
The bacteriophage T4 *regA* gene: primary sequence of a translational repressor

Maria Trojanowska, Eric S. Miller⁺, Jim Karam, Gary Stormo⁺ and Larry Gold⁺

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

Received 8 June 1984; Accepted 12 July 1984

ABSTRACT

The *regA* gene product of bacteriophage T4 is an autogenously controlled translational regulatory protein that plays a role in differential inhibition (translational repression) of a subpopulation of T4-encoded "early" mRNA species. The structural gene for this polypeptide maps within a cluster of phage DNA replication genes, (genes 45-44-62-*regA*-43-42), all but one of which (gene 43) are under *regA*-mediated translational control. We have cloned the T4 *regA* gene, determined its nucleotide sequence, and identified the amino-terminal residues of a plasmid-encoded, hyperproduced *regA* protein. The results suggest that the T4 *regA* gene product is a 122 amino acid polypeptide that is mildly basic and hydrophilic in character; these features are consistent with known properties of *regA* protein derived from T4-infected cells. Computer-assisted analyses of the nucleotide sequences of the *regA* gene and its three upstream neighbors (genes 45, 44, and 62) suggest the existence of three translational initiation units in this four-gene cluster; one for gene 45, one for genes 44, 62 and *regA*, and one that serves only the *regA* gene. The analyses also suggest that the gene 44-62 translational unit harbors a stable RNA structure that obligates translational coupling of these two genes.

INTRODUCTION

Several protein translational repressors have been identified in prokaryotic organisms (1). These include ribosomal proteins of *Escherichia coli* (2-5), the coat and replicase proteins of the RNA bacteriophages (6-10), the gene V protein of the filamentous DNA phages (11,12), the bacteriophage P22 scaffolding protein (13-15), the single-stranded DNA binding (*ssb*) protein of bacteriophage T4 (16-18), and the T4 *regA* protein (19,20). The T4 *ssb* protein (encoded by gene 32), is known to be essential for phage DNA replication, DNA repair and genetic recombination (21-23). In addition to binding DNA, the T4 *ssb* protein can interact specifically with its own mRNA to effect autogenous translational repression (16-18). Physiological and genetic studies suggest that the T4 *regA* protein is also a translational repressor of its own synthesis (19,20) as well as the synthesis of several other T4-encoded proteins (24,25). That is, mutations in the *regA* gene result in

overproduction of phage DNA replication proteins as well as of the defective regA polypeptides (26). One regA repressor-recognition site, on T4 rIIB mRNA, has been identified genetically (20); the site lies within the rIIB ribosome binding site. The regA-insensitive mutations in rIIB probably define a translational operator site at which the regA product competes with ribosomes for rIIB mRNA (20). The T4 regA gene maps within a cluster of DNA replication genes and may, conceivably, play a role in replication in addition to its role in regulating the synthesis of a set of replication enzymes (26,27). The roles of regA protein in translational regulation and replication could be elucidated via its purification and biochemical characterization.

Three properties of the T4 gene 32 system have facilitated cell-free characterization of translational repression by that protein: (a) phage infections can be manipulated to hyperproduce the protein (16,17), (b) the DNA-binding properties of the protein allow for its isolation and purification by simple chromatographic procedures (18), and (c) unlike the majority of T4-encoded mRNAs, the mRNA from gene 32, which bears the only known oligoribonucleotide target for repression by gene 32 protein, is intrinsically stable and can be isolated in biologically active form (18). By contrast, the T4 regA protein is synthesized in small amounts in phage-infected cells (19,24,25) and the substrate mRNA species that are repressed via regA protein constitute a heterogeneous population (28,29) which is difficult to fractionate into individual species. As a first step in developing sources of T4 regA protein for biochemical studies, we have constructed several recombinant plasmids that bear different regA alleles, some of which can be induced to overproduce the regA gene product. We report here the nucleotide sequence of the T4 regA gene and the amino-terminal sequence of an overproduced mutant regA protein encoded by a recombinant plasmid. We have noted as well a probable example of translational coupling for a polycistronic mRNA regulated by the regA protein.

MATERIALS AND METHODS

Bacterial strain and plasmids

Escherichia coli BE strain NapIV (hsdR κ ⁻, hsdM κ ⁺, hsdS κ ⁺, thi⁻; ref. 30) was used as the recipient in transformations with recombinant plasmids and as the host for overproduction of regA protein. Plasmids used to clone the regA gene for sequence analysis and for constructing gene fusions were pBH20 (Ap^R, Tc^R, lacPO; ref. 31), pUC8 (Ap^R, lacPO; ref. 32) and pGW7 (Ap^R, P₁ci1857; G. Wilson, N. E. Biolabs, Inc.).

Sources of DNA for cloning the *regA* gene

Several portions of the T4 genome have been cloned in lambdoid phage vectors in the laboratory of N. Murray (33). Dr. Murray provided us with an *imm21* recombinant phage designated NM761-4, which carries intact T4 genes 45, 44, 62 and *regA* as well as segments of T4 genes 46 and 43. Phage NM761-4 expressed these T4 genes under control of the *imm21* leftward promoter (P_L) (unpublished observations) and served as the source of the DNA that we used for subcloning *regA* in plasmids suitable for DNA sequence determinations.

