JOURNAL OF BACTERIOLOGY, Nov. 1993, p. 7150–7159
0021-9193/93/227150-10$02.00/0

Vol. 175, No. 22

Cloning and Characterization of the RNA Polymerase α-Subunit Operon of *Chlamydia trachomatis*

MING TAN,[1,2] RUSSELL KLEIN,[1,2]† RICHARD GRANT,[2] DON GANEM,[1,2] AND JOANNE ENGEL[2]*

Departments of Microbiology and Immunology[1] and Medicine,[2] University of California, San Francisco, San Francisco, California 94143

Received 26 January 1993/Returned for modification 3 February 1993/Accepted 8 September 1993

We have cloned the chlamydial operon that encodes the initiation factor IF1, the ribosomal proteins L36, S13, and S11, and the α subunit of RNA polymerase. The genes for S11 and α are closely linked in *Escherichia coli*, *Bacillus subtilis*, and plant chloroplast genomes, and this arrangement is conserved in *Chlamydia* spp. The S11 ribosomal protein gene potentially encodes a protein of 125 amino acids with 41 to 42% identity over its entire length to its *E. coli* and *B. subtilis* homologs; the gene encoding the α subunit specifies a protein of 322 amino acids with 25 to 30% identity over its entire length to its *E. coli* and *B. subtilis* homologs. In a T7-based expression system in *E. coli*, the chlamydial α gene directed the synthesis of a 36-kDa protein. Mapping of the chlamydial mRNA transcript by RNase protection studies and by a combination of reverse transcription and the polymerase chain reaction demonstrates that IF1, L36, S13, S11, and α are transcribed as a polycistronic transcript.

Chlamydia infection is the most prevalent sexually transmitted disease and the leading cause of preventable sterility in the United States and the most common cause of noncongenital blindness in third world countries (reviewed in reference 35). Together with the rickettsiae, chlamydiae are the only known prokaryotes that are obligate intracellular parasites (reviewed in references 23 and 34). They replicate via an unusual developmental cycle involving the serial alteration between two distinct forms within the cytoplasm of the infected cell. Chlamydial development proceeds according to a strict program which is likely to reflect the temporally regulated activation of specific gene sets. Although this cycle has been well described morphologically, the molecular basis of developmentally regulated gene expression in *Chlamydia* spp. is undefined. This is largely due to the lack of convenient systems for gene transfer into this organism and the paucity of information about the nature of the signals and machinery that govern chlamydial gene expression. Initial studies aimed at determining the transcription initiation sites of several chlamydial mRNAs have shown typical eubacterial promoter sequence elements at the expected positions for the small transcript of the *Chlamydia trachomatis* plasmid (12) and the P1 putative promoter of the *C. trachomatis* cysteine-rich protein (20). However, analysis of other chlamydial genes has failed to show typical −10 and −35 sequences upstream of the mapped start sites (1, 7, 10, 11, 33). Furthermore, at least two chlamydial promoters are not correctly recognized by the transcription apparatus of *Escherichia coli* (11, 33).

A fuller understanding of chlamydial gene regulation will require a more detailed characterization of chlamydial RNA polymerase (RNAP), the central component of the transcriptional apparatus. Eubacterial RNAP is a multisubunit enzyme composed of α, β, β′, and σ subunits (reviewed in reference 13). The core enzyme, $\alpha_2\beta\beta'$, is a nonspecific DNA-binding protein that is competent for the elongation and termination of RNA transcripts. Holoenzyme, formed by the association of the σ subunit with core enzyme, demonstrates sequence-specific DNA recognition, permitting the specific binding of RNAP to promoter sequences and the initiation of mRNA transcription. Direct biochemical purification of RNAP from many bacterial species for use in in vitro transcription systems has been relatively straightforward, in part because of the ability to grow large quantities of these microorganisms. Such an approach is not practical in *Chlamydia* spp.; the poor growth of this bacterium in culture makes it exceedingly difficult to generate the necessary starting material for large-scale purifications. Instead, we have turned our efforts towards the cloning and overexpression in *E. coli* of the subunits of chlamydial RNAP. This approach will facilitate their further purification for subsequent use in vitro in the analysis of promoter structure and the identification of protein factors important in the control of chlamydial gene expression.

We have previously cloned and characterized the β and β′ subunits of RNAP (9), as well as a homolog of σ^{70} (8), from a murine strain of *C. trachomatis*. We used degenerate polymerase chain reaction (PCR) primers to regions conserved in bacterial and plant chloroplast RNAP subunits to amplify related sequences from the chlamydial genome, but such an approach was not successful for cloning the α subunit (3a). Instead, we have used a PCR to amplify the gene for the ribosomal protein S11, an upstream gene found in the same operon as the α subunit of RNAP in both *E. coli* and *Bacillus subtilis*, and have walked down the chromosome to the α gene. We report the sequences of the genes for the initiation factor IF1, the ribosomal proteins L36, S13, and S11, and the α subunit of RNAP from a murine strain of *C. trachomatis* and show that they are part of an operon. Furthermore, we demonstrate that when overexpressed in *E. coli*, the chlamydial α gene directs the synthesis of a 36-kDa protein.

MATERIALS AND METHODS

Reagents. Products were obtained from the following sources and used as described in the manufacturers' specifications. Restriction enzymes, bacterial alkaline phosphatase, exonuclease III, DNA polymerase I, T4 DNA ligase, and RNase were from New England BioLabs, Inc. (Beverly, Mass.); T4 polynucleotide kinase was from Boehringer Mannheim Biochemicals (Indianapolis, Ind.); SP6 polymerase,

* Corresponding author.
† Present address: Department of Medicine, University of California, Los Angeles, CA 90024.

RQ1 RNase, and RNasin were from Promega Biotech (Madison, Wis.); ^{32}P-containing radioisotopes were from Amersham Corp. (Arlington Heights, Ill.); [^{35}S]methionine was from ICN (Irvine, Calif.); *Thermus aquaticus* DNA polymerase was from Cetus Corp. (Emeryville, Calif.); Superscript reverse transcriptase (RT) was from GIBCO-BRL (Gaithersburg, Md.); SeaPlaque and SeaKem agarose were from FMC Bioproducts (Rockland, Me.); ampicillin, kanamycin, and rifampin were from Sigma Chemical Co. (St. Louis, Mo.); the RPAII kit was from Ambion (Austin, Tex.); and the Sequenase kit was from U.S. Biochemical Corp. (Cleveland, Ohio).

Nucleic acid preparation and analysis. Chlamydial DNA and mRNA from the mouse pneumonitis strain of *C. trachomatis* were prepared as described previously (11). Standard recombinant DNA methods were used for nucleic acid preparation and analysis (31). Restriction fragments were subcloned into a pGEM7Zf (Promega Biotech) plasmid vector. pMT70 was generated by subcloning the 1.4-kb internal *Pvu*II fragment from pMT20 (see Fig. 1) into the *Sma*I site of pGEM7Zf, so that the coding region of the α gene lay downstream of the T7 promoter. pMT103, pMT104, and pMT114 were generated by exonuclease III digestion of the 3′ end of the pMT70 insert (see Fig. 5B). pMT153 consists of the 5′ *Eco*RI-*Cla*I fragment of pMT20 (see Fig. 1) that includes the gene for IF1 and the 5′ portion of L36, cloned into the *Eco*RI and *Cla*I sites of pGEM7Zf. Southern blotting was carried out as described previously (11). Radioactive DNA probes were labelled by nick translation or by 5′-end labelling with T4 polynucleotide kinase (31).

