 8.0), 80 mM KCl, then dried and counted in Ecolume™ scintillation fluid. Controls were done in the absence of regA protein.

Cloning, overexpression, and purification of regA protein

Cloning

DNA primers were synthesized for cloning the T4 wild type allele of regA from bacteriophage T4D by the polymerase chain reaction (PCR) using the Taq DNA polymerase. The 3' PCR primer anneals to the region 25 to 58 nucleotides 3' to the stop codon of regA and adds a BamHI site at the 3' end. The 5' PCR primer anneals to bases 1–31 of the protein coding region of regA, and adds a strong synthetic *E. coli* ribosome-binding site (RBS) (3) and a HindIII restriction site at the 5' end (see Fig. 2).

The PCR reaction was done under the following conditions: 5 µg/ml purified T4D DNA, 0.4 µM each primer, 250 µM dNTP's (each), 50 mM KCl, 10 mM Tris-Cl pH 8.4, 2.5 mM MgCl₂, 100 µg/ml BSA, 25 u/ml Taq DNA polymerase. A reaction cycle consisted of 60 sec at 95°, 45 sec at 50°, and 90 sec at 65°. The reaction was run for 35 cycles. The ~500 bp PCR product was band purified from a 1% agarose gel, digested with HindIII and BamHI, and cloned into the HindIII and BamHI sites of the plasmid expression vector pDIP19D (see Fig. 2). The resultant plasmid pDIP19D:regA was maintained in *E. coli* host strain MC1061 which contains no T7 RNA polymerase (T7 RNAP).

Plasmid pDIP19D was constructed in this laboratory, and details of its general usefulness as part of a bacteriophage T4 expression system will be given elsewhere (33). This plasmid has the T7 gene 10 promoter directed toward a pUC19 polylinker (Fig. 2). Downstream from the T7 promoter and polylinker is a T7 transcriptional terminator sequence. These sequences are flanked by T4 rII sequences that allow DNA cloned into the pDIP19D polylinker to be incorporated into an appropriate T4 bacteriophage by homologous recombination. Upstream from the rII sequence is the *E. coli* *rpoC* transcriptional terminator sequence which reduces read-through transcription.

The DNA sequence of the cloned regA gene was determined from the T7 promoter of the vector through the translational stop codon of regA, by dideoxy sequencing of alkali-denatured plasmid DNA (34). Sequencing verified the presence of the synthetic RBS followed by the complete wild type T4 regA coding sequence.

Figure 2. Construction of pDIP19D:regA. Plasmid pDIP19D:regA is wild-type T4 regA sequence cloned in the HindIII and BamHI sites of pDIP19D, with a synthetic Shine and Dalgarno sequence 8 basepairs upstream from the initiation codon (fmet).

Expression of regA protein from pDIP19D:regA

Bacteriophage lambda DE3 carries the T7 gene 1 coding for T7 RNAP under lacUV5 promoter control. Induction of genes under T7 promoter control in *E. coli* can thus be achieved by infecting IPTG-induced cells with DE3. Plasmid pDIP19D:regA could not be maintained in the DE3 lysogen host BL21(DE3) (16).

General techniques for growing lambda bacteriophage were followed (28,35). A plate lysate stock was made from a single plaque of DE3 grown on *E. coli* strain C600, yielding a bacteriophage titer of 6×10^9 pfu/ml. To expand and concentrate this stock, 200 μ l (1.2×10^9) was added to 10 ml of an overnight C600 culture (L broth + 10 mM MgSO₄) in a 2 liter culture flask. After adsorption for 10 min at room temperature, 500 ml of L broth with 10 mM MgSO₄ was added and incubated at 37° with vigorous shaking for 7 hrs. CHCl₃ (1 ml) and 14.5 g NaCl was added and incubation continued for 15 min. Cell debris was removed by centrifugation, and the bacteriophage precipitated by dissolving 25 g. PEG 6000 in the cleared lysate, followed by a 30 min incubation on ice. The bacteriophage were collected by centrifugation at 10000 rpm in a Beckman JA10 rotor, and resuspended in 10 ml SM buffer (28). The titer of this 50-fold concentrated lysate was 2×10^{11} pfu/ml.

To express regA protein from pDIP19D:regA in host MC1061, a single colony was inoculated into 10 ml L broth + 0.2% maltose + 30 μ g/ml chloramphenicol and grown overnight at 37°. This culture was diluted 1:100 into the same medium, then grown to $A_{600} \approx 0.4$, IPTG added to 1.0 mM and incubation continued to $A_{600} \approx 0.8$. DE3 was added to give 2×10^9 phage per ml of culture (a multiplicity of infection of about 5 phage particles per cell). One ml samples of infected culture were taken at timed intervals and treated for separating proteins on SDS-PAGE gels, or the entire culture was used for regA protein purification 3–4 hours after infection.