Expression vectors, employing lambda P_L and *ci857*, were constructed from a pACYC184 *regA* clone previously identified in a T4 *HindIII* genomic bank (34). The *regA*-containing 2.0 kb *HindIII* fragment in this latter clone, pPG3, was derived from T4 phage that harbored mutations in several DNA synthesis genes. Although the original *regA* allele was wild-type, this DNA did yield a *regA* clone that failed to show autogenous translational repression (35; see Figure 3). Plasmid pEM104 is representative of this series of plasmids.

Enzymes and reagents

Restriction endonucleases, T4 polynucleotide kinase, T4 DNA ligase, and *E. coli* DNA polymerase I (Klenow fragment) were purchased from Bethesda Research Laboratories (Gaithersburg, MD), New England Biolabs (Beverly, MA), and New England Nuclear (Boston, MA) and were used in accordance with protocols from these suppliers. Calf intestine alkaline phosphatase, purchased as an ammonium sulfate suspension from Boehringer-Mannheim (Indianapolis, IN), was used at 60°C in 90 μ l of assay buffer containing 110 mM Tris-Cl, pH 8, 28 mM NaCl, 1 mM MgCl₂, 0.2 mM each of MgSO₄ and ZnSO₄, and 50 μ g linear DNA. *Bam*HI oligonucleotide linkers, purchased from Collaborative Research Inc. (Waltham, MA) were dissolved in TE (10 mM Tris-Cl, pH 8 and 0.1 mM EDTA) buffer at a concentration of 8 A₂₆₀ units/ml and were phosphorylated before use in ligation reactions. Phosphorylations were carried out on 500-600 pmol quantities of linkers at 37°C for 1 hr in 75 μ l of 60 mM Tris-Cl pH 7.8, 10 mM MgCl₂, 10 mM ATP, 25 mM DTT, and 10-60 units T4 polynucleotide kinase. The linker ligation reaction mix contained 3 mM ATP, 10 mM DTT, 150 pmole phosphorylated *Bam*HI linker, 0.2-1 μ g blunt-ended (or *BAL*-31-digested) plasmid DNA (0.5-3 pmole ends) and 5 units T4 DNA ligase. Ligations were carried out at 16°C overnight.

DNA sequence determinations

About 50 μ g of recombinant plasmid were linearized with a suitable restriction endonuclease (*Bam*HI or *Ava*I), dephosphorylated by treatment with calf-intestine alkaline phosphatase and then rephosphorylated in the presence

of [γ - 32 P]ATP and T4 polynucleotide kinase. The end-labeled DNA was then digested with a second restriction endonuclease to release the DNA fragment for purification by gel electrophoresis and use in nucleotide sequence determinations. The methods used were as described by Maxam and Gilbert (36).

Purification and characterization of a *regA* protein

Plasmid-encoded *regA* protein was purified from heat-induced cells containing the P_L-*regA* fusion plasmid pEM104 (see Figure 3). The purification scheme is detailed elsewhere (35). Briefly, the mutant *regA* protein encoded by pEM104 was found to be membrane-associated and was subsequently extracted with a hydrophobic solvent mixture of 40% acetonitrile:1-propanol (1:1), 0.5% n-octyl- β -D-glucopyranoside and 0.63% trifluoroamine- triethylamine. The solvent-extracted *regA* protein was then resolved to purity by high performance liquid chromatography on a C₁₈ reverse phase column. Proteins from induced cells, and from various stages of the purification were analyzed by SDS/urea-polyacrylamide gel electrophoresis according to described procedures (37). For amino acid analysis, samples of HPLC-purified *regA* protein (20 μ g) were dried in a Spin Vac Concentrator (Savant Instruments) at 42°C and then hydrolyzed at 105°C for 24, 48 and 72 hr in 0.5 ml 6N HCl plus 0.05 ml redistilled phenol. Samples were also hydrolyzed with performic acid (38) and then hydrolyzed as above for 24 hr in order to determine cysteine (cysteic acid) and methionine (methionine sulfone). All analyses were performed on a Durrum D500 amino acid analyzer. The *regA* protein (180 nmol) was treated prior to Edman degradation with [3 H]-iodoacetic acid in order to carboxymethylate free cysteine (39), and was then reisolated by reverse phase chromatography as described (35). The product was subjected to successive rounds of automated Edman degradation (40) on a Beckman 890D sequenator. Phenylthiohydantoin amino acids were identified by at least 2 out of 3 methods: reverse phase chromatography, gas chromatography or thin layer chromatography, performed at the University of California, Davis Protein Structure Facility. Although the protein is homogeneous on HPLC and SDS gels, only ca. 5% of the expected yield was obtained for each cycle. Apparently 95% of the *regA* protein contained a blocked amino-terminus, which might be N-formylmethionine.