Synthetic oligonucleotides. The following single-stranded oligonucleotide primers were synthesized at the Biomedical Resource Center at the University of California, San Francisco (nucleotides in parentheses indicate the bases used in the degenerate positions): S11 (5′ primer), GGGAATTCTT(CT)AA(CT)AA(CT)AC(AGCT)AT(ACT)GTT; S11 (3′ primer), GGCTCGAGC(TG)(AG)CA(AGCT)CC(GA)TT(GA)TG(AGCT)GGT; IF1-1 (5′ primer), GAAGCTTACATATTAT; IF1-4 (5′ primer), CCAAAATGCATTCAA; S13-2 (3′ primer), ACATTAGCATCAGCC; α-3 (3′ primer), TCTTCGTCGCTTAAGCC. Oligonucleotides homologous to the T7 and SP6 promoters were from Promega Biotech.

RT. One to two micrograms of chlamydial mRNA was pretreated with RQ1 DNase at 37°C for 1 h. Control samples were also pretreated with RNase during this step. First-strand cDNA was reverse transcribed by using the α-3 primer and Superscript RT as described in the manufacturer's specifications (see Fig. 6C). The total reaction volume was 20 μl, and at the end of the incubation, 100 μl of 10 mM Tris HCl (pH 8.0)–1 mM EDTA was added, and the reaction was stopped by heat inactivation. Five microliters of the reaction mix was used for each subsequent PCR.

PCR. PCR was performed with a Perkin-Elmer Cetus DNA thermocycler. Reactions were carried out in a volume of 100 μl and contained 100 pmol each of the 5′ and 3′ primers, 100 ng of DNA, all four deoxyribonucleotides (0.2 mM each), 10 mM Tris HCl (pH 8.0), 50 mM KCl, 1.5 to 2.5 mM $MgCl_2$, 0.1% gelatin, and 2.5 U of *T. aquaticus* DNA polymerase. Amplification was carried out for 25 cycles (denaturation at 94°C for 1 min, annealing at 37°C for 1 min, and polymerization at 72°C for 1 min). After analysis of the PCR product on a 1.5% low-melting-point agarose gel (SeaPlaque), the amplification product was recovered from the gel, digested with *Eco*RI and *Xho*I, and cloned into pGEM7Zf that had been previously digested with the same restriction enzymes and dephosphorylated with bacterial alkaline phosphatase.

RNase protection studies. RNase protection experiments were carried out with the RPAII kit as described in the manufacturer's specifications. An RNA transcript, synthesized by SP6 RNA polymerase and labelled with [α-^{32}P]rCTP (100,000 cpm), was hybridized to approximately 5 μg of chlamydial RNA at 45°C for 18 h and then subjected to RNase digestion. The products were electrophoresed with a sequencing ladder generated from an unrelated single-stranded M13 template.

Preparation and screening of a chlamydial DNA library. Construction and screening of a 1,000-colony chlamydial *Eco*RI DNA library were carried out as described previously (8). Hybond (Amersham) filters bearing colonies were hybridized to a 5′-end-labelled probe made from the S11 PCR product. Clones demonstrating hybridization to this probe were mapped by restriction endonuclease digestion.

DNA sequencing. The dideoxy chain termination method of DNA sequencing (32) was carried out with the Sequenase kit on double-stranded fragments cloned into pGEM7Zf. Sequencing reactions were primed with oligonucleotides homologous to the T7 and SP6 promoters flanking the cloned inserts in the pGEM7Zf vector. Nested deletions were generated by using exonuclease III as described in the manufacturer's specifications. The entire sequence of the IF1, L36, S13, S11, and α genes and their intergenic regions was determined on both strands. DNA and protein sequence comparisons were carried out by using the FASTA, FASTN, and TFASTA computer programs (27). Conservative changes in amino acid residues between two proteins are the same as those defined in those programs (27).

Overproduction of bacterially encoded proteins. Overproduction of the protein encoded by the cloned α gene was carried out by using a T7 RNAP expression system under the control of a thermolabile lambda repressor (41). Metabolic labelling with [^{35}S]methionine and separation of products by polyacrylamide gel electrophoresis under denaturing conditions were carried out as described previously (8).

Nucleotide sequence accession numbers. The nucleotide sequences of the IF1, L36, S13, S11, and RNAP α-subunit genes (see Fig. 2) have been deposited in the GenBank data base under accession numbers L23478 and L09636.

RESULTS

Cloning of the chlamydial S11 gene. We were unable to amplify the α subunit of *C. trachomatis* RNAP by PCR, and therefore we chose to clone a closely linked gene to find the gene for the α subunit of RNAP. In *B. subtilis*, the gene for the ribosomal protein S11 is just upstream of the α-subunit gene, separated by an intergenic region of 177 nucleotides (40). In *E. coli*, there is another ribosomal protein, S4, between S11 and the α-subunit gene, but fewer than 700 nucleotides separate the S11 and α genes (2). S11 and α are also closely linked in plant chloroplasts. If this conservation of operon structure is maintained in *Chlamydia* spp. we should be able to walk down the chromosome from the S11 to the α gene.

Given the evolutionary conservation of ribosomal proteins, we used a PCR-based approach to clone the S11 gene from *C. trachomatis*. We designed synthetic degenerate oligonucleotide PCR primers on the basis of two conserved regions between the S11 proteins from *E. coli* (2), *B. subtilis* (40), *Bacillus stearothermophilus* (18), and the chloroplasts of the common tobacco plant (36), maize (22), and liverwort (25) (see Materials and Methods). In a PCR containing these primers and DNA from the murine strain of *C. trachomatis*, a product of the expected size (300 bp) was generated and cloned into the vector pGEM7Zf. The DNA sequence of the cloned PCR

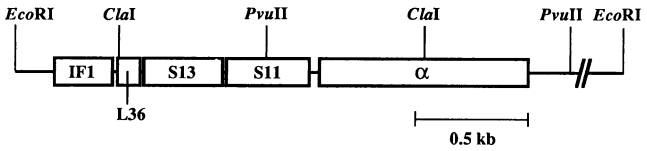

FIG. 1. Restriction map of clone pMT20 containing the full-length genes for IF1, L36, S13, S11, and α.

product was determined and found to bear 44% identity to *E. coli* S11, suggesting that the PCR product corresponds to the chlamydial homolog of prokaryotic S11 (data not shown).

To prove that the PCR product was of chlamydial origin, it was 5′ end labelled with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP and used as a hybridization probe for chlamydial DNA digested with *Eco*RI. A 4-kb *Eco*RI fragment hybridized to the probe (data not shown). Hybridization to Southern blots of chlamydial DNA digested with either of the infrequently cutting enzymes *Cla*I and *Nsi*I also revealed a single hybridization band for each, suggesting that the sequences detected are present in a single copy in the chlamydial genome. The 5′-end-labelled probe was then hybridized to an *Eco*RI library of chlamydial DNA cloned into the plasmid vector pGEM7Zf. Five clones, containing chlamydial DNA that hybridized to the S11 probe, were found after screening 1,000 colonies. Restriction enzyme analysis showed that all five clones contained the same 4-kb *Eco*RI insert and that this fragment comigrated with the genomic *Eco*RI fragment observed in the Southern blot analysis (data not shown).

The gene encoding the chlamydial ribosomal protein S11 is interrupted by a TGA codon. Figure 1 shows a restriction map of one of the five positive clones, pMT20. Hybridization of the 5′-end-labelled PCR probe demonstrated that the S11 coding region encompassed the upstream *Pvu*II site. DNA sequencing showed an open reading frame that potentially encodes a protein of 125 amino acids (Fig. 2) with 41% identity and 86% similarity (identical and conservatively substituted residues) to *E. coli* S11 (Fig. 3) and 42% identity and 86% similarity to *B. subtilis* S11 (data not shown). Of note, however, is the presence of a TGA termination codon at residue 37 (Fig. 2, nucleotides 1046 to 1048). To rule out the possibility that this TGA codon was a cloning artifact, this region was sequenced from three other genomic clones containing the 4-kb *Eco*RI insert, and the TGA codon was shown to be present in each. The existence of a second copy of the S11 gene in the chlamydial genome is unlikely, since digests of chlamydial DNA cleaved separately with *Eco*RI, *Cla*I, or *Nsi*I hybridized with an S11-specific probe and gave a single band of hybridization in each case, even under nonstringent hybridization conditions (data not shown). The significance of this TGA codon will be discussed later.

