Volume 12 Number 14 1984 Nucleic Acids Research

Nucleotide sequence of the *rpoA-rplQ* DNA of *Escherichia coli:* a second regulatory binding site for protein S4?

David W.Meek and Richard S.Hayward

Department of Molecular Biology, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, UK

Received 17 May 1984; Accepted 21 June 1984

ABSTRACT

The "α-operon" of E.coli is a unit of regulation comprising the following known genes, mostly encoding ribosomal proteins (in order of transcription, and with their products named in brackets): rpsM (S13), rpsK (S11), rpsD (S4), rpoA (α-subunit of RNA polymerase), rplQ (L17). There is evidence that S4 tightly regulates all of these genes, except rpoA, by repressing translation of the polycistronic mRNA. Binding of S4 to the S13 start-site is thought to regulate the first three genes. We have extended the 'rpsD-rpoA' sequences previously determined by others, to include all of rpoA and rplQ. The rpoA-rplQ intercistronic region shows strong primary, and potential secondary structural homologies with the S4-binding sites on 16S rRNA and S13 mRNA. We suggest that S4 represses L17 translation directly.

INTRODUCTION

Most of the operons encoding ribosomal proteins in E.coli show an elegant form of feedback regulation at the level of translation, and in one case additionally through attenuation of transcription; see for example the reviews by Nomura and colleagues (1) and Lindahl and Zengel (2), and recent papers from both groups (3, 4). One interesting case, discussed in particular in ref.1, is the "α-operon" (Fig.1). This is located near 72.4 min on the genetic map of E.coli, and comprises at least five genes (mostly encoding ribosomal proteins): in order of transcription they are rpsM (S13), rpsK (S11), rpsD (S4), rpoA (RNApolymerase subunit α), and rplQ (L17), their protein products being named in brackets. Although an inherently strong promoter, P_{α}, is located immediately upstream of rpsM (5) the α-operon is not normally a unit of transcription because it is mainly transcribed from P_{spc} (Fig.2) with strong or complete occlusion of Pα (5). Nevertheless, it is a (complex) unit of regulation, because translation of all four ribosomal genes is repressed by S4 protein if in significant excess over the amount which can actually be assembled into ribosomes (1, 2, 7, 8).

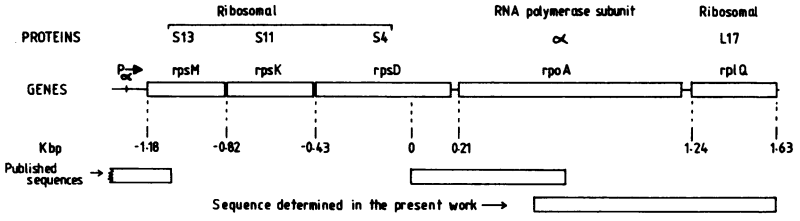

Fig.1. The α-operon of E.coli drawn to scale. The genetic map, and the sequences of the P_α (promoter) and 'rpsD-rpoA' regions are given in refs. 5 and 11. Position 1 of the kilobasepair (Kbp) scale is that of ref. 11.

Nomura and colleagues have presented evidence that translation of the first three genes (Fig.1) is repressed by binding of S4 protein to the mRNA at the S13 translational start-site; and that this mRNA region shows striking homologies of primary and putative secondary structure with the S4-binding site on 16S ribosomal RNA (9). There is, however, evidence that the rpoA gene is less strongly repressed by S4 in vivo (1); and that it can be translationally uncoupled from the neighbouring ribosomal genes, e.g. by rifampicin treatment of E.coli (10). Accordingly, as discussed by Nomura et al. (1), the apparently tighter regulation of the distal rplQ gene by S4 requires explanation.

In the course of our studies on rpoA we have extended the 'rpsD-rpoA' sequence published by Post and Nomura (11) to include the whole of rpoA and rplQ. Our results suggest that S4 probably regulates rplQ directly by binding to the mRNA start-site for L17 translation, as hypothesised by Nomura et al, (1).