RESULTS

Nucleotide sequence of the *regA* gene

Figure 1 shows a restriction map, derived from restriction enzyme digestion and partial nucleotide sequencing, of the T4 gene 45-44-62-*regA*-43

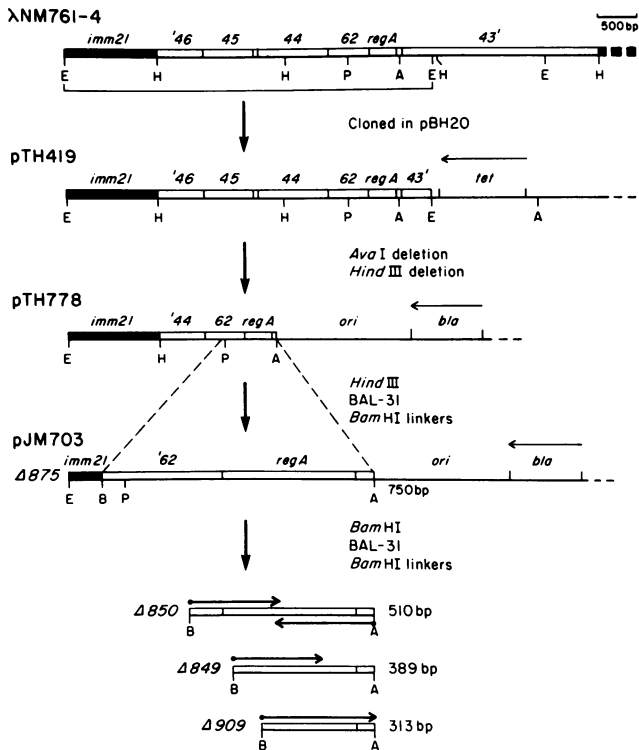


Figure 1

A diagram of the *regA* cloning scheme used for DNA sequence determinations. A portion of the T4 insert in lambdaoid recombinant phage NM761-4 was cloned in the *EcoRI* site of pBH20 to yield the recombinant plasmid pTH419. The pTH419 recombinant was subsequently deleted in two steps to remove the DNA segments bracketed by the *AvaI* and *HindIII* sites. The resulting plasmid, pTH778, was linearized by *HindIII* digestion and then subjected to a series of digestions by *BAL-31* exonuclease followed by introduction of synthetic *BamHI* endonuclease recognition sites at the termini of the *BAL-31*-generated deletions (see Materials and Methods). Sites of ³²P-end-labeling and the directions of sequencing are indicated by closed circles and bold arrows, respectively. The following designations are used to indicate sites of cleavage by restriction endonucleases: A (*AvaI*), B (*BamHI*), E (*EcoRI*), H (*HindIII*), P (*PstI*).

cluster. We noted that the *regA* gene fell between *PstI* and *AvaI* sites and we used these sites to subclone the gene and to construct overlapping *BAL-31* exonuclease-generated deletions that terminated at different distances from and into the structural gene. *BamHI* oligonucleotide linkers were introduced at the exonuclease-generated deletion termini. These clones facilitated the determination of the *regA* nucleotide sequence. By using the methods devised

```

      Δ850
      AAAGGTTTAGTCACGGATGATTTTTTGAAGAAGTGACTAAGAACGTAAGAAGAACAGAAACAACCTCAAAAA
      CTAGCATTGGAATGGTAAA 62 regA ATG ATT GAA ATT ACT CTT AAA AAA CCT GAA Δ849 GAT TTT CTG AAA
      LeuAlaLeuGluTrpEND Met Ile Glu Ile Thr Leu Lys Lys Pro Glu Asp Phe Leu Lys
      GTA AAA GAA ACT TTG ACT CGT ATG GGA ATT GCT AAT AAT AAA GAT AAA GTT CTA TAT
      Val Lys Glu Thr Leu Thr Arg Met Gly Ile Ala Asn Asn Lys Asp Lys Val Leu Tyr
      CAG TCC Δ909 TGT CAT ATT CTT CAG AAA AAA GGA CTA TAC TAT ATC GTT CAT TTT AAA GAA
      Glu Ser Cys His Ile Leu Glu Lys Lys Gly Leu Tyr Tyr Ile Val His Phe Lys Glu
      ATG CTT CGT ATG GAT GGT CGT CAA GTT GAA ATG ACA GAA GAA GAT GAA GTT CGT CGT
      Met Leu Arg Met Asp Gly Arg Glu Val Glu Met Thr Glu Asp Glu Val Arg Arg
      GAT TCG ATT GCA TGG CTA TTA GAA GAT TGG GGA CTA ATT GAA ATC GTT CCT GGT CAA
      Asp Ser Ile Ala Trp Leu Leu Glu Asp Trp Gly Leu Ile Glu Ile Val Pro Gly Glu
      AGA ACT TTT ATG AAA GAT TTA ACT AAT AAC TTC CGA GTT ATT TCT TTT AAA CAA AAA
      Arg Thr Phe Met Lys Asp Leu Thr Asn Asn Phe Arg Val Ile Ser Phe Lys Glu Lys
      CAT GAA TGG AAG CTC GTT CCT AAA TAT ACG ATT GGT AAT TAA GCAAGGCTTCGCCCCCTTA
      His Glu Trp Lys Leu Val Pro Lys Tyr Thr Ile Gly Asn END
      TTTGGAGTATAATATATCAAGAGCCTAATAACTCGGG
  