The chlamydial α-subunit gene is located 34 bp downstream of the S11 gene. We determined the DNA sequence of the region downstream of the S11 open reading frame and compared the derived amino acid sequences with the GenBank data base (release number 72) for homology to the α subunit of RNAP. There is a large open reading frame (Fig. 2) beginning just 34 nucleotides downstream of the end of the S11 gene that could encode a protein of 322 amino acids with 25% identity and 68% similarity to the α subunit of RNAP in *E. coli*. Similar sequence conservation with *B. subtilis* (30% identity and 70% similarity) and plant chloroplast RNAP α subunits is also noted (Fig. 4). These findings strongly suggest that this open reading frame downstream of the S11 gene encodes the α subunit of *C. trachomatis* RNAP. Preceding the methionine initiator codon of the α gene is a potential partial Shine-Dalgarno sequence (AGAGG); no identifiable ribosome binding site is present upstream of the S11 gene (Fig. 2). The 34-bp intergenic region contains no significant stem-loop structures suggestive of an attenuator. No T-rich region that would be indicative of a rho-independent transcription terminator is present downstream of the α coding region. No significant open reading frames initiated by a methionine codon are identifiable in the 180 bp we have sequenced downstream of the α coding region.

The chlamydial IF1, L36, and S13 genes are located upstream of the S11 gene. Nine hundred thirty base pairs were sequenced upstream of the S11 coding region, and three open reading frames with respective homology to the initiation factor IF1 and ribosomal proteins L36 and S13 were identified. The chlamydial IF1 gene potentially encodes a protein of 90 amino acids with 33% identity and 63% similarity to *E. coli* IF1 (Fig. 3) and 33% identity and 71% similarity to *B. subtilis* IF1 (data not shown). The chlamydial ribosomal protein L36 gene could encode a protein of 37 amino acids that is very highly conserved compared with prokaryotic L36 proteins (70% identity and 92% similarity to its *E. coli* homolog as shown in Fig. 3; 62% identity and 99% similarity to *B. subtilis* L36). The putative product of the chlamydial S13 gene is a protein of 119 amino acids with 47% identity and 83% similarity to *E. coli* S13 (Fig. 3) and 54% identity and 86% similarity to *B. subtilis* S13. Of these three upstream genes, only the L36 gene is preceded by a potential partial Shine-Dalgarno sequence (Fig. 2, AGGAG). These genes are all closely spaced to each other and to S11; the largest intergenic distance is the 19 bp between IF1 and L36. No significant stem-loop structures are apparent within these very short intergenic regions. The arrangement of the five genes, IF1, L36, S13, S11, and α, is the same as it is in *B. subtilis* (3) and will be discussed further.

The α-subunit gene of chlamydial RNAP directs the synthesis of a 36-kDa protein in *E. coli*. The 1.4-kb internal *Pvu*II fragment from pMT20 (Fig. 1) was subcloned into pGEM7Zf so that the coding region of the α gene lay downstream of the bacteriophage T7 promoter (Fig. 5B, plasmid pMT70). Two other plasmids containing C-terminal deletions of the α gene were also constructed: pMT103 is missing the C-terminal 27% of the coding region and pMT104 is missing the C-terminal 58% (Fig. 5B). These plasmids and a pGEM7Zf control plasmid were separately introduced into a strain of *E. coli* (41) containing a pGEM-compatible plasmid, pGP1-2, that expresses the T7 RNAP gene under the control of a lambda promoter and a thermolabile lambda repressor (*c*I857). Upon thermoinduction, the plasmid harboring wild-type (wt) α sequences, pMT70, directed the synthesis of an approximately 36-kDa protein (Fig. 5A, lane 3) that was not seen without heat induction (Fig. 5A, lane 4) or with the control pGEM7Zf plasmid alone (Fig. 5A, lane 1). The C-terminal deletion plasmids produced correspondingly smaller proteins than did the wild-type α plasmid upon heat induction. pMT103 (27% C-terminal deletion, 241-amino-acid open reading frame) directed the synthesis of an approximately 25-kDa protein (Fig. 5A, lane 5), and pMT104 (58% C-terminal deletion, 140-amino-acid open reading frame) produced an approximately 16-kDa protein (Fig. 5A, lane 7). In each case, the truncated protein was present in larger quantities than the wild-type protein and was absent without heat induction (Fig. 5A, lanes 6 and 8). Thus, the observed molecular weight of the chlamydial α protein overexpressed in *E. coli* corresponds well with the predicted molecular weight.

The genes for IF1, L36, S13, S11, and α comprise an operon that is actively transcribed. We were unable to detect tran-

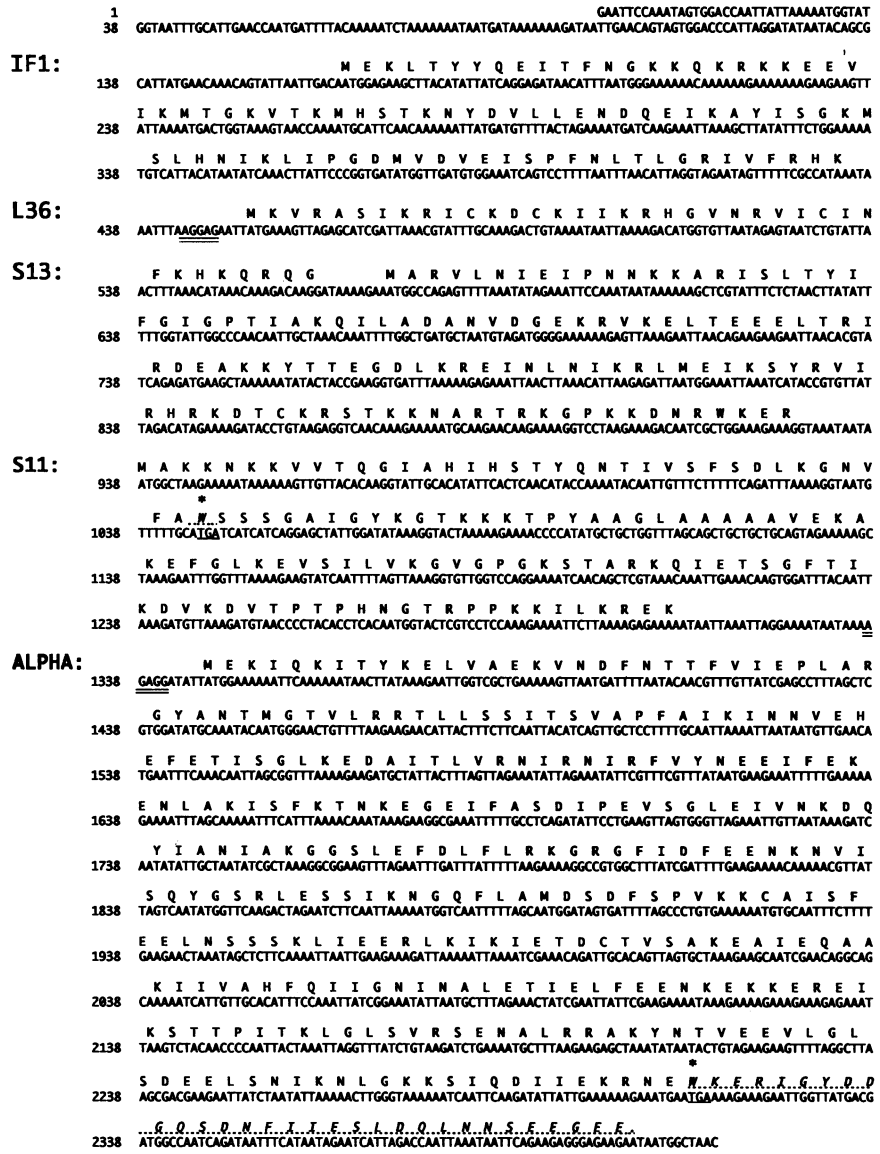

FIG. 2. Nucleotide and predicted amino acid sequence of the *C. trachomatis* initiation factor IF1, the ribosomal proteins L36, S13, and S11, and RNAP α-subunit (ALPHA) genes. Only the coding strand is shown. Possible ribosome binding sites 4 nucleotides upstream of the start codon of the L36 and 5 nucleotides upstream of the start codon of the α genes are doubly underlined. Two tryptophan residues putatively coded for by TGA are shown by an asterisk at amino acid 37 of S11 and at amino acid 323 of the α coding region. These putative residues and the putative 31-amino-acid C-terminal tail of α are shown in italics and underlined with dashes. The TGA codons are underlined. The numbers to the left refer to the nucleotide sequence.