MATERIALS AND METHODS

λspc1 (12, 13) was kindly provided (in the lysogen NO 1267) by Prof. M. Nomura; and the plasmid pKO4 by Dr. K. McKenney. pKO4 is a derivative of pKO1 (14), having a BamHI linker inserted into the SmaI target. DNA polymerase I (large fragment) was purchased from Boehringer Corp. Ltd., dideoxy- and deoxyribonucleoside triphosphates from PL, and labelled nucleotides from Amersham International. Other materials, and bacteriophage, plasmid, and DNA manipulations were as reported previously (15).

RESULTS

The transducing bacteriophage λspc1 carries a 20 Kbp E.coli DNA insert, including the α-operon; we have mapped it with several extra

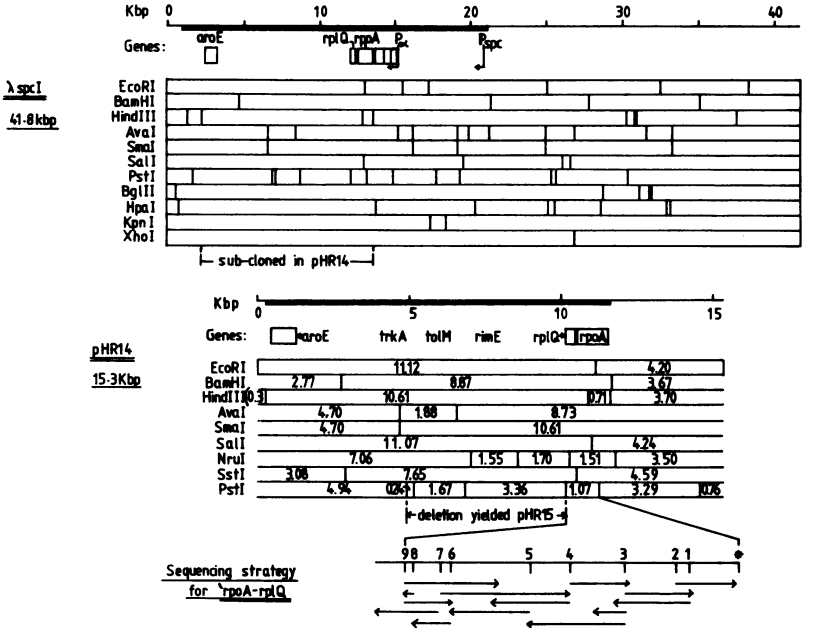

Fig.2. Restriction and partial genetic maps of λspc1 and pHR14, showing the regions of DNA sequenced in this work. The α-operon DNA inserts are indicated by heavy lines. The map of λspc1 is derived from refs. 5, 11, 12, 13 and 18, plus our own restriction data (not shown) and unpublished results of Dr. J.C. Ma (Edinburgh and Guangdong) and, for aroE, Dr. I.A. Anton (Glasgow). The map of pHR14, linearised at an EcoRI site and with segment sizes in Kbp, is from the same sources, together with ref. 14 (and pers. comm. from Dr. K. McKenney, Bethesda) for pKO4, and ref. 19 for the bacterial genes known to be present (the relative position of tolM is uncertain). The bottom lines show the PstI fragments deleted in constructing pHR15; and (in expansions) the origins, extents and directions of the DNA sequencing runs used in the present study. Sites 1 to 9 are origins of sequencing, * and 9 being targets for PstI; 1, EcoRI; 2, SalI; 3, HindIII; 4, SstI; 5,6, and 8, Sau 3A; and 7, NruI. All drawings are to scale.

restriction enzymes (Fig.2). In the course of our studies on the rpoA gene we cloned a HindIII-generated partial-restriction fragment of λspc1 into the plasmid pKO4, to generate pHR14 (Fig.2). We then simplified pHR14 by in vitro deletion, using PstI endonuclease, to generate pHR15 (Fig.2). For DNA sequencing, suitable restriction fragments were sub-cloned from pHR15 (or, in a few cases, pHR14) into M13 mp10 and -mp11 (16), and were sequenced by the latest modification (17) of Sanger's dideoxynucleotide chain termination method. The origins and extents of the sequences determined are shown in Fig.2. The final sequence is