```

Figure 2

DNA sequence of the T4 regA gene and the inferred amino acid sequence of its protein product. Only the sequence of the mRNA strand is shown. The nucleotide sequence was determined by the Maxam and Gilbert (36) method after ³²P-labeling of cloned DNA fragments at the 5'-end of the AvaI site near the 3'-terminal region of regA (pTH778 and Δ875, Figure 1) and at BamHI-terminated deletions near the 5'-terminal end of the gene (Δ850, Δ849, and Δ909, Figure 1). Underlined amino acids are residues whose identity was confirmed by N-terminal sequencing of purified regA protein; all others were inferred from the nucleotide sequence, but are consistent with the amino acid composition of the purified protein (see text).

by Maxam and Gilbert (36), almost the entire sequence of the sense strand of the regA gene could be determined with [³²P]5'-end-labeled, AvaI-digested pTH778 DNA or its derivatives (Figure 1). Sequence determinations on the BamHI-digested plasmids bearing the deleted T4 fragments yielded overlapping data for the anti-sense strand of regA and also served to confirm the sequence obtained from the AvaI side of the gene. Figure 2 summarizes the sequence for the mRNA strand of regA derived from all the sequences obtained. Our determinations revealed an open reading frame of 122 amino acids preceded by nucleotide sequences suggestive of a translational initiation domain (see Discussion). Computer-assisted analysis of the open reading frame in Figure 2 indicates that the regA gene has a pattern of codon usage typical of bacteriophage T4 coding regions; it is comprised of 69% A+T and 81% of the third positions of each codon are A or T. The initiator AUG on the regA transcript is separated from the gene 62 termination codon, UAA, by one nucleotide. The C-terminal amino acid sequence of the T4 gene 62 protein and

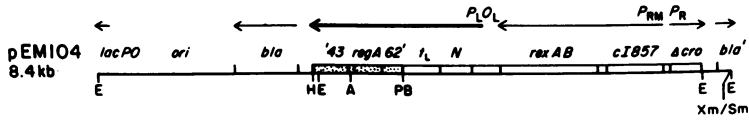


Figure 3

a) A diagram of the 8.4 kb expression vector pEM104 (35) used for overproduction of a *regA* protein. The plasmid (thin lines) is shown linearized at an *EcoRI* site. Lambda phage DNA sequences (open box) were joined to cloned T4 phage DNA (stippled box) at *PstI* and *BamHI* sites derived from the multiple cloning site of pUC8. The structure is drawn to scale, with gene boundaries indicated with vertical bars. Symbols used are described in the legend to Figure 1, plus Sm (*SmaI*) and Xm (*XmnI*).

b) An SDS/urea-polyacrylamide gel (16%) run according to the published procedure (37) and stained with Coomassie brilliant blue R. Samples are a) total protein from *E. coli* NapIV containing pGW7, incubated at 42°C, b) total protein from NapIV/pEM104 incubated at 30°C, c) total protein from NapIV/pEM104 incubated at 42°C, d) protein extracted from membranes isolated from cells as shown in lane c (see Materials and Methods), and e) ca. 4 μ g of HPLC peak fraction of purified *regA* protein (35).

its DNA sequence have been determined (41) and our nucleotide sequence for the region upstream from the regA initiator AUG matches this sequence perfectly.

Overproduction and characterization of a T4 regA gene product

Figure 3 diagrams pEM104, the recombinant plasmid that was used in the production of regA protein for amino acid composition and sequence analyses. This plasmid, a member of the pEM100-series of regA gene clones (Materials and Methods), harbored a mutant regA allele that rendered regA protein synthesis insensitive to autogenous translational repression (35). In these clones, transcription of the regA gene under control of the lambda cI857 P_LN genetic elements in pGW7 could be induced at 42°C (Figure 3a) and, because of transcriptional and translational derepression, large quantities of the regA gene product could be accumulated (Figure 3b). We estimate that under the conditions described in Figure 3, regA protein synthesis could be amplified to comprise at least 10% of the total cell protein. The properties of the pEM100-series of clones proved to be practical for generating sufficient quantities of protein for purification and use in amino-terminal sequencing studies and for preparation of antibody reactive against wild-type regA protein (35). Typically, 4-5 mg of regA protein were recovered from 500 ml of induced culture at 3×10^8 cells per ml.

Automated Edman degradation provided unambiguous identification of the seventeen N-terminal amino acids. The amino terminus agrees with the nucleotide sequence, and specifies the correct open reading frame. We also obtained an amino acid composition for the purified protein. The determined and predicted compositions are in agreement, with a correlation coefficient of 0.97. Lysine is the single most abundant residue (14 residues) and cysteine the least (1 residue). The presence of one cysteine residue was unequivocally confirmed by quantification of [³H]-carboxymethylation (35).