scripts from the S11 and/or α genes by Northern (RNA) blot analysis (data not shown); this negative finding probably reflects the low abundance of these transcripts. However, RNase protection and RT-PCR studies, described below, suggest that IF1, L36, S13, S11, and α mRNA sequences are present on a single mRNA molecule and thus constitute part or all of an operon.

Initial RNase protection studies used a ^{32}P-labelled RNA transcript produced in vitro from a template extending from the *Pvu*II site in S11 through the first 163 bp of the α gene (Fig. 1). This product was fully protected from RNase digestion when hybridized to chlamydial mRNA (data not shown).

RT-PCR studies showed that cDNA primed from chlamydial mRNA with a primer complementary to the very 3′ end of the α gene (α-3) could serve as a template for PCR by using a 5′ primer as far upstream as the IF1 gene (Fig. 6C). Thus, for example, a PCR with the IF-1 (5′) and S13-2 (3′) primers produced a 500-bp product (Fig. 6B, lane 2), and a PCR with the IF1-4 (5′) and S13-2 (3′) primers showed a 450-bp product (Fig. 6B, lane 5). In each case, the PCR product obtained from this cDNA template comigrated on agarose electrophoresis with the PCR product generated from a chlamydial genomic DNA template (Fig. 6B, lanes 1 and 4). No product was observed when the mRNA was pretreated with RNase before RT (Fig. 6B, lanes 3 and 6). This rules out the possibility that the cDNA template was derived from chlamydial chromosomal DNA that might be contaminating the RNA preparation rather than from chlamydial mRNA. In separate PCRs, the

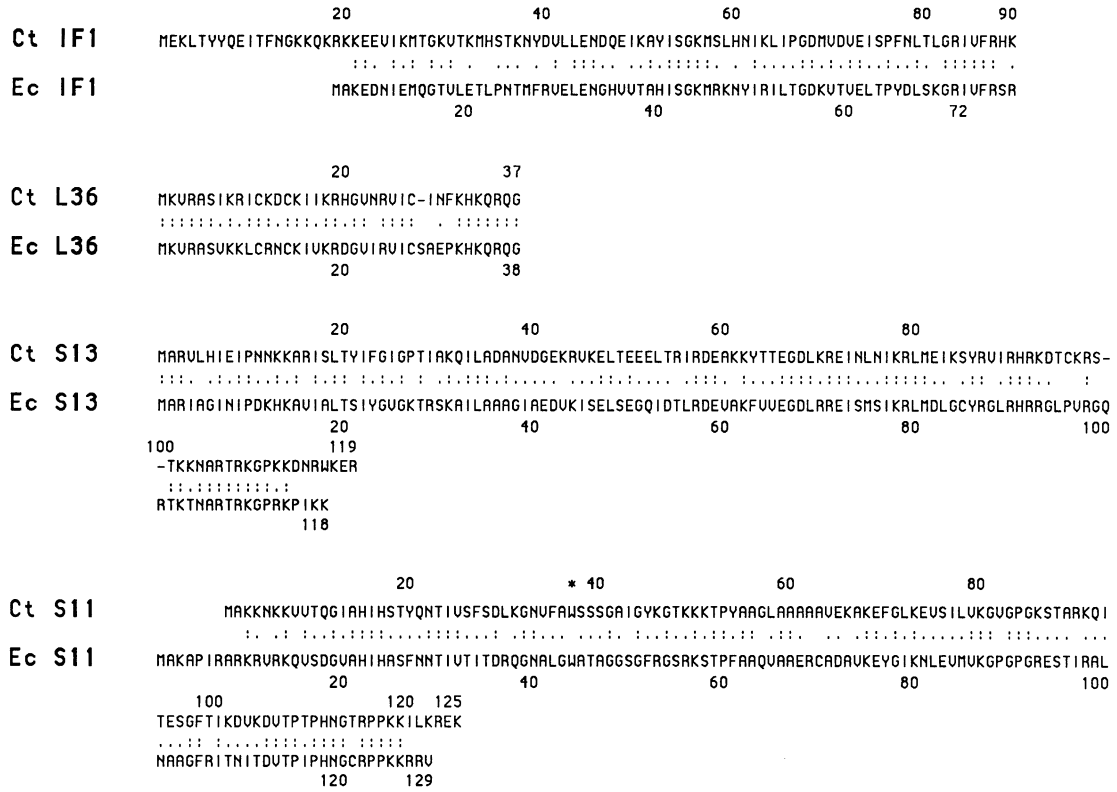

FIG. 3. Comparison of the amino acid sequences derived from the genes for initiation factor 1 (IF1) and the ribosomal proteins L36, S13, and S11 of *C. trachomatis* (Ct) and their *E. coli* (Ec) homologs. Symbols: :, an identical residue; ., a conserved amino acid substitution; –, a gap that was introduced for purposes of alignment. The putative tryptophan residue at position 37 of *C. trachomatis* S11 is marked with an asterisk. Accompanying numerals refer to the numbering of amino acid residues for each sequence.

entire distance between IF1 and α could be amplified from the α-3-primed cDNA by using pairs of primers that gave overlapping PCR products (data not shown), suggesting the existence of a continuous mRNA that includes the transcript for IF1, L36, S13, S11, and α.

We have attempted to map the exact transcription initiation site of this polycistronic mRNA in the following way. RNase protection experiments were carried out by hybridizing a radiolabelled RNA transcript, extending upstream from the *Cla*I site of L36, to isolated chlamydial mRNA (Fig. 6C, pMT153). Two protected sets of transcripts were observed: a set of products 245 to 247 bp in length, and a fainter set of five products 223 to 227 bp in length (Fig. 6A, lane 1). No RNA was protected when uninfected HeLa cell RNA alone was used for hybridization (Fig. 6A, lane 2). Thus, the most 5′ end of the chlamydial message detected by RNase digestion analysis is located 225 to 227 bp upstream of the start of the coding sequence for L36, within the coding region of IF1 (Fig. 6D). No obvious *E. coli* consensus promoter sequences were present at the usual −10 and −35 positions. However, this putative transcription initiation site lies downstream of the IF1-1 PCR primer that was used successfully as the 5′ primer in the RT-PCR experiment described above (Fig. 6D). This discrepancy will be discussed below.