A.

```
 200  gaggacacaa tgcagggttc tgtgacagag tttctaaaac cgcgcctggt tgatatcgag
       START-α Fmetglnglyser valthrglu pheleulyspro argleuval aspileglu

 260  caagtgagtt cgacgcacgc caaggtgacc cttgagcctt tagagcgtgg ctttggccat
      glnvalserser thrhisala lysvalthr leugluproleu gluarggly pheglyhis

 320  actctgggta acgcactgcg ccgtattctg ctctcatcga tgccgggttg cgcggtgacc
      thrleuglyasn alaleuarg argileleu leusersermet proglycys alavalthr

 380  gaggttgaga ttgatggtgt actacatgag tacagcacca aagaaggcgt tcaggaagat
      gluvalgluile aspglyval leuhisglu tyrserthrlys gluglyval glngluasp

 440  atcctggaaa tcctgctcaa cctgaaaggg ctggcggtga gagttcaggg caaagatgaa
      ileleugluile leuleuasn leulysgly leualavalarg valglngly lysaspglu

 500  gttattctta ccttgaataa atctggcatt ggccctgtga ctgcagccga tatcacccac
      valileleuthr leuasnlys serglyile glyprovalthr alaalaasp ilethrhis

 560  gacggtgatg tcgaaatcgt caagccgcag cacgtgatct gccacctgac cgatgagaac
      aspglyaspval gluileval lysprogln hisvalilecys hisleuthr aspgluasn

 620  gcgtctatta gcatgcgtat caaagttcag cgcggtcgtg gttatgtgcc ggcttctacc
      alaserileser metargile lysvalgln argglyarggly tyrvalpro alaserthr

 680  cgaattcatt cggaagaaga tgagcgccca atcggccgtc tgctggtcga cgcatgctac
      argilehisser glugluasp gluargpro ileglyargleu leuvalasp alacystyr

 740  agccctgtgg agcgtattgc ctacaatgtt gaagcagcgc gtgtagaaca gcgtaccgac
      serprovalglu argileala tyrasnval glualaalaarg valglugln argthrasp

 800  ctggacaagc tggtcatcga aatggaaacc aacggcacaa tcgatcctga agaggcgatt
      leuasplysleu valileglu metgluthr asnglythrile aspproglu glualaile

 860  cgtcgtgcgg caaccattct ggctgaacaa ctggaagctt tcgttgactt acgtgatgta
      argargalaala thrileleu alaglugln leuglualaphe valaspleu argaspval

 920  cgtcagcctg aagtgaaaga agagaaacca gagttcgatc cgatcctgct gcgccctgtt
      argglnproglu vallysglu glulyspro glupheasppro ileleuleu argproval

 980  gacgatctgg aattgactgt ccgctctgct aactgcctta aagcagaagc tatccactat
      aspaspleuglu leuthrval argserala asncysleulys alagluala ilehistyr

1040  atcggtgatc tggtacagcg taccgaggtt gagctcctta aaacgcctaa ccttggtaaa
      ileglyaspleu valglnarg thrgluval gluleuleulys thrproasn leuglylys

1100  aaatctctta ctgagattaa agacgtgctg gcttcccgtg gactgtctct gggcatgcgc
      lysserleuthr gluilelys aspvalleu alaserarggly leuserleu glymetarg

1160  ctggaaaact ggccaccggc aagcatcgct gacgagtaac cggatcacag gttaagcgtt
      leugluasntrp proproala serileala aspgluSTOP-α

1220  ttactgagaa ggataaggtc atgcgccatc gtaagagtgg tcgtcaactg aaccgcaaca
              START-L17   Fmetarghisarg lyssergly argglnleu asnargasn
```

```
1280   gcagccatcg ccaggctatg ttccgcaata tggcaggttc actggttcgt catgaaatca
       serserhisarg glnalamet pheargasnmet alaglyser leuvalarg hisgluile

1340   tcaagacgac tctgcctaaa gcgaaagagc tgcgccgcgt agttgagccg ctgattactc
       ilelysthrthr leuprolys alalysgluleu argargval valglupro leuilethr

1400   ttgccaagac tgatagcgtt gctaatcgtc gtctggcatt cgcccgtact cgtgataacg
       leualalysthr aspserval alaasnargarg leualaphe alaargthr argaspasn

1460   agatcgtggc aaaactgttt aacgaactgg gcccgcgttt cgcgagccgt gccggtggtt
       gluilevalala lysleuphe asngluleugly proargphe alaserarg alaglygly

1520   acactcgtat tctgaagtgt ggcttccgtg caggcgacaa cgcgccgatg gcttacatcg
       tyrthrargile leulyscys glypheargala glyaspasn alapromet alatyrile

1580   agctggttga tcgttcagag aaagcagaag ctgctgcaga gtaa
       gluleuvalasp argserglu lysalagluala alaalaglu STOP-L17
```