The nucleotide sequence of the regA gene (Figure 2) predicts a molecular weight of 14,600 for the encoded protein. This agrees with molecular weight estimates from SDS-gel electrophoretic assays on T4-encoded regA protein and from amino acid composition determinations on the pEM104-derived protein. Isoelectric focusing analysis of regA protein from T4-infected *E. coli*, from pEM104 and from other plasmid clones gives an experimentally determined pI of about 8.2 (Alford and Karam, unpublished observations). This is consistent with the slight predominance of basic amino acids (lysine and arginine) over acidic amino acids (aspartate and glutamate) in the protein (21 vs. 18, respectively). The occurrence of 17% basic amino acids in the regA protein, as compared to an average distribution of 11% in most proteins (42), is

perhaps related to an RNA binding function of the protein. It is hydrophilic, with only a short region between residues 80 and 91 of potential hydrophobic character (43). This also supports previous observations that, during T4 infection, the wild-type regA protein is membrane associated via an RNA component and can be released with ribonuclease or 0.2 M NaCl (26).

DISCUSSION

As a translational repressor, the T4 regA protein appears to be specific for a subclass of T4-encoded early (prereplicative) mRNAs (44). The regA protein probably binds to an RNA sequence and interferes with translation initiation and/or triggers message destruction. In the case of T4 rIIB mRNA, where the target sequence (translational operator) for regA-mediated repression was identified by genetics, the proposed operator includes the first three codons of the message, i.e. AUGUACAAU (20). Mutations that alter this sequence result in translational derepression of rIIB protein synthesis. In all the other cases of regA-mediated translational control, the target sites for regA protein action have not been identified. There is as yet no evidence that mRNA secondary structure plays a role in regA-mediated regulation and sequence differences among the operators of regA-regulated mRNAs may underlie their differential responses to this translational repressor (20,26). In the case of the T4 regA gene, which is a nonessential function under most physiological conditions, the operator for autogenous translational control (or a portion of it) may even reside within the gene 62 mRNA, directly upstream of the regA gene. Without speculating on the precise sequence of the regA target site for autogenous regulation, low levels of regA protein made from induced P_L vectors containing the wild-type gene also suggest it lies at the translation initiation domain or immediately 5' to it in the gene 62 coding region (manuscript in preparation). Other regA-regulated genes are being cloned in expressible form in bacterial plasmids and it should ultimately be possible to use in vitro site-directed mutagenesis to create alterations in presumed target domains and to assay for effects on translation of the plasmid encoded mRNAs and on binding of purified regA protein.

The role of the T4 regA translational repressor in phage development has not yet been explained. Most T4-encoded proteins that appear during the early stages after phage infection play either direct or indirect roles in the control of viral DNA replication (45). Although regA⁻ mutations do not exhibit obvious adverse effects on phage replication, they do result in

altered molar ratios of some of the phage DNA replication proteins. Also, some regA⁻ mutants are defective in phage-induced host DNA degradation (24) and so differ from the wild-type in their access to nucleotide precursors for DNA synthesis. Possibly, the regA protein constitutes a part of the multiprotein complexes that regulate the channelling of precursors from specific intracellular nucleotide pools to sites of DNA synthesis (46,27). We are particularly intrigued by the close genetic linkage between the T4 regA gene and genes 45, 44, and 62 (see Figure 1), the structural genes for three regA-regulated proteins that are known to be central components of the phage DNA replication complex. The translation initiation codon for regA is separated from the termination codon of its upstream neighbor, gene 62, by only one base (Figure 2) and it is possible that the two genes share regulatory signals with each other and with genes 44 and 45. Some insights about this possibility may be gained from computer-assisted analyses of the known nucleotide sequence of the T4 gene 45-44-62-regA-43 cluster (41 and Figure 2). We searched the entire sequence for potential ribosome binding sites by using a weighting function derived from the perceptron algorithm (47). An analysis using three different matrix weighting functions revealed eight potential ribosome binding sites on the sense strand. These exhibited positive values by at least two of the three matrices. The values from the analysis are listed in Table 1. The highest average matrix value (mv) was obtained for gene 44 (mv = 158), followed by gene 43 (mv = 110) and gene 45 (mv = 80). The regA gene and gene 62 each had significantly lower values (mv = 16 and mv = -29, respectively), as did two other potential ribosome binding sites that are followed by very short open reading frames (nucleotide positions 989 and 2168, Table 1). One site, at nucleotide position 196 (Table 1), could be an initiation site that yields an internal gene 45 product; although this site lacks an AUG or GUG for translational initiation it does have an AUU which, as for the E. coli IF-3 structural gene, may function as an initiation codon (48). Interestingly, the two weak ribosome binding sites for genes 62 and regA are separated by only one nucleotide from the preceding translation terminators. Translational reinitiation has been shown to occur when stop and start codons are adjacent or separated by only a few nucleotides (49-52). Thus, the data in Table 1 suggest that the regA gene and gene 62 initiation sites are potential reinitiation sites rather than independent ribosome binding sites (see also below). T4 genes 45, 44 and 43 on the other hand, should be well translated by ribosomes that bind directly to their initiation domains.