DISCUSSION

The genes for IF1, L36, S13, S11, and α are arranged in an operon in which the gene order is conserved between *C. trachomatis* and *B. subtilis* (3). An additional gene for the ribosomal protein L17 is located downstream of α in *B. subtilis* as part of the same operon (3). Sequencing 180 bp 3′ to the chlamydial α gene has failed to reveal any open reading frames, but we have not carried out an exhaustive DNA sequence analysis to determine whether L17 is present downstream of α in *Chlamydia* spp. The equivalent *E. coli* operon is similar but not identical, with the gene order SecY-L36-S13-S11-S4-α-L17 (2). The gene for the ribosomal protein S4 separates the genes for S11 and α in *E. coli* but is absent in both *C. trachomatis* and *B. subtilis*. S4 has been shown to be important for the translational regulation of all of the genes in this operon in *E. coli* with the exception of the α cistron (42, 45). In plant chloroplast genomes, the S11 gene is also located a short distance upstream of the α gene (14, 25, 28, 36, 37). Of the chloroplasts, the arrangement of the chlamydial IF1-α region is most similar to that of chloroplasts of rice, in which the gene order is IF1-L36-S11-α, with the genes for S13 and L17 absent. Conservation of operon structure across bacterial species has been noted for other operons, including the *rpsU-dnaG-rpoD* operon which contains the genes for the ribosomal protein S21, primase, and σ^{70}, respectively (44), and suggests a requirement for coordinate regulation of the genes in the conserved operon.

Sequence analysis of the S11 gene surprisingly reveals the presence of a TGA codon at amino acid 37. We have shown that this is not a sequencing or cloning artifact and that we can detect only one copy of the S11 gene in the chlamydial

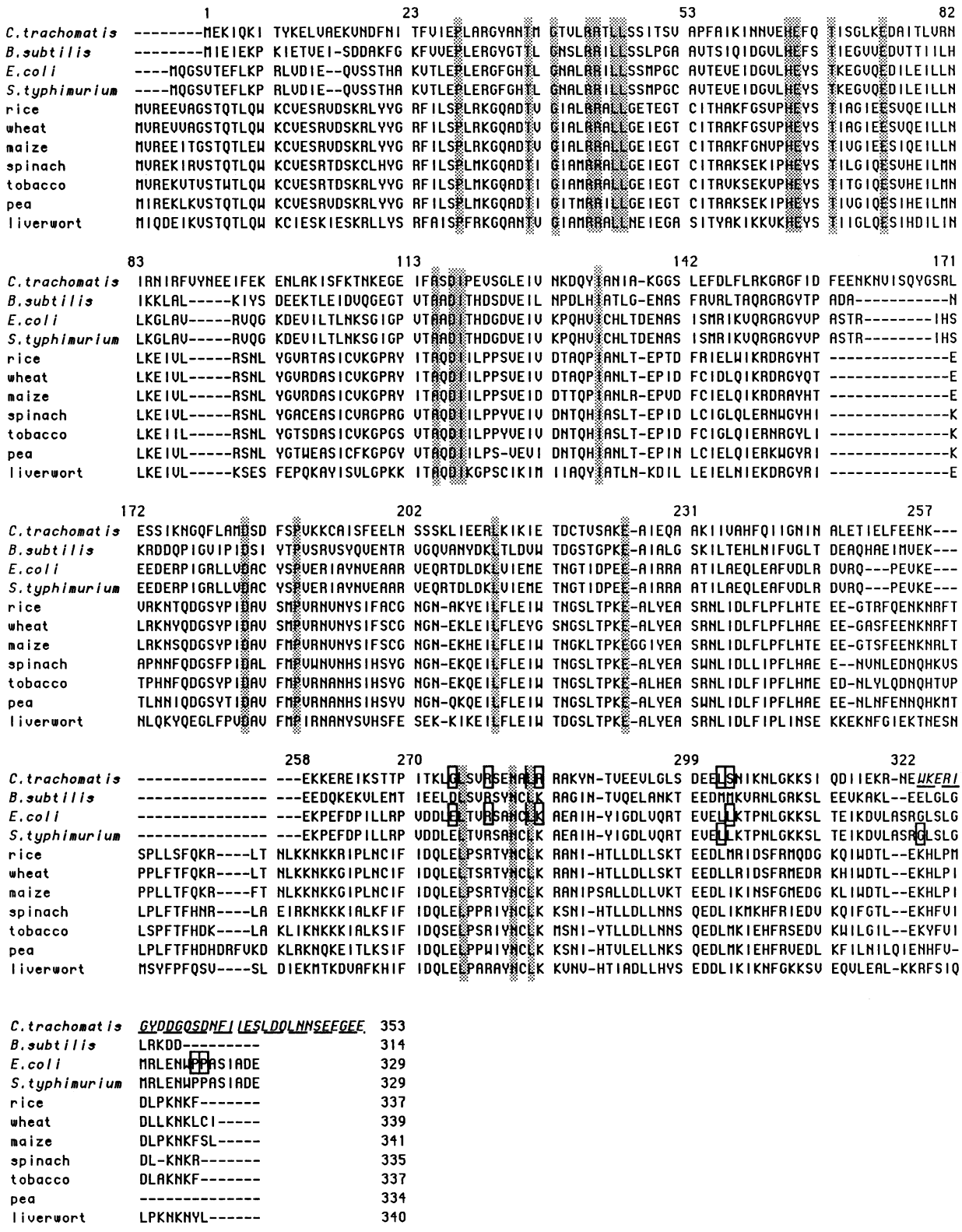

FIG. 4. Comparison of the amino acid sequences derived from the RNAP α genes of *C. trachomatis*, *B. subtilis* (3), *E. coli* (2), *S. typhimurium* (18), and the chloroplasts of rice (14), wheat (15), maize (29), spinach (37), tobacco (36), pea (28), and liverwort (25). Amino acid residues identical at a given position in all of these organisms are shaded. A gap that was introduced for purposes of alignment is marked with a dash. Numbers above the amino acid sequences refer to the *C. trachomatis* α protein sequence. Numbers at the end of each amino acid sequence refer to the total number of amino acids in the respective protein. Individual residues that were discussed in the text are boxed. The putative 31-amino-acid C-terminal tail of α is shown in italics and underlined with dashes.

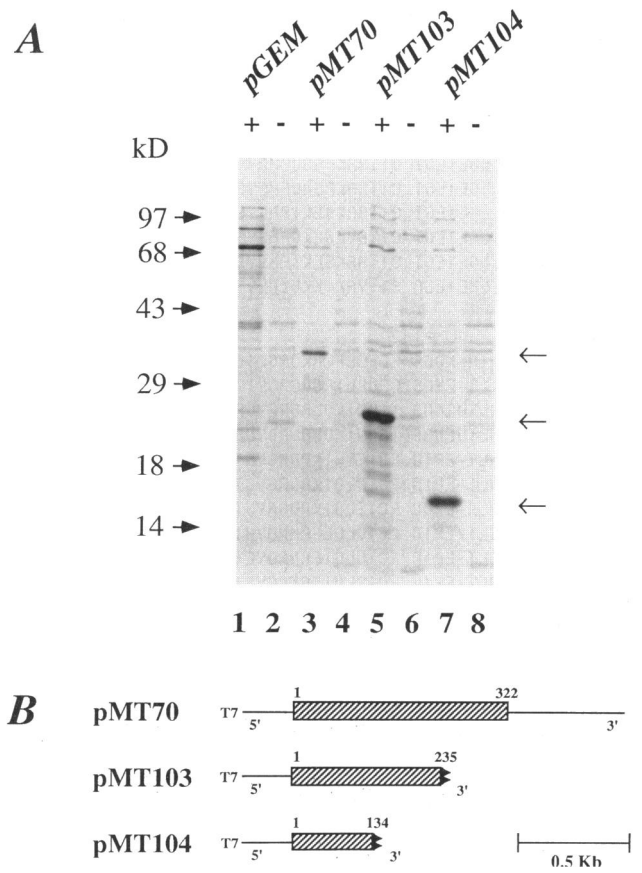

FIG. 5. (A) Autoradiograph showing expression of the α protein in *E. coli*. Plasmid-encoded gene products were expressed in a strain of *E. coli* containing pGP1-2 and one of the following plasmids: pGEM7Zf (lanes 1 and 2), the full-length α clone pMT70 (lanes 3 and 4), and two truncated α clones, pMT103 (lanes 5 and 6) and pMT104 (lanes 7 and 8). The proteins in lanes 1, 3, 5, and 7 were from bacteria that had been thermoinduced for plasmid protein expression from the T7 polymerase promoter in pGEM7Zf (indicated by + above the lanes), while lanes 2, 4, 6, and 8 were from bacteria that were not thermoinduced (indicated by − above the lanes). A 1-ml portion of cells was pelleted, the cells were suspended in 100 μl of Laemmli buffer, and 10 μl was electrophoresed on a 12% polyacrylamide gel. The gel was dried down and exposed to X-Omat film. Protein sizes in kilodaltons are indicated on the left. Arrows on the right indicate the wild-type and truncated α proteins in their respective lanes. (B) Map of the clones used for expression of the α protein in *E. coli*. The shaded region represents the coding region of the α gene; the hatched ends mark the extent of the remaining coding region in the truncated constructs. The construction of the clones is described in the text.

chromosome. RNase protection studies demonstrate that this gene is transcribed.