Expression of regA protein from T4DIP19D:regA

Bacteriophage T4(rIIΔ26, SaΔ9, 42⁻, 56⁻, 46⁻, 33⁻, 55⁻) was grown on pDIP19D:regA/CR63 and the resulting lysate plated on *E. coli* hosts CR63 and CR63(lambda). Plating on the host lysogenic for lambda selects for T4 phage that have become rII⁺ by homologous recombination with the pDIP19D:regA plasmid; the regA construct cloned between the plasmid regions of rII homology is passively carried along (Fig. 2). About 10^{-4} phage were pseudo-wild for rII; one was chosen for further study and named T4DIP19D:regA. T4DIP19D was constructed in the same manner from pDIP19D/CR63, for use as a control phage that carries the T7 promoter and pUC19 cloning region without the regA gene.

Expression of regA protein from T4DIP19D:regA was accomplished by infection of IPTG-induced pAR1219/NapIV cells. Plasmid pAR1219 (16) is a variant of pBR322 with the *E. coli* lacI gene and the lacUV5 promoter driving T7 gene 1 (RNAP). One ml of T4DIP19D:regA lysate at 1×10^{10} pfu/ml (in supplemented M9 medium + 100 μ g/ml ampicillin) was added per 3 ml of IPTG (1 mM) induced cells grown to $A_{600} = 0.4$ in the same medium at 37°. One ml samples of infected culture (10–15 ml) were taken at timed intervals for analysis on SDS-PAGE gels or the entire culture (1–2 L) was harvested for regA protein purification.

Purification of regA protein

Buffers used in the purification of regA protein from T7 expression vectors were as follows. A: 20 mM Tris-Cl pH 7.5, 100 mM NaCl, 1 mM EDTA, 10 mM MgCl₂, 1 mM

β -mercaptoethanol; B: 50 mM Tris-Cl pH 7.5, 1 mM EDTA, 1 mM β -mercaptoethanol, 50 mM NaCl, 10% (v/v) glycerol; C: 20 mM K₂HPO₄ pH 7.0, 1 mM EDTA, 1 mM β -mercaptoethanol; D: 50 mM Na₂HPO₄ pH 7.5, 1 mM EDTA, 1 mM DTT; E: 50 mM Tris-acetate pH 7.5, 0.1 mM EDTA, 1 mM DTT, 50% (v/v) glycerol.

RegA protein synthesis was induced by DE3 infection of a one liter culture of cells containing pDIP19D:regA, or by infection with T4DIP19D:regA as described above. The cells were harvested by centrifugation at 6000g for 10 min and resuspended in 50 ml of buffer A. All procedures were carried out at 4° unless otherwise indicated. The cell suspension was stirred for 2 hrs and then passed through a French press as described (14). Triton X-100 was added to 2% and NaCl to 0.5M and stirred for 3 hrs (releases regA protein bound to membranes). The cell lysate was then centrifuged at 35000 $\times$ g for 30 min. The resulting supernatant was centrifuged again for 2 hrs at 100,000 $\times$ g. The pellet was discarded and the supernatant was dialysed against buffer B and chromatographed on a 1.0 $\times$ 10 cm. DEAE sephacel column equilibrated with buffer B. Most of the regA protein was in the flow-through fractions, which was then precipitated with 50% (w/v) ammonium sulfate. The ammonium sulfate precipitate was collected by centrifugation at 5000 $\times$ g for 30 min and resuspended in buffer C, and then chromatographed on a Mono S column equilibrated with buffer C using an FPLC system. RegA protein was eluted with 20 ml of a linear 0–0.5 M KCl gradient. The regA protein eluted as a sharp peak at 70–90 mM KCl. The fractions containing regA were dialyzed against buffer D and applied to a poly(U) agarose column and eluted as described (12,14). The poly(U) fraction was loaded on a DEAE sephacel column to remove any poly(U) contamination. Analysis of samples by SDS-PAGE after each stage of purification is shown in Figure 4. RegA protein was then dialyzed against buffer E and stored at –20°C.

RESULTS**Overexpression of regA protein**

The PCR cloning techniques used for regA allowed us to make several desired changes at once. HindIII and BamHI cloning sites were introduced flanking the gene and the natural RBS was replaced by a synthetic one simply by adding these sequences to the PCR primers that are otherwise homologous to the wild-type T4 regA gene (Fig. 2). The entire regA coding region of pDIP19D:regA was wild-type sequence, confirmed by DNA sequencing.

The overexpression of regA protein from the pDIP19D:regA plasmid is shown by Coomassie-stained SDS-PAGE gel (Fig. 3, lane 11). Transcription is from a T7 promoter, with the T7 RNAP introduced into the IPTG-induced host cell by infection with phage DE3. The pDIP19D:regA plasmid could not be transferred into an *E. coli* host [BL21(DE3)] with an uninduced source of the T7 RNAP. There must be a background level of T7 RNAP in this cell sufficient to express a killing level of regA protein. Translation is initiated from a synthetic RBS that should preclude reduction in expression by autoregulation that occurs with the natural regA RBS (7). Expression of regA protein is abundant from this plasmid, and increases for 3–4 hours after infection.