Table 1. Ribosome binding sites identified by perceptron weighting functions. Both sense and anti-sense strands of the contiguous 3,221 nucleotide sequence (from gene 45 through the first 450 nucleotides of gene 43) were analysed using the previously described algorithm (47). Only those sites identified with positive values by at least 2 of 3 matrices, in addition to gene 62, are included. Positions are listed in 5' to 3' order for both strands.

Nucleotide Position ^a		Matrix Values (mv)			
<u>Sense</u>	<u>Gene</u>	<u>Average</u>	<u>1</u>	<u>2</u>	<u>3</u>
67	45	80	-40	54	226
196		52	4	99	52
801	44	158	123	174	178
989		-31	18	3	-115
1762	62	-29	-17	-52	-17
2168		8	40	8	-25
2327	<u>regA</u>	16	9	-53	93
2772	43	110	129	150	50
Anti-Sense					
1685		46	52	-15	102
1269		-51	22	18	-194
944	?	161	177	216	91

^aThe nucleotide positions designated refer to the nucleotide count starting from the first base of the published gene 45 sequence (ref. 41).

We also scanned and analyzed the anti-sense strand RNA from these regions. One site, at nucleotide position 944 (Table 1), i.e. within the gene 44 coding region, was revealed by all three matrix weighting functions, giving an average mv of 161, the largest numerical assignment for any site on either strand (Table 1). This initiation site and open reading frame would encode a polypeptide comprised of 56 amino acids, of which 21 are basic or aromatic residues. Although the codon usage of this open reading frame is not correlated with other T4 genes, it could be constrained by the gene 44 coding sequence. We do not yet know if this strand is transcribed (53), but it is interesting to consider the possibility of overlapping, opposite orientation genes in T4. The two additional sites identified generate very short or no

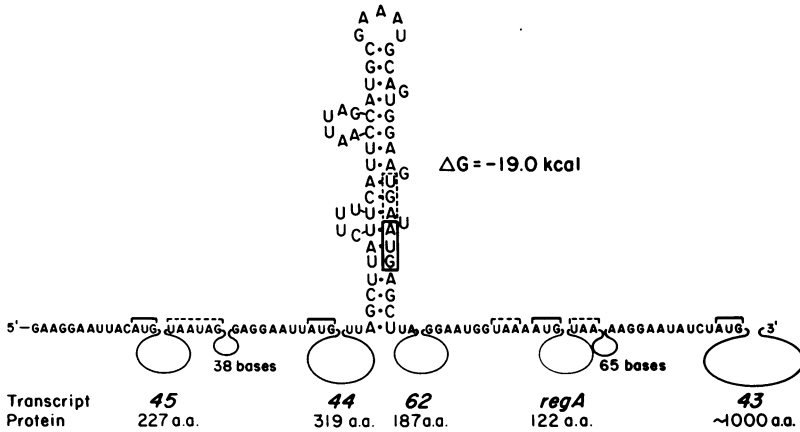


Figure 4

A postulated RNA secondary structure involved in the sequential translation of bacteriophage T4 genes 44 and 62. The gene 62 initiation codon is sequestered in the base-paired portion of the structure, which would be disrupted by ribosomes translating gene 44. Termination at the gene 44 UGA would position ribosomes for reinitiation at the adjacent AUG of gene 62. Translation of gene 62 is therefore dependent upon translation of the preceding gene 44 region. Gene starts (closed boxes), terminators (dashed boxes) and coding regions for the entire mRNA are shown.

open reading frames.

Secondary structure predictions

By using the Delila System (54,55) program HELIX, several potential secondary structures were identified in the T4 gene 45-44-62-regA-43 sequence. We focused on those having substantial stem regions, with nominal unpaired bases, and ΔG values of -15 Kcal or less. One striking structure was evident, whose basepairing pattern is shown in Figure 4. This 17 base-pair stem, with few unpaired bases, has a ΔG of -19.0 Kcal and could exist in vivo (56). The hairpin occurs near the terminus of the gene 44 coding region and, most significantly, sequesters the gene 44 stop codon and the gene 62 Shine and Dalgarno and AUG sequences within the base-paired portion of the stem. In addition to the structure shown in Figure 4, several pieces of evidence strongly suggest that genes 44 and 62 are translationally coupled: a) The perceptron matrix analysis (Table 1) gives an overall low value for the ribosome binding site of gene 62; it is thus not likely to function independently, b) the gene 62 initiation site is only one base pair removed from the gene 44 stop codon (41); the ribosome is precisely poised for reinitiation (49-52), and c) amber mutations in gene 44 are polar on gene 62