Several hypotheses might explain the presence of a TGA codon within the S11 gene. First, S11 may not be synthesized in *C. trachomatis*. We do not know whether S11 is an essential gene in *Chlamydia* spp., although *E. coli* 30S ribosomal subunits assembled without S11 can function in vitro (24). However, omission of S11 alone, but no other single 30S ribosomal protein, caused decreased translational fidelity. A second possibility is that only truncated S11 proteins are translated. Arguing against this is the observation that the similarity between the chlamydial S11 protein and those of other prokaryotes is equally conserved both proximal and distal to the TGA codon. In addition, the truncated protein terminating at the TGA codon would be less than one-third of the length of full-length S11. A third hypothesis is that full-length S11 protein is synthesized, but this possibility requires a mechanism for the read-through of the TGA codon. This could occur by means of RNA editing or translational read-through.

Suppression of the TGA codon by an RNA editing process would involve changing the UGA to another codon (4). For example, the UGA could be changed to a UGG and thus code for a tryptophan residue. However, sequencing of the chlamydial RNA transcript demonstrates that the UGA codon is present in the mRNA (41a).

Translational read-through of the TGA codon could occur by misreading, mediated by either the ribosome or tRNA anticodon wobble, by a suppressor tRNA, or by an altered genetic code. As mentioned earlier, ribosomes lacking S11 have been shown to have a lower translational fidelity. The accumulation of ribosomes lacking S11 or containing truncated S11 proteins could increase the incidence of misreading of the internal TGA stop codon within the S11 gene. Misincorporation of an amino acid at the TGA codon could then allow for some translation of full-length S11 protein. Such a negative feedback loop might represent a form of regulatory control of S11. However, it should be noted that this decreased translational fidelity would most likely be quite general and not limited to the misreading of a TGA stop codon. This model awaits experimental verification.

Translational read-through mediated by a tRNA raises several interesting possibilities. While wobble at the third position of the anticodon has been well described, it would be an inefficient method of suppression, especially for a ribosomal protein that can be expected to be highly expressed throughout most of the chlamydial life cycle. However, a suppressor tRNA that can recognize the TGA codon and incorporate an amino acid at this position could function as a relatively efficient opal (TGA) suppressor. An example of a tRNA species that recognizes TGA by means of a novel CCA anticodon and is aminoacylated by a tryptophan residue has been described in some mycoplasma species (26) and in mitochondria (17). In light of these findings, the presence of a conserved tryptophan at the residue corresponding to the TGA codon in all of the characterized prokaryotic S11 proteins is intriguing (Fig. 3, *E. coli*) (2, 22, 40).

One of the main distinctions between suppression and an altered genetic code would be whether the phenomenon occurs all or just some of the time. No coding region for a chlamydial protein is known unequivocally to end in a TGA codon, and by far the most common termination codon in *Chlamydia* spp. is TAA (41b). We note with interest that the coding region of the chlamydial α RNAP gene ends in a TGA codon. If this TGA codon is not recognized as a termination codon, but instead is read through, an additional 31 amino acids of α would be translated before the next in-frame termination codon (TAA; Fig. 4). Arguing against this possibility is the complete lack of homology of this α tail to any prokaryotic α homologs in the current data base. However, this putative C-terminal tail has a very high proportion of acidic residues that might be involved in protein-protein interactions with other holoenzyme subunits such as σ or with transcription activators. In particular, the vegetative chlamydial σ factor, σA, is noticeably more basic (calculated pI, 9.37) than its *E. coli* homolog, σ70 (8, 19). We are currently carrying out experiments to determine if the chlamydial α protein contains the C-terminal tail. It should be noted that expression of the chlamydial α subunit in *E. coli* (Fig. 5A) would give a protein

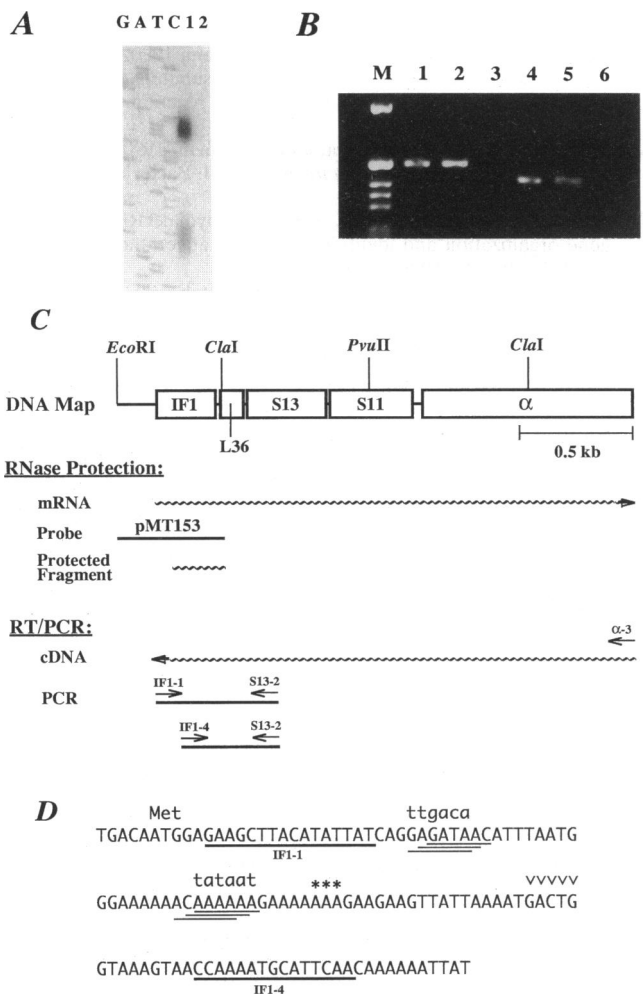

FIG. 6. Mapping of the transcription initiation site of the IF1-L36-S13-S11-α operon. (A) RNase protection experiments were carried out as described in Materials and Methods. Chlamydial RNA isolated at 18 h postinfection (lane 1) or RNA isolated from uninfected HeLa cells (lane 2) was hybridized to a radiolabelled mRNA transcript generated by SP6 transcription of pMT153. The products of the RNase digestion reaction were electrophoresed on a 5% acrylamide–8 M urea sequencing gel next to a sequencing ladder generated from an unrelated single-stranded M13 template (lanes GATC). (B) Gel electrophoresis of PCR products on a 1.5% low-melting-point agarose gel. Primers for lanes 1 to 3 were IF1-1 (5′) and S13-2 (3′). Primers for lanes 4 to 6 were IF1-4 (5′) and S13-2 (3′). The templates were chlamydial genomic DNA (lanes 1 and 4), cDNA reverse transcribed from chlamydial mRNA by using the α-3 primer (lanes 2 and 5), and cDNA prepared from chlamydial mRNA pretreated with RNase prior to RT with the α-3 primer (lanes 3 and 6). The molecular weight marker lane is indicated by an M. (C) Genomic map of the full-length genes for IF1, L36, S13, S11, and α. The results shown in panels A and B are diagrammed. The polycistronic mRNA and plasmid pMT153 that were used in the RNase protection experiment, together with the fragment protected from RNase digestion, are shown. For the RT-PCR experiments, the cDNA reverse transcribed from the polycistronic mRNA by using the α-3 primer, together with the primers used for PCR with this cDNA template and the products obtained, is shown. (D) DNA sequence of the initiator methionine and upstream coding region of the IF1 gene. The location of the 5′ end of the larger transcript as measured by RNase protection and shown in panel A is indicated by asterisks. The location of the 5′ end of the smaller RNase protection products is indicated by inverted carets. The putative −10 and −35 regions are underlined, and the σ^{70} consensus sequences are shown in small type above them. The locations and sequences of the 5′ primers used for RT-PCR, i.e., IF1-1 and IF1-4, are underlined with a bold line and labelled.

that is terminated at the TGA stop codon, regardless of how this codon is read in *Chlamydia* spp.