The pDIP plasmids were designed to allow incorporation of any cloned gene (along with the plasmid T7 promoter and terminator) into a T4 bacteriophage by homologous recombination (see Materials and Methods and Fig. 2). RegA

Figure 3. Time-course of regA protein synthesis after infection of IPTG-induced pAR1219/NapIV with T4 phage. Details of infection conditions given under Materials and Methods. From left to right in the Fig., lanes 1–5 represent equal volumes of culture sampled after 0, 1, 2, 3 and 4 hrs after infection with T4DIP19D. Lanes 6–10 are the same times, but cells infected with T4DIP19D:regA. Lane 11 is pDIP19D:regA 4 hrs after infection with DE3. The positions of regA protein and T4 proteins rIIB and gp52 are indicated. Coomassie blue stained 15% SDS-PAGE gel.

from pDIP19D:regA (and pDIP19D sequence without regA as a control) was 'lifted' into T4 phage as described. Overexpression of regA protein by T4DIP19D:regA infection of cells containing induced T7 RNAP from plasmid pAR1219 is shown in Figure 3. Infection of induced cells with T4DIP19D yields no increase in regA protein beyond the wild-type level, but two much larger proteins are produced with time after infection (Fig. 3, lanes 1–5). They are the same molecular weight as proteins from T4 genes 52 and rIIB, which are downstream from the T7 promoter in T4DIP19D. Expression of regA protein from T4DIP19D:regA is at a high level and increases for at least 3 hrs after infection (Fig. 3, lanes 6–10). Gene products from T4 genes 52 and rIIB are also overexpressed from T4DIP19D:regA, but to a lesser extent than from the same phage without regA.

Purification of regA protein was by a combination of methods reported earlier (5,7) with some modifications. RegA protein purification was monitored by Coomassie blue staining of samples analyzed by SDS-PAGE. Samples analyzed at different stages of purification from regA plasmid containing cells and from T4DIP19D:regA infected cells are shown in Figure 4. The major difference in the stages of purification between expression from the plasmid and expression from T4 phage was that T4 gene 52 protein, also overexpressed from T4DIP19D:regA, co-purified with regA protein until the poly(U) agarose affinity chromatography step (Fig. 4). The yield of regA protein was approximately 15 mg per liter of T4DIP19D:regA infected culture, and only slightly less from DE3 infected regA plasmid bearing cells.

The purification process reported here can likely be further simplified to a protocol similar to that of Adari and Spicer (13). The 100000×g supernatant could be directly applied to a DEAE sephacel column and the regA protein fraction precipitated with ammonium sulfate and loaded on a poly(U) column (see Fig. 4). A second run through DEAE sephacel would remove any contaminating poly(U).

The expression systems described above produce high levels of regA protein. This enabled us to purify large quantities of regA

Figure 4. Stages in the purification of T4 regA protein from T4DIP19D:regA infected cells. The position of the regA protein band (regA), the T4 gene 52 protein band (arrow), and rIIB protein band (*) are indicated. See Materials and Methods section for details of the different purification steps. Samples taken after each indicated step were analyzed on this Coomassie blue stained 15% SDS polyacrylamide gel.

rIIB genotype	Nuclease Protection by reg A	Toeprint Inhibition by reg A	Sensitive to reg A <i>in vivo</i>
rIIB ⁺	strong	strong	yes
P53rev1	weak	weak	no
X504rev7	N/D	weak	no
FC301	no protection	N/D	no
P53	no protection	no	N/D
z37	no protection	no	N/D
HD263	no protection	no	no
zAP10	no protection	no	no

Figure 5. Sequence of the rIIB translational operator and locations of the regA constitutive mutations. The nucleotide represent the final result of the mutations in the sequence of rIIB. A summary of nuclease protection, toeprint inhibition and *in vivo* translational repression data for the rIIB mutants is also shown. N/D: not determined.

protein from small volumes of cell culture. Overexpression plasmids for regA have been reported (12–14) but the yield of regA protein from such systems have not been high. The system we describe produces regA as the major protein, which is nearly 20% of the total cell protein. We have obtained between 10 and 15 mg of regA protein from one liter of culture, using rather modest cell densities. The ability to purify large amounts of regA protein to homogeneity should now allow crystallographic studies of regA protein with and without bound ligands.