protein synthesis (57). We envisage that the traversing ribosome previously initiated at gene 44 would need to disrupt the secondary structure in order to allow reinitiation at gene 62. That is, even the weak ribosome binding site on the gene 62 mRNA (Table 1) is not available for initiation by free ribosomes. The gene 44 and 62 proteins comprise a single stranded DNA-dependent ATPase activity that is required for efficient T4 DNA replication (58,59). The two proteins are isolated from phage-infected cells as a tight complex consisting of gene 44 and gene 62 protein subunits in a relative ratio of 5:1 (58,60). Apparently, a mechanism has evolved whereby coordinate synthesis of the gene 44 and gene 62 proteins is assured. Whereas nonsense mutations in gene 44 are polar on gene 62 expression, amber mutations in gene 45 do not depress the synthesis of the gene 44 and gene 62 proteins. In addition, significantly more gene 45 protein is made than gene 44 and 62 proteins (57) and it was recently shown that gene 45 and 44 proteins are encoded by electrophoretically separable mRNA species (61). These observations support our view that genes 44 and 62 are co-transcribed and translationally coupled and we predict that the regA operator site for gene 44 will be found to translationally regulate synthesis of gene 62 protein. Although regA protein might also be made via reinitiation after gene 62 termination, our plasmid constructs express regA and show autogenous repression when the protein is wild-type (manuscript in preparation). Thus regA must also have an independent (and independently regulated) translational initiation domain.

ACKNOWLEDGEMENTS

We are grateful to Myra Dawson for her help with the DNA sequencing, to Scott Power and Al Smith for their expert assistance with the protein analyses and to Kathie Piekarski for her help with the preparation of the manuscript. This work was supported by N.S.F grant #PCM82-01869 and N.I.H. grants #GM28685 and #GM18842. E.S.M. was supported by N.I.H. Postdoctoral Fellowship #GMO9149.

REFERENCES

1. Campbell, K.M., Stormo, G.D. and Gold, L. (1983) Davies, J. and Gallant, J.A., eds., pp. 185-210, Gene Function in Prokaryotes, Beckwith, J., Cold Spring Harbor Laboratory, New York.
2. Baughman, G. and Nomura, M. (1983) *Cell* **34**, 979-988.
3. Johnsen, M., Christensen, T., Dennis, P. and Fiil, N.P. (1982) *EMBO J.* **1**, 999-1004.
4. Nomura, M., Dean, D. and Yates, J.L. (1982) *Trends Biochem. Sci.* **7**, 92-95.

5. Nomura, M., Jinks-Robertson, S. and Miura, A. (1982) Interaction of Translational and Transcriptional Controls in the Regulation of Gene Expression, Grunberg-Manago, M. and Safer, B., eds., pp. 91-104, Elsevier/North-Holland, Amsterdam.
6. Carey, J., Cameron, V., de Haseth, P.L. and Uhlenbeck, O.C. (1983) *Biochem.* 22, 2601-2610.
7. Spahr, P.F., Farber, M. and Gesteland, R.F. (1969) *Nature* 222, 455-458.
8. Kozak, M. and Nathans, D. (1972) *Bacteriological Revs.* 36, 109-134.
9. Meyer, F., Weber, H. and Weissmann, C. (1981) *J. Mol. Biol.* 153, 631-660.
10. Weber, H., Billeter, M.A., Kahane, S., Weissmann, C., Hindley, J. and Porter, A. (1972) *Nature New Biol.* 237, 166-170.
11. Model, P., McGill, C., Mazur, B. and Fulford, W.D. (1982) *Cell* 29, 329-335.
12. Yen, T.S.B. and Webster, R.E. (1982) *Cell* 29, 337-345.
13. Wyckoff, E. and Casjens, S. (1984) *J. Virol.*, submitted.
14. Casjens, S. and Adams, M.B. (1984) *J. Virol.*, submitted.
15. Casjens, S., Adams, M.B., Hall, C. and King, J. (1984) *J. Virol.*, submitted.
16. Gold, L., O'Farrell, P.Z. and Russel, M. (1976) *J. Biol. Chem.* 251, 7251-7262.
17. Russel, M., Gold, L., Morrisett, H. and O'Farrell, P.Z. (1976) *J. Biol. Chem.* 251, 7263-7270.
18. LeMaire, G., Gold, L. and Yarus, M. (1978) *J. Mol. Biol.* 126, 73-90.
19. Cardillo, T.S., Landry, E.F. and Wiberg, J.S. (1979) *J. Virol.* 32, 905-916.
20. Karam, J., Gold, L., Singer, B.S. and Dawson, M. (1981) *Proc. Natl. Acad. Sci. USA* 78, 4669-4673.
21. Tomizawa, J., Anraku, N. and Iwama, Y. (1966) *J. Mol. Biol.* 21, 247-253.
22. Alberts, B.M. and Frey, L. (1970) *Nature* 227, 1313-1318.
23. Huberman, J.A., Kornberg, A. and Alberts, B.M. (1971) *J. Mol. Biol.* 62, 39-52.
24. Wiberg, J.S., Mendelsohn, S., Warner, V., Hercules, K., Aldrich, C. and Munro, J.L. (1973) *J. Virol.* 12, 775-792.
25. Karam, J.D. and Bowles, M.G. (1974) *J. Virol.* 13, 428-438.
26. Karam, J.D., Dawson, M., Gerald, W., Trojanowska, M. and Alford, C. (1982) Interaction of Translational and Transcriptional Controls in the Regulation of Gene Expression. Grunberg-Manago, M. and Safer, B., eds., pp. 83-89, Elsevier/North-Holland, Amsterdam.
27. Campbell, K. and Gold, L. (1982) Interaction of Translational and Transcriptional Controls in the Regulation of Gene Expression. Grunberg-Manago, M. and Safer, B., eds., pp. 69-83, Elsevier/North-Holland, Amsterdam.
28. Trimble, R.B. and Maley, F. (1976) *J. Virol.* 17, 538-549.
29. Karam, J., McCulley, C. and Leach, M. (1977) *Virol.* 76, 685-700.
30. Nelson, M.A., Ericson, M., Gold, L. and Pulitzer, J.F. (1982) *Mol. Gen. Genet.* 188, 60-68.
31. Itakura, K., Hirose, T., Crea, R. and Riggs, A.D. (1977) *Science* 198, 1056-1063.
32. Viera, J. and Messing, J. (1982) *Gene* 19, 259-268.
33. Wilson, G.G. and Murray, N.E. (1979) *J. Mol. Biol.* 132, 471-491.
34. Doherty, D.H., Gauss, P. and Gold, L. (1982) *Mol. Gen. Genet.* 188, 77-90.
35. Miller, E.S., Power, S.D. and Gold, L. (1984) Manuscript in preparation.