Prior studies of point mutants and deletion mutants of *E. coli* α suggest that the N-terminal portion of the α protein is involved in the assembly of the core RNAP, while the C-terminal 94 residues appear to define a region that contacts some positive regulators of gene transcription (reviewed in references 16 and 30). Point mutations in *E. coli* α that affect the activation of transcription by transcription factors have identified individual residues of α as potential contact sites for these activators. Most of these residues are conserved in *C. trachomatis* α with one notable exception (Fig. 4, see the derived amino acid sequence of *C. trachomatis* α). Thus, amino acids 261, 265, and 270 of *E. coli* α are essential for the correct interaction of α with cyclic AMP activator protein (CAP) at class I CAP-dependent promoters (6). While the residues corresponding to positions 265 and 270 of *E. coli* α are unchanged in *C. trachomatis* α (Arg-278 and Leu-283), there is a glycine at amino acid 274 of chlamydial α that corresponds to glutamic acid at position 261 in *E. coli* α. This substitution is of particular interest because the replacement of glutamic acid at position 261 with glycine is one of several single-amino-acid-substitution mutants in *E. coli* α that exhibits a loss of response to CAP at a class I CAP-dependent promoter in *E. coli* (6). Whether this substitution has an effect in *Chlamydia* spp. is unknown, since to date neither a homolog of CAP nor CAP-dependent promoters have been shown to exist in *Chlamydia* spp.

Single-amino-acid substitutions in *E. coli* α have also been shown to affect transcription activation by other factors. Lys-271 of *E. coli* α is involved in interactions with the activators CysB, AraC, and MelR in *E. coli* (43), and its corresponding residue in *C. trachomatis* α is Arg-284, a conservative change. A mutation in Leu-290 of *E. coli* α prevents growth of bacteriophage P2 by inhibiting the expression of the late phage genes which are normally regulated by the *ogr* gene product. This mutation also inhibits activation of the functionally related δ gene product of the satellite phage P4 (Christie and Ayres, cited in reference 21). The residue corresponding to *E. coli* Leu-290 is Ser-303 in *C. trachomatis* α, a nonconservative amino acid substitution. Leu-289 of *Salmonella typhimurium* α (identical to *E. coli* α) is involved in interactions with OxrA (21), an activator protein for genes induced by anaerobiosis. It is conserved as Leu-302 in *C. trachomatis* α. We cannot comment on *S. typhimurium* α residue Gly-311, which is also involved in activation by OxrA, because it is located at the extreme C terminus of *E. coli* α, beyond the C terminus of *C. trachomatis* α (as defined by the location of the TGA codon). Similarly, we cannot make any inferences about Pro-322 and Pro-323 of *E. coli* α, which are involved in OmpR activation (38). The studies of mutant proteins in *E. coli*, in conjunction with protein modeling, suggest that some activators directly contact the C terminus of α at particular sites (16, 30). Although no homologs of these activators have been described to date in *Chlamydia* spp., the work in *E. coli* to define activator contact sites on α can help define the region of chlamydial α that might interact similarly with transcription activators.

The results of RT-PCR and RNase protection experiments to map the start of the α transcript gave conflicting results. We note that we have successfully used RT-PCR analysis to

confirm the location of the chlamydial major outer membrane protein transcript (6a). The initiation site for this outer membrane transcript had previously been mapped by a combination of RNase protection studies, S1 nuclease analysis, and primer extension studies in several chlamydial species and strains (7, 39). Taken together, however, the RT-PCR and RNase protection results can be explained if (i) the IF1, L36, S13, S11, and α genes are part of an operon in *C. trachomatis*, (ii) transcription initiation occurs upstream of IF1, either from a putative promoter for this operon or by read-through from an upstream gene, and (iii) the protected transcript, whose 5′ end mapped within the IF1 gene in the RNase protection experiments, represents a stable, processed intermediate of a larger transcript or a second transcription initiation site. In *E. coli*, a larger proportion of the α transcript has been reported to originate from the strong *spc* promoter of the upstream *spc* operon (5) than from a secondary promoter that immediately precedes S13 (20a). Our results would also be consistent with this model of two promoters for the α gene. In *B. subtilis*, the transcript initiation site for the polycistronic message lies upstream of IF1 but has not been mapped (3).

Now that we have cloned all of the subunits of chlamydial RNAP, we are directing our efforts towards their overexpression in *E. coli* with the goal of reconstituting chlamydial RNAP in vitro from the purified subunits. Such a system will allow us, in the absence of a viable chlamydial genetic system, to study the promoter elements and *trans*-acting factors that underlie the developmental regulation of the chlamydial life cycle. We are also conducting further experiments to determine whether TGA can code for an amino acid in chlamydial translation.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (AI24436 to J.N.E. and D.G.), the University of California, San Francisco, academic senate (J.N.E.), and the Lucille P. Markey Charitable Trust (88-36 to J.N.E.). J.N.E. is a Lucille P. Markey scholar. M.T. is a recipient of a Merck, Sharp and Dohme-University of California, San Francisco, Molecular Medicine Fellowship.

REFERENCES

1. **Allen, J. E., and R. S. Stephens.** 1989. Identification by sequence analysis of two-site posttranslational processing of the cysteine-rich outer membrane protein 2 of *Chlamydia trachomatis* serovar L2. J. Bacteriol. **171:**285–291.
2. **Bedwell, D., G. Davis, M. Gosink, L. Post, M. Nomura, H. Kestler, J. M. Zengel, and L. Lindahl.** 1985. Nucleotide sequence of the alpha ribosomal protein operon of *Escherichia coli*. Nucleic Acids Res. **13:**3891–3903.
3. **Boylan, S. A., J.-W. Suh, S. M. Thomas, and C. W. Price.** 1989. Gene encoding the alpha core subunit of *Bacillus subtilis* RNA polymerase is cotranscribed with the genes for initiation factor 1 and ribosomal proteins B, S13, S11, and L17. J. Bacteriol. **171:** 2553–2562.

3a. **Brodsky, M., M. Tan, and J. Engel.** Unpublished results.

4. **Cattaneo, R.** 1990. Messenger RNA editing and the genetic code. Experientia **46:**1142–1148.
5. **Cerretti, D. P., D. Dean, G. R. Davis, D. M. Bedwell, and M. Nomura.** 1983. The *spc* ribosomal protein operon of *Escherichia coli*: sequence and contranscription of the ribosomal protein genes and a protein export gene. Nucleic Acids Res. **11:**2599–2616.
6. **Ebright, R. H.** 1993. Transcription activation at class I CAP-dependent promoters. Mol. Microbiol. **8:**797–802.

6a. **Engel, J.** Unpublished observations.