Analysis of T4 rIIB mutations which diminish regA binding

The effect of mutations within the initiation codon of the T4 rIIB cistron on binding by regA protein was analyzed by using the

primer extension inhibition assay [toeprinting (1,6,30,31)]. In this assay, a 5' end-labeled DNA primer is annealed to mRNA that is bound in a translation initiation complex with 30S ribosomal subunits and tRNA. The primer is then extended by reverse transcriptase, which is blocked by the initiation complex at about the fifteenth nucleotide 3' to the adenosine of the initiation codon. In these experiments, the initiation complex contains the minimal required components; only 30S ribosomal subunits and tRNA are needed to produce a toeprint on most mRNAs (30,31).

Previous results from such assays showed that the addition of regA protein inhibits the formation of an initiation complex on rIIB⁺ mRNA and eliminates the toeprint when added with or before (but not after) the 30S subunits (1). The toeprint of a 30S/tRNA complex bound to the mRNA prior to the addition of regA protein is insensitive to the repressor (unpublished observations). Mutations within and flanking the rIIB initiation

codon were analyzed for toeprint inhibition by regA protein. Fig. 5 summarizes the mutations in the rIIB translation initiation domain which affect sensitivity to regA protein and lists the results of nuclease protection, toeprint inhibition, and *in vivo* repression experiments with those rIIB mutations. P53rev1 and X504rev7 are mutations which flank the initiation codon, while HD263 and zAPI0 alter the nucleotides of the initiator AUG itself. All of these mutations render rIIB expression insensitive to regA mediated repression *in vivo* (9,23,24). Toeprint inhibition assays showed that rIIB⁺ is strongly repressed by 5 μ M regA, P53rev1 and X504rev7 are only weakly repressed by 10 μ M regA, and HD263 and zAPI0 are unresponsive to even 10 μ M regA (1).

We extended analysis of the rIIB binding site to include the rIIB mutations P53 and z37. P53 is a deletion of the T in the initiator ATG and z37 is a T $\rightarrow$ C transition mutation at the same position (Fig. 2; refs. 22–24). We have not been able to determine if P53 and z37 are regA-sensitive *in vivo* because the mutations within the initiation AUG eliminate translation. However, regA sensitivity *in vitro* was easily measured by using the tRNA cognate to the second codon of the rIIB message (tRNA^G) for toeprint formation. The toeprint from tRNA^G on rIIB⁺ mRNA was eliminated with the addition of 5 μ M regA protein, while the addition of up to 10 μ M regA protein had little or no effect on the tRNA^G toeprints of P53 and z37 mRNA (Fig. 6).

In footprinting experiments, P53 and z37 mRNAs were not protected by regA against nuclease attack; the ribosome binding site of P53rev1 mRNA was only weakly protected by 10 μ M regA protein (data not shown).

Finally, we examined regA binding to the rIIB mRNA isolated from cells infected with the mutant FC301. This mutation was originally isolated as a revertant of FC6 (18). Although the initiator AUG and the flanking sequences are unaltered in FC301, this mutation is not repressed in a regA⁺ infection (9). Consistent with *in vivo* observations, the addition of 5 μ M regA gave no protection against nuclease attack and 10 μ M regA gave only minor protection (data not shown).

Figure 6. Toeprint inhibition by regA protein on rIIB⁺, P53 and z37. Toeprint inhibition was performed as described in the Materials and Methods. Total RNA nucleic acid was purified from *E. coli* strain NAPIV infected with T4⁺ phage and strain BB infected with P53 or z37 phage. RegA protein was added to the reactions prior to the addition of 10 pmol of 30S ribosomes and 100 pmol of tRNA^G. For rIIB⁺, P53 and z37 we show (from left to right), primer extension in the absence of regA, 30S ribosomes and tRNA^G; primer extension in the presence of 5 μ M regA protein alone; and primer extension of reactions containing 30S ribosomes tRNA^G and 0, 1, 5 and 10 μ M regA protein. Brackets mark the position of the toeprint.

Figure 7. Nuclease protection analysis of regA bound to the mRNA encoding T4 proteins 45, 44 and regA itself. Primer extension analysis was performed as described in Materials and Methods. RNA was isolated from cells containing the plasmids pJM1168 (genes 44 and 45) and pEM104(regA). Each autoradiogram panel shows the mRNA sequence in the translation initiation region; primer extension without nuclease in the presence and absence of regA; partial digestion by ribonuclease T1; partial digestion by RNase A. When used, regA was added at 5 μ M. [RNase T1 cleaves RNA 3' to guanosine and ribonuclease A cleaves 3' to pyrimidines, showing a preference to pyr-A linkages. (A) Nuclease protection analysis of mRNA of gene 45. (B) Nuclease protection analysis of mRNA of gene 44. (C) Nuclease protection analysis of the regA message.

The translation initiation domains of other T4 mRNAs are protected by regA against nuclease digestion

We extended our analysis of regA protein-RNA interactions to include other regA-sensitive T4 mRNAs. We sought to determine if regA binds to the translation initiation region of each message and to identify the sequences required for binding.