B.

```
   1   ctgcaggctg atgttatggt cgtcgtcgcc tatggtttaa ttctgccgaa agcagtgctg
  61   gagatgccgc gtcttggctg tatcaacgtt catggttcac tgctgccacg tggcgcggtg
 121   ctgcaccaat ccaacgctca ctatgggcgg gtgatgcaga aactggtgtg accattatgc
 181   aaatggatgt cggtttagac accggtgata gctctataag ctctcctgcc cgattactgc
 241   ag
```

Fig.3. A. Nucleotide sequence of rpoA-rplQ of E.coli K12, map position 72.4 min (19). Nucleotides 200-685 are as reported by Post and Nomura (11) whose scale we have adopted; nucleotides 550-1623 have been determined independently in the present work. The predicted amino acid sequences of the gene products α (RNA polymerase subunit) and L17 (ribosomal component) are shown below the DNA sequence. All but nucleotides 1619-1623 have been determined on both strands.

B. Nucleotide sequence of a small PstI-generated fragment of E.coli K12, mapping between aroE and rplQ at about 72.3 min (19; and see Fig.2). The orientation and genetic function are unknown. Nucleotides 15-204, inclusive, were determined on both strands.

presented (as the DNA strand corresponding to the mRNA) in Fig.3, together with the predicted amino acid sequences of the α and L17 polypeptides. For completeness we have included the upstream rpoA' sequence (200-686) determined by Post and Nomura (11), whose scale we have adopted. We have confirmed independently virtually all of their rpoA sequence which lies between the EcoRI and upstream Pst1 targets (Fig.2). We have determined both DNA strands throughout the new sequence, except for the final five basepairs of rplQ.

A 242 bp Pst1 fragment mapping between rplQ and aroE (Fig.2) was inadvertently sequenced (on both strands between nucleotides 15 and 204, inclusive) during the course of this work. Its sequence is included in Fig.3, in case of possible usefulness to other workers.

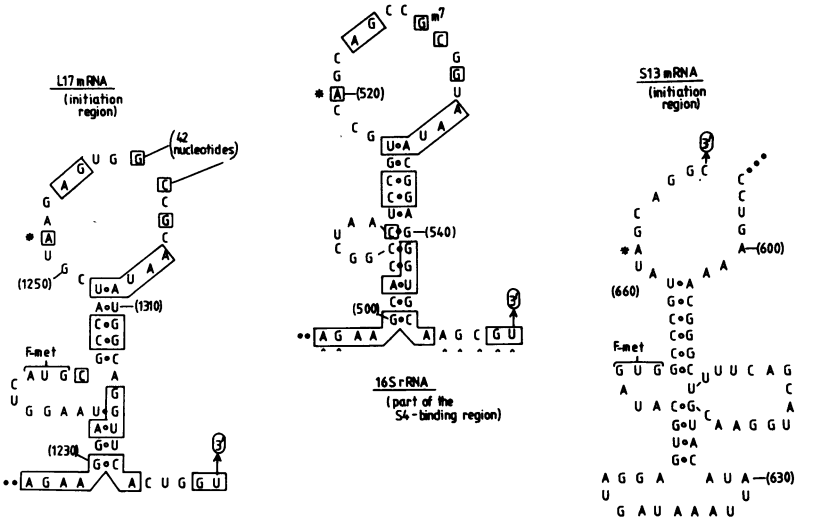

FIG.4. The primary and secondary structures of the proposed binding sites for ribosomal protein S4 on 16S rRNA and S13 mRNA (after Nomura et al. (9); see also Brimacombe et al. (23), and on L17 mRNA (our hypothesis). The F-met start-codons for S13 and L17 translation are indicated. Primary sequence identities between the 16S rRNA and L17 mRNA sites are shown by boxes; for those between 16S rRNA and S13 mRNA, see ref. 9. Proposed base pairs are indicated thus (e.g.) GoC; the nucleotides marked ^ in 16S rRNA are believed to be paired to other regions of the same molecule. The A marked with an asterisk may be essential for S4-binding (9: see text). The nucleotide numbering systems used are as in refs. 9 and 23.