7. **Engel, J., and D. Ganem.** 1990. Identification and comparison of putative chlamydial promoter elements, p. 245–260. *In* L. Van der Ploeg (ed.), Immune recognition and evasion: molecular aspects of host parasite interaction. Academic Press, Inc., San Diego, Calif.
8. **Engel, J., and D. Ganem.** 1990. A polymerase chain reaction-based approach to cloning sigma factors from eubacteria and its application to the isolation of a sigma-70 homolog from *Chlamydia trachomatis*. J. Bacteriol. **172:**2447–2455.
9. **Engel, J., J. Pollack, F. Malik, and D. Ganem.** 1990. Cloning and characterization of RNA polymerase core subunits of *Chlamydia trachomatis* by using the polymerase chain reaction. J. Bacteriol. **172:**5732–5741.
10. **Engel, J., J. Pollack, E. Perara, and D. Ganem.** 1990. Heat shock response of murine *Chlamydia trachomatis*. J. Bacteriol. **172:**6959–6972.
11. **Engel, J. N., and D. Ganem.** 1987. Chlamydial rRNA operons: gene organization and identification of putative tandem promoters. J. Bacteriol. **169:**5678–5685.
12. **Fahr, M. J., K. S. Sriprakash, and T. P. Hatch.** 1992. Convergent and overlapping transcripts of the *Chlamydia trachomatis* 7.5-kb plasmid. Plasmid **28:**247–257.
13. **Helmann, J. D., and M. J. Chamberlin.** 1988. Structure and function of bacterial sigma factors. Annu. Rev. Biochem. **57:**839–872.
14. **Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamato, M. Mori, C. Kondo, Y. Honji, C.-R. Sun, B.-Y. Meng, Y.-Q. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki, and M. Sugiura.** 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. **217:**185–194.
15. **Hird, S. M., T. A. Dyer, and J. C. Gray.** 1989. Nucleotide sequence of the *rpoA* gene in wheat chloroplast DNA. Nucleic Acids Res. **17:**6394.
16. **Ishihama, A.** 1992. Role of the RNA polymerase α subunit in transcription activation. Mol. Microbiol. **6:**3283–3288.
17. **Jukes, T. H., and S. Osawa.** 1990. The genetic code in mitochondria and chloroplasts. Experientia **46:**1117–1125.
18. **Kimura, M., J. Kimura, and T. Hatakeyama.** 1988. Amino acid sequences of ribosomal proteins S11 from *Bacillus stearothermophilus* and S19 from *Halobacterium marismortui*. FEBS Lett. **240:**15–20.
19. **Koehler, J. E., R. R. Burgess, N. E. Thompson, and R. S. Stephens.** 1990. *Chlamydia trachomatis* RNA polymerase major σ subunit. J. Biol. Chem. **265:**13206–13214.
20. **Lambden, P. R., J. S. Everson, M. E. Ward, and I. N. Clarke.** 1990. Sulfur-rich proteins of *Chlamydia trachomatis*: developmentally regulated transcription of polycistronic mRNA from tandem promoters. Gene **87:**105–112.

20a. **Lindahl, L., F. Sor, R. H. Archer, M. Nomura, and J. M. Zengel.** 1990. Transcriptional organization of the S10, pc and alpha operons of *Escherichia coli.*, Biochim. Biophys. Acta **1050:**337–342.

21. **Lombardo, M.-J., D. Bagga, and C. G. Miller.** 1991. Mutations in *rpoA* affect expression of anaerobically regulated genes in *Salmonella typhimurium*. J. Bacteriol. **173:**7511–7518.
22. **Markmann-Mulisch, U., and A. R. Subramanian.** 1988. Nucleotide sequence of maize chloroplast *rps11* with conserved amino acid sequence between eukaryotes, bacteria and plastids. Biochem. Int. **17:**655–664.
23. **Moulder, J. W.** 1991. Interaction of chlamydiae and host cells in vitro. Microbiol. Rev. **55:**143–190.
24. **Nomura, M., S. Mizushima, M. Ozaki, P. Traub, and C. V. Lowry.** 1969. Structure and function of ribosomes and their molecular components. Cold Spring Harbor Symp. Quant. Biol. **34:**49–61.
25. **Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, U. Kazuhiko, Y. Shiki, M. Takeuchi, Z. Chang, S.-I. Aota, H. Inokuchi, and H. Ozeki.** 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature (London) **322:**572–574.
26. **Osawa, S., A. Muto, T. Ohama, Y. Andachi, R. Tanaka, and F. Yamao.** 1990. Prokaryotic genetic code. Experientia **46:**1097–1106.
27. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
28. **Purton, S., and J. C. Gray.** 1989. The plastid *rpoA* gene encoding a protein homologous to the bacterial RNA polymerase alpha subunit is expressed in pea chloroplasts. Mol. Gen. Genet. **217:** 77–84.

29. **Ruf, M., and H. Kossel.** 1988. Structure and expression of the gene coding for the α-subunit of DNA-dependent RNA polymerase from the chloroplast genome of *Zea mays*. Nucleic Acids Res. **16:**5741–5754.
30. **Russo, F. D., and T. J. Silhavy.** 1992. Alpha: the Cinderella subunit of RNA polymerase. J. Biol. Chem. **267:**14515–14518.
31. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
32. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
33. **Sardinia, L. M., J. N. Engel, and D. Ganem.** 1989. Chlamydial gene encoding a 70-kilodalton antigen in *Escherichia coli*: analysis of expression signals and identification of the gene product. J. Bacteriol. **171:**335–341.
34. **Schachter, J.** 1988. The intracellular life of *Chlamydia*, p. 109–139. *In* W. Goebel (ed.), Current topics in microbiology and immunology. Springer-Verlag, New York.
35. **Schachter, J., and H. D. Caldwell.** 1980. Chlamydiae. Annu. Rev. Microbiol. **34:**285–309.
36. **Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B. Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada, and M. Sugiura.** 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. **5:**2043–2049.
37. **Sijben-Muller, G., R. B. Hallick, J. Alt, P. Westhoff, and R. G. Herrmann.** 1986. Spinach plastid genes coding for initiation factor IF-1, ribosomal protein S11 and RNA polymerase α-subunit. Nucleic Acids Res. **14:**1029–1044.
38. **Slauch, J. M., F. D. Russo, and T. J. Silhavy.** 1991. Suppressor mutations in *rpoA* suggest the OmpR controls transcription by direct interaction with the α subunit of RNA polymerase. J. Bacteriol. **173:**7501–7510.
39. **Stephens, R. A., E. A. Wagar, and U. Edman.** 1988. Developmental regulation of tandem promoters for the major outer membrane protein gene of *Chlamydia trachomatis*. J. Bacteriol. **170:**744–750.
40. **Suh, J.-W., S. A. Boylan, and C. W. Price.** 1986. Gene for the alpha subunit of *Bacillus subtilis* RNA polymerase maps in the ribosomal protein gene cluster. J. Bacteriol. **168:**65–71.
41. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82:**1074–1078.
41a. **Tan, M.** Unpublished data.
41b. **Tan, M., and J. Engel.** Unpublished results.
42. **Thomas, M. S., D. M. Bedwell, and M. Nomura.** 1987. Regulation of α operon gene expression in *Escherichia coli*: a novel form of translational coupling. J. Mol. Biol. **196:**333–345.
43. **Thomas, M. S., and R. E. Glass.** 1991. *Escherichia coli rpoA* mutation which impairs transcription of positively regulated systems. Mol. Microbiol. **5:**2719–2725.
44. **Versalovic, J., T. Koeuth, R. Britton, K. Geszvain, and J. R. Lupski.** 1993. Conservation and evolution of the *rpsU-dnaG-rpoD* macromolecular synthesis operon of bacteria. Mol. Microbiol. **8:**343–355.
45. **Yates, J. L., A. E. Arfsten, and M. Nomura.** 1980. *In vitro* expression of *Escherichia coli* ribosomal protein genes: autogenous inhibition of translation. Proc. Natl. Acad. Sci. USA **77:**1837–1841.