We first studied the regA target sites on the messages of T4 genes 45, 44, and regA. During infections with T4 regA⁺ phage infections, the synthesis of gp45, gp44, and gpregA are rather low, while in a regA⁻ infection these proteins are overproduced (5,15,26). Similarly, *in vitro* transcription/translation systems have been used to demonstrate repression of these regA-regulated gene products by added regA protein *in vitro* (1,13). These observations, combined with results from rIIB studies, suggested that regA protein interferes with the translation by binding to the translation initiation domain and preventing the formation of the initiation complexes. Indeed, regA protein at a concentration of 5 μ M protected a region within the mRNA ribosome binding sites for genes 45 (Fig. 7A), 44 (Fig. 7B), and regA (Fig. 7C).

Although information about regA-mediated repression of T4-encoded mRNAs from genes dexA and rpbA during phage infection is unavailable, repression data for these genes has been obtained from *in vitro* transcription/translation and *in vivo* two plasmid assays (1,15). We found that regA protein protects the ribosome binding sites of both dexA (data not shown) and rpbA (data not shown) against nuclease digestion. The area of protection on rpbA mRNA was much larger than the protected regions observed for other regA targets. A repeated sequence (UAUUAUGACUAAA) lies upstream of the ribosome binding site for rpbA, and although the initiator AUG is downstream of this sequence, protection was stronger at the AUG within this sequence (data not shown). rpbA is one of the most sensitive to repression (maybe more than one binding site).

Figure 8. Nuclease protection analysis of an internal regA site on the mRNA of gene 43. Experimental methodology and panel display are the same as in Figs. 3 and 4. Nucleic acid was isolated from cells containing the plasmid pEM104.

Ribonuclease analysis of regA sites within the coding region for gene 43

The rIIB data showed that AUG is essential for regA binding, and all of the regA targets examined above contain an AUG sequence. Previous studies have shown that regA protein binds tightly to polyU-agarose columns (12) and so we hypothesized that the sequence U₄AUGU₄, which is contained within the coding region for gene 43, would bind the regA protein. However, results from nuclease protection assays showed that 5 μ M regA protein gave no protection and 10 μ M regA gave very minor protection from nuclease attack. In the same experiment, we noticed a small region of protected RNA upstream of the U₄AUGU₄ sequence (Fig. 8); the sequence of the protected region is UAUG.

Analysis of regA binding sites using a filter binding assay

To measure the strength of regA/operator interactions, nitrocellulose filter binding assays were performed on internally labeled RNAs transcribed from synthetic templates (assay described in Materials and Methods). For the interaction between regA protein and RNA R (Fig. 1), the equilibrium dissociation constant (K_d) was calculated to be 4.2×10^{-7} M. This RNA is homologous with wild type rIIB mRNA over the sequence AAAUUAUGUACAA. The transcripts RNA N and RNA C are identical to RNA R in length and sequence, except that the three nucleotides flanking the AUG on each side are Ns and Cs, respectively (Fig. 1). The K_d s for regA binding were calculated to be 7.9×10^{-7} M for RNA N, and 9.8×10^{-7} M for RNA C. The averaged data points from the assays and best-fit theoretical binding curves generated by the program described in Materials and Methods are shown in Fig. 9.

Figure 9. Binding of regA to internally labeled synthetic RNAs (see Fig. 1; experimental details in Materials and Methods). RNA R: theoretical curve (—), and averaged experimental data points from three experiments (+). RNA N: theoretical curve (---), and averaged experimental data points from three experiments (o). RNA C: theoretical curve (···), and experimental data points (□) from one experiment.

Competition between regA binding sites

Competition experiments between internally ^{32}P -labeled rIIB mRNA and other cold RNAs were done to determine the specificity of interaction between regA protein and its binding sites. Fig. 10 shows the competition between ^{32}P -labeled rIIB mRNA and the RNAs transcribed from the synthetic templates (see Fig. 1 and Materials and Methods) for regA binding. The transcript RNA R was able to compete strongly for regA binding; the transcripts RNA N and RNA C also competed, but to a lesser extent.

Competition experiments with internally ^{32}P -labeled rIIB mRNA were also done using unlabeled rIIB mRNA, HD263 mRNA, and the trinucleotide AUG. The results are shown in Fig. 11. The unlabeled rIIB mRNA competes strongly with itself as expected. The HD263 mRNA and the trinucleotide AUG do not compete at all.

Figure 10. Competition of regA binding to internally labeled wild type rIIB mRNA with unlabeled synthetic RNA R (□), synthetic RNA N (▲) and synthetic RNA C (●). The regA protein concentration used was $6\mu\text{M}$ (Materials and Methods).