36. Maxam, A.M. and Gilbert, W. (1980) *Methods in Enzymology*, Grossman, L. and Moldave, K., eds., vol. 65, 499-560.
37. Merle, P. and Kadenbach, B. (1980) *Eur. J. Biochem.* 105, 499-507.
38. Hirs, C.H.W. (1967) *Methods in Enzymology*, Colowick, S.P. and Kaplan, N.O., eds., vol. 11, pp. 59-62, Academic Press, New York.
39. Allen, G. (1981) *Laboratory Techniques in Biochemistry and Molecular Biology*, Work, T.S. and Burdon, R.H., eds., vol. 9, pp. 30-31, Elsevier/North-Holland, Amsterdam.
40. Edman, P. and Begg, G. (1967) *Eur. J. Biochem.* 1, 80-91.
41. Spicer, E.K. and Konigsberg, W.H. (1983) *Bacteriophage T4*, Mathews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B., eds., pp. 291-301, American Society for Microbiology, Washington, DC.
42. Doolittle, R.F. (1981) *Science* 214, 149-159.
43. Kyte, J. and Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132.
44. Wiberg, J.S. and Karam, J.D. (1983) *Bacteriophage T4*, Mathews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B., eds., pp. 193-201, American Society for Microbiology, Washington DC.
45. Mosig, G. (1983) *Bacteriophage T4*, Mathews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B., eds., pp. 120-130, American Society for Microbiology, Washington, DC.
46. Mathews, C.K., North, T.W. and Reddy, G.P.V. (1979) *Adv. Enz. Regulation*, Vol. 17, pp. 133-156, Pergamon Press, New York.
47. Stormo, G.D., Schneider, T.D., Gold, L. and Ehrenfeucht, A. (1982) *Nucl. Acids Res.* 10, 2997-3011.
48. Sacerdot, C., Fayat, G., Dessen, P., Springer, M., Plumbridge, J.A., Grunberg-Manago, M. and Blanquet, S. (1982) *EMBO J.* 1, 311-315.
49. Steege, D.A. (1977) *Proc. Natl. Acad. Sci. USA* 74, 4163-4167.
50. Oppenheim, D.S. and Yanofsky, C. (1980) *Genetics* 95, 785-795.
51. Napoli, C., Gold, L. and Singer, B.S. (1981) *J. Mol. Biol.* 149, 433-449.
52. Aksoy, S., Squires, C.L. and Squires, C. (1984) *J. Bacteriol.* 157, 363-367.
53. Frederick, R.J. and Snyder, L. (1977) *J. Mol. Biol.* 114, 461-476.
54. Schneider, T.D., Stormo, G.D., Haemer, J.S. and Gold, L. (1982) *Nucl. Acids Res.* 10, 3013-3024.
55. Schneider, T.D., Stormo, G.D., Yarus, M.A. and Gold, L. (1982) *Nucl. Acids Res.* 12, 129-140.
56. Tinoco, I. Jr., Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) *Nature New Biol.* 246, 40-41.
57. Karam, J., Bowles, M. and Leach, M. (1979) *Virology* 94, 192-203.
58. Piperno, J.R. and Alberts, B.M. (1978) *J. Biol. Chem.* 253, 5174-5179.
59. Nossal, N.G. and Alberts, B.M. (1983) *Bacteriophage T4*, Mathews, C.K., Kutter, E.M., Mosig, G. and Berget, P.B., eds., pp. 71-81, American Society for Microbiology, Washington, DC.
60. Nossal, N.G. (1979) *J. Biol. Chem.* 254, 6026-6031.
61. Gerald, W.L. and Karam, J.D., *Genetics*, in press.