DISCUSSION

Our DNA sequence predicts, for α and L17, precisely the amino acid sequences which have already been published (20,21). Examination of the codon usage in the rpoA and rplQ sequences shows that it is typical of genes coding for ribosomal proteins and other highly expressed polypeptides, with a strong preference for codons recognised by abundant tRNA species (cf. 6,22). Thus the University of Wisconsin Genetics Computer Group's Program "Frame" recognised no rare codons in either gene (and no likely significant translation products in any of the other five possible reading frames).

We have no information on the orientation, in the E.coli map, of the short PstI fragment sequenced (Fig. 3B). The most plausible translational reading frame runs left to right, ending with TGA at 169. A possible rightward promoter sequence lies beyond this (TTTAGA at 194; a 17 bp space; then TAAGCT at 217). The position of this fragment, and the known

genes lying in the space between rplQ and aroE, are shown in Fig. 2.

The most interesting aspect of our results is illustrated in Fig. 4. Nomura and colleagues proposed in 1980 that protein S4 can repress translation of S13 (and through coupling, of S11 and S4) by binding to the S13 mRNA start site, which shows strong primary and secondary structural homologies with part of the known 16S ribosomal RNA binding site for protein S4 (9). The relevant structures are shown in Fig. 4: for the detailed homologies noted by Nomura and colleagues, see ref. 9. Since S4 also regulates the translation of L17, but appears not to affect that of α, Nomura and colleagues (1) recently suggested that there might be a second target site for S4 at the start site for L17 translation. Our results show that the L17 mRNA ribosome-binding site could indeed adopt a configuration (Fig.4) very closely resembling those previously proposed for the S13 mRNA-, and especially for the 16S rRNA-targets for S4 protein-binding (Fig. 4). The most striking similarities are: (i) the basic stem and loop plus side-loop region of the 16S rRNA site, which is one of the most highly conserved features in small-subunit RNA (23).

Although this is interrupted by 42 extra nucleotides (of undefined structure) in the L17 site, the already proposed S13-regulation site is "interrupted" at the same point; (ii) 29 nucleotides in the L17 site (boxed in Fig. 4) are identical to 29 at the corresponding positions in a 56-nucleotide sequence of the 16S rRNA site; (a partially overlapping set of 28 nucleotides in the S13 site has the same property (9); (iii) the L17 start codon and putative Shine-Dalgarno sequence are located in almost identical positions, with respect to the proposed secondary structure, to those of the corresponding S13 initiation signals; and (iv) the A marked with an asterisk, whose presence is considered by Nomura et al (9) to be critical (on the grounds of an intriguingly unusual codon choice in the S13 gene) is also found in the L17 site.

We therefore propose that S4 regulates L17 translation directly by binding to the L17 mRNA initiation region, and sterically excluding ribosomes, or fostering a secondary structure which is unfavourable for initiation. It should be noted that the α-mRNA terminates well upstream of the proposed S4 binding site, and is therefore unlikely to be affected. In their very interesting study of an S4 mutant which is defective in translational repression, Jinks-Robertson and Nomura (7) noted that L17 is less strongly derepressed than S13, S11, and S4. They suggested that this could be an artefact arising from polarity of the S4 mutation. Our results

suggest, as an alternative, that the S4 mutation may interfere to a greater extent with binding to the S13 mRNA than to the L17 mRNA site (or, presumably, the 16SrRNA site). Direct studies of S4/mRNA binding are clearly called for.

It remains to be seen whether or not expression of the rpoA gene is wholly independent of that of the upstream genes. The distance between the S4 stop codon and the α-initiator (25 nucleotides) is probably too great to allow sequential translation by a single ribosome. However, this would not exclude possible sequestration of the α-initiator in a structure formed by pairing with upstream mRNA, which would have to be disrupted (to activate the α-start) by upstream translation, as proposed for other systems). Similarly, the inhibition of upstream translation might reduce transcriptional readthrough into rpoA, and/or the stability of α-mRNA, unless the system has evolved in such a way as to shield the α-sequences from such effects.