Figure 11. Competition of regA binding to internally labeled wild type rIIB mRNA with unlabeled wild type rIIB mRNA (▲), unlabeled HD263 RNA (●), and the trinucleotide AUG (□). The regA protein concentration used was $6\mu\text{M}$ (Materials and Methods).

DISCUSSION

On the basis of previous studies that utilized the regA-regulated T4 rIIB cistron, we proposed that regA protein binds to a translational operator within the translation initiation domain of its target mRNA and prevents the formation of the initiation complex (1). The *in vitro* experiments described in this paper refine this model. We have observed that the translational operators of all the regA-sensitive genes we examined were protected from nuclease digestion when prebound with regA protein. RegA protein prevented the formation of translation initiation complexes only when it was added before the 30S/tRNA complex had been formed (unpublished observations).

Results from footprinting experiments illustrate that regA protein binds to discrete sequences in T4 mRNAs and that the initiator AUG is an essential component of these binding sites. A binding site that contains mutations within the initiator AUG is no longer protected from RNase digestion by regA protein. Interestingly, regA binding sites show no strong homology in their primary sequence other than for the presence of an AUG. AUG cannot be the only essential element for binding, since we have found several AUG sequences in T4 mRNAs which are not bound by regA protein. Some of these are initiation codons for regA-insensitive genes and some are AUGs internal to the coding regions of regA-sensitive genes.

Table I shows the protected regions of the different mRNA targets from nuclease digestion (ref. 1, and Figs. 7,8). Except in the case of the rIIB mRNA, the Shine-Dalgarno regions of all the mRNAs are protected, in addition to the initiator AUG. Another interesting observation from Table I is that the protected sites of all the mRNAs (except rpbA) are 16 to 24 nucleotides in length, which is quite large considering the small size of the regA protein. The rpbA mRNA has a 13 nucleotide sequence UAUUAUGACUAAA which occurs two times in the protected region. As compared to the other mRNAs, a much larger portion of the rpbA sequence is protected by regA. This may be because more than one molecule of regA binds to the rpbA mRNA.

TABLE I

Gene	mRNA Sequence (5' to 3')
rIIB	CCCUUGCGGCCUAAUAGgA AAAAUUUAUGACA UUAUAAUgCCUGAC
44	ACACUAAAAACUUGAAUg GAGGAAUUUAUGAUUA CUgUAAUUGAAAAAGA
45	UCUcAUUU GAAUUGAAGGAAUUUAUGAUA AACUgUCUAAAGAUACUAC
dexA	UUCGUGAUUUAg CGAGGAAAAUUUAUGUUUGA UUUUuAUUAUAGAUUU
rpbA	gUU ACUGACUUAUUUAUGACUAAAGGUGUUAUUUGACUAAAAUUACUGUGA AAUUA
regA	AAAAAACUAg CAUUGGAUUGGUAUAUAUGAUAUGAUAUA CUCUuAAAAA
43 internal	CAgAA UAUGA AAUUgAUGCUAUCA

Table 1. mRNA sequences of the regA-sensitive genes showing the regions protected by regA from ribonucleases. The initiator AUG and Shine-Dalgarno sequences are noted by *. The protected regions are shown as bold letters; the protected regions include the GN, UA, and CA dinucleotides attacked by T_1 and RNaseA. Each protected region is noted by lower case letters corresponding to the closest GN, UA, or CA dinucleotides that are not protected by regA protein, and hence delineate the extreme ends of the sites. The direct repeat of the rpbA mRNA is —.

Recently Webster et al. (2) published a detailed study of the regA protein binding site on the gene 44 translation initiation domain. Their work included both T1 and RNaseA footprinting of regA protein, thus overlapping the experiments shown in Figure 4B. We see protection (Table 1) that indeed includes the Shine and Dalgarno, as noted by Webster et al., but in our experiments protection extends 3' to the initiation AUG (as is found in every other site for regA binding). Our data differ from those published by Webster et al. in two critical nuclease sites on the gene 44 mRNA: the T1 site that includes the G of the initiation codon is protected in our experiments, as is the UA sequence bridging the second and third codons. The critical experiments in Webster et al. are shown in Figures 7 and 8 of their paper (2). Two differences are apparent in the experiments. We include an incubation at high temperature in our footprinting because the reverse transcriptase reaction (used to identify the RNase cuts) occurs in the presence of both nuclease and regA protein. The other difference is that the footprinting of gene 44 mRNA/regA protein complexes in Webster et al. utilizes 5' end labeled RNA which has a high background of bands in the RNA (for most of their data) in the very region of the gel corresponding to the G of the AUG and just 3' to that position. We note these differences because the conclusions of Webster et al. include a notion that the regA protein may bind to some of its sites without contacting the G of the initiation codon. Both the mutations in the rIIB initiation codon and all of our nuclease protection data suggest that the initiating AUG of any target is protected against T1 attack by the repressor. What is remarkable about the data for gene 44 mRNA (2 and Table 1) and the other four initiation domains (for genes 45, dexA, rpbA, and regA) is the protection of the G's of the Shine and Dalgarno, a protection not seen on the rIIB mRNA. The failure to protect the G's of the rIIB message can not be explained by sequence inspection: the unprotected rIIB Shine and Dalgarno G's are six nucleotides 5' to the A of the AUG, while in the other targets a pair of protected G's sits at every spacing upstream of the A in the range five to nine nucleotides. The important observation of Webster et al. (now extended to other mRNA targets) is that the regA protein can prevent inspection of the Shine and Dalgarno region by ribosomes; our work on the rIIB mRNA completely underestimated that potential activity of the repressor.

The results of the filter binding assays and competition experiments further emphasize the importance of the initiator AUG in regA binding. RegA protein binds well to the *in vitro* transcript RNA R, which contains the wild type rIIB sequence AAAAUUAUGUACAA; the equilibrium dissociation constant (K_d) for this interaction is $0.4 \mu\text{M}$. RNA R competes with wild type rIIB mRNA for regA binding as well as rIIB mRNA itself. RNA N and RNA C are identical in sequence to RNA R except that the three nucleotides flanking the AUG on each side are Ns and Cs, respectively. Surprisingly, the regA binding K_d s for these transcripts are within a factor of three of that for RNA R, and both compete with wild type rIIB mRNA for binding, albeit not as well as RNA R. In striking contrast, the HD263 mRNA does not compete at all. Thus, a mutation in the initiation codon virtually eliminates regA binding; base changes flanking the initiation codon inhibit binding to a lesser degree. Similar conclusions are drawn from toeprinting and nuclease footprinting experiments done on various rIIB mutations, as described in this study and in reference 1. As expected, the trinucleotide AUG did not compete with rIIB RNA for binding. AUG is thus necessary but insufficient for regA binding.

The role that mRNA structure plays in regA binding must be

examined further. Our explanation for why FC301 is regA-insensitive is that the mutation allows formation of an RNA structure which prevents regA protein from recognizing its binding site. FC301 mRNA can form a weak hairpin (Fig. 12) with a 4 base loop with the AUG contained in a hairpin. The wild type mRNA might form the same hairpin, but with only three nucleotides in the loop. Hairpins with four nucleotide loops are more stable than those with three nucleotide loops (36). Although FC301 mRNA probably forms a hairpin structure, the Shine-Dalgarno region is outside the hairpin and is accessible to the 30S ribosomes. Therefore 30S particles should bind to FC301 mRNA. When the hairpin breathes, there is probably a competition between the initiator tRNA [which may be already present on the 30S] and regA protein; tRNA must bind to the AUG faster than regA, so no repression occurs.

We still don't know the preferred RNA sequence for recognition by regA protein. Webster et al. (2) proposed that the sequence AAAUUA is a significant site. However, in the targets studied so far that sequence is both upstream and downstream from the AUG; additionally, the constitutive rIIB mutation P53rev1 (Fig. 5) retains the AAAUUA. In part because of our data for FC301, we think the recognition site will be unstructured and hence composed of conserved primary sequences, even if we can not identify them at this time. Simple alignment of the known targets by the initiation codons yields a weak consensus sequence (eleven target sequences exist; upper case letters are found between six and eight times, lower case letters are found five times, and the AUG is used to align the sequences and is thus invariant):

```

aGaAaAaUAUGAaUAaUAUA
g   u   u           u

```

The sequence appears interesting until one realizes that the consensus conforms to generalized rules about ribosome binding sites and that T4 DNA is about two thirds A/T. The contacts between regA protein and its targets, however, might include sequence elements (in addition to the AUG and perhaps a portion of the AAAUUA) that can be variably off-set from each other and still be bound by a single protein active site. This is similar in principle to the manner in which ribosome binding sites can be recognized by ribosomes even though the spacings between the Shine and Dalgarno domain and the initiation codon can be quite different (37). For the moment we remain quite uncertain how the small regA protein chooses a target over a non-target, and how one target is chosen with higher affinity over another. Our data suggest a need for more detailed mutational analysis of individual mRNA targets, including structural determinations from co-crystals of specific RNAs with the regA protein.

Figure 12. Putative hairpin structure formed by FC301 mRNA. The Shine-Dalgarno sequence is overscored.

ACKNOWLEDGEMENTS

We wish to thank Craig Tuerk, Sean Eddy and Mike Yarus for discussions throughout the course of this work. We also thank Doug Irvine for help with the K_d determination. This work was supported by L.G. NIH grants GM28685 and GM19963 and NSF grant BCS-8912259 and J.K. NSF grant DMB-8810062. We also thank the W.M. Keck Foundation for their generous support of RNA science on the Boulder campus.