Our preliminary data for the DNA sequence downstream of rplQ suggest that a strong transcriptional terminator may be present shortly beyond this gene. We are pursuing this question by more direct analyses.

ACKNOWLEDGEMENTS

We thank Drs. I.A. Anton, J.-C. Ma, and K. McKenney for unpublished information, Prof. M. Nomura for λspc1, G. Brown, E. McCready, and B.Morgan for other help, and the Medical Research Council for personal (DWM) and Project Grant (RSH) support.

REFERENCES

1. Nomura,M., Jinks-Robertson,S. and Miura,A. (1982) in Interaction of Translational and Transcriptional Controls in the Regulation of Gene Expression, Grunberg-Manago,M. and Safer,B. Eds. pp. 91-104, Elsevier Science Publg. Co. Inc., Amsterdam.
2. Lindahl,L. and Zengel,J.M. (1982) Adv. in Genet. 21, 53-121.
3. Baughman,G. and Nomura,M. (1983) Cell 34, 979-988.
4. Lindahl,L., Archer,R. and Zengel,J.M. (1983) Cell, 33, 241-248.
5. Post,L.E., Arfsten,A.E., Davis,G.E. and Nomura,M. (1980) J. Biol. Chem. 255, 4653-4659.
6. Ceretti,D.P., Dean,D., Davis,G.R., Bedwell,D. and Nomura,M. (1983) Nucleic Acids Res. 11, 2599-2616.
7. Jinks-Robertson,S. and Nomura,M. (1982) J. Bacteriol., 151, 193-202.
8. Yates,J.L., Arfsten,A.E. and Nomura,M. (1980) Proc. Natl. Acad. Sci. USA 77, 1837-1841.
9. Nomura,M., Yates,J.L., Dean,D. and Post,L.E. (1980) Proc. Natl. Acad. Sci. USA 77, 7084-7088.
10. Pedersen,S., Reeh,S. and Friesen,J.D. (1978) Molec. Gen. Genet. 166, 329-336.
11. Post,L.E. and Nomura,M. (1979) J. Biol. Chem. 254, 10604-10606.

12. Fiandt,M., Szybalski,W., Blattner,F.R., Jaskunas,S.R., Lindahl,L. and Nomura,M. (1976) J. Mol. Biol. 106, 817-835.
13. Lindahl,L., Zengel,J. and Nomura,M. (1976) J. Mol. Biol. 106, 837-855.
14. McKenney,K., Shimatake,H., Court,D., Schmeissner,U., Brady,C. and Rosenberg,M. (1981) in Gene Amplification and Analysis, Chirikjian, J.G. and Papas,T.S. Eds., Vol II, Elsevier, North-Holland.
15. Newman,A.J., Ma,J.-C., Howe,K.M., Garner,I. and Hayward,R.S. (1982) Nucleic Acids Res. 10, 7409-7424.
16. Messing,J. and Vieira,J. (1982) Gene 19, 269-276.
17. Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
18. Sanger,F., Coulson,A.R., Hong,G.F., Hill,D.F. and Petersen,G.B. (1982) J. Mol. Biol. 162, 729-773.
19. Bachmann,B.J. (1983) Microb. Rev. 47, 180-230.
20. Ovchinnikov,Yu.A., Lipkin,V.M., Modyanov,N.N., Chertov,O.Yu. and Smirnov,Yu.V. (1977) FEBS Lett. 76, 108-111.
21. Rombauts,W., Feytons,V. and Wittmann-Liebold,B. (1982) FEBS Lett., 149, 320-327.
22. Gouy,M. and Gautier,C. (1982) Nucleic Acids Res. 10, 7055-7074.
23. Brimacombe,R., Maly,P. and Zwieb,C. (1983) Prog. in Nucleic Acid Res. and Molec. Biol. 28, 1-48.
24. Yates,J.L., Dean,D., Strycharz,W.A. and Nomura,M. (1981) Nature 294, 190-192.
25. Weissman,C. (1974) FEBS Lett. 40, S10-S18.