REFERENCES

1. Winter, R.B., Morrissey, L., Gauss, P., Gold, L., Hsu, T. and Karam, J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7822–7826.
2. Webster, K.R., Adari, H.Y. and Spicer, E.K. (1989) *Nucl. Acids Res.* **17**, 10047–10068.
3. Gold, L. and Stormo, G.D. (1990) *Methods Enzymol.* **185**, 89–93.
4. Kassavetis, G.A. and Geiduschek, E.P. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5101–5105.
5. Wiberg, J.S. and Karam, J.D. (1983) in Mathews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B. (eds.), *The Bacteriophage T4*. Am. Soc. Microbiol., Washington D.C., pp. 193–201.
6. McPheeters, D.S., Stormo, G.D. and Gold, L. (1988) *J. Mol. Biol.* **201**, 517–534.
7. Andrade, M., Guild, N., Hsu, T., Gold, L., Tuerk, C. and Karam, J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7942–7946.
8. Tuerk, C., Eddy, S., Parma, D. and Gold, L. (1990) *J. Mol. Biol.* **213**, 749–761.
9. Karam, J., Gold, L., Singer, B.S. and Dawson, M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4669–4673.
10. Trojanowska, M., Miller, E.S., Karam, J., Stormo, G. and Gold, L. (1984) *Nucl. Acids Res.* **12**, 5979–5993.
11. Adari, H.Y., Rose, K., Williams, K.R., Konigsberg, W.H., Lin, T.-C. and Spicer, E.K. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1901–1905.
12. Miller, E.S., Winter, R.B., Campbell, K.M., Power, S.D. and Gold, L. (1985) *J. Biol. Chem.* **260**, 13053–13059.
13. Adari, H.Y. and Spicer, E.K. (1986) *Proteins: Structure, Function and Genetics* **1**, 116–124.
14. Liang, Y., Wei, R., Hsu, T., Alford, C., Dawson, M. and Karam, J. (1988) *Genetics* **119**, 743–749.
15. Miller, E.S., Karam, J., Dawson, M., Trojanowska, M., Gauss, P. and Gold, L. (1987) *J. Mol. Biol.* **193**, 397–410.
16. Studier, F.W. and Moffatt, B.A. (1986) *J. Mol. Biol.* **189**, 113–130.
17. Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) *Methods Enzymol.* **185**, 60–89.
18. Nelson, M.A., Erickson, M., Gold, L. and Pulitzer, J.F. (1982) *Mol. Gen. Genet.* **188**, 60–68.
19. Appleyard, R.K., McGregor, J.F. and Baird, K.M. (1956) *Virology* **2**, 565–574.
20. Appleyard, R.K. (1954) *Genetics* **39**, 440–452.
21. Karam, J.D. and Bowles, M.G. (1974) *J. Virol.* **13**, 428–438.
22. Barnett, L., Brenner, S., Crick, F.H.C., Shulman, R.G. and Watts-Tobin, R.J. (1967) *Phil. Trans. Roy. Soc. Ser. B.* **252**, 487–560.
23. Shinedling, S., Gayle, M., Pribnow, D. and Gold, L. (1987) *Mol. Gen. Genet.* **207**, 224–232.
24. Singer, B.S., Gold, L., Shinedling, S.T., Colkitt, M., Hunter, L.R., Pribnow, D. and Nelson, M.A. (1987) *J. Mol. Biol.* **149**, 405–432.
25. Gauss, P., Gayle, M., Winter, R.B. and Gold, L. (1987) *Mol. Gen. Genet.* **206**, 24–34.
26. Hsu, T., Wei, R., Dawson, M. and Karam, J.D. (1987) *J. Virol.* **61**, 366–374.
27. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucl. Acids Res.* **12**, 7035–7056.
28. Maniatis, T., Fritsch, E.F., Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor University Press, Cold Spring Harbor.
29. Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) *Nucl. Acids Res.* **15**, 8783.
30. Hartz, D., McPheeters, D.S., Traut, R. and Gold, L. (1988) *Methods Enzymol.* **164**, 419–425.
31. Hartz, D., McPheeters, D.S. and Gold, L. (1989) *Genes and Develop.* **3**, 1899–1912.
32. Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1988) *Numerical Recipes in C*. Cambridge University Press, Cambridge.
33. Singer, B.S. and Gold, L., in preparation.
34. Schneider, T.D. and Stormo, G.D. (1989) *Nucl. Acids Res.* **17**, 659–674.
35. Silhavy, T.J., Berman, M.L. and Enquist, L.W. (1984) *Experiments with Gene Fusions*. Cold Spring Harbor University Press, Cold Spring Harbor.
36. Groebe, D.R. and Uhlenbeck, O.C. (1988) *Nucl. Acids Res.* **16**, 11725–11735.
37. Gold, L. (1988) *Ann. Rev. Biochem.* **57**, 199–233.