Volume 13 Number 12 1985 Nucleic Acids Research

Structure of the *Escherichia coli* S10 ribosomal protein operon

Gerard Zurawski and Sandra Marvo Zurawski

DNAX Research Institute of Molecular and Cellular Biology, 901 California Avenue, Palo Alto, CA 94304, USA

Received 5 April 1985; Revised 20 May 1985; Accepted 24 May 1985

ABSTRACT

The complete structure of the *Escherichia coli* S10 ribosomal protein operon is presented. Based on the DNA sequence, the deduced order of the 11 genes in the operon is *rpsJ, rplC, rplD, rplW, rplB, rpsS, rplV, rpsC, rplP, rpmC, rpsQ*. The estimated transcribed length of the operon is 5181 base pairs. Putative sequences involved in ribosome binding are discussed. The DNA sequence data corrects several errors in previously determined protein sequence data.

INTRODUCTION

The *Escherichia coli* S10 operon encodes eleven ribosomal protein genes and maps at 73 minutes on the chromosome (1). Previous cloning, restriction mapping, and *in vitro* transcription and translation studies (2,3) have established a physical and genetic map of the operon. The deduced order of genes within the S10 operon was *rpsJ, rplC, rplD, rplW, rplB, (rpsS, rplV), rpsC, rplP, rpmC, rpsQ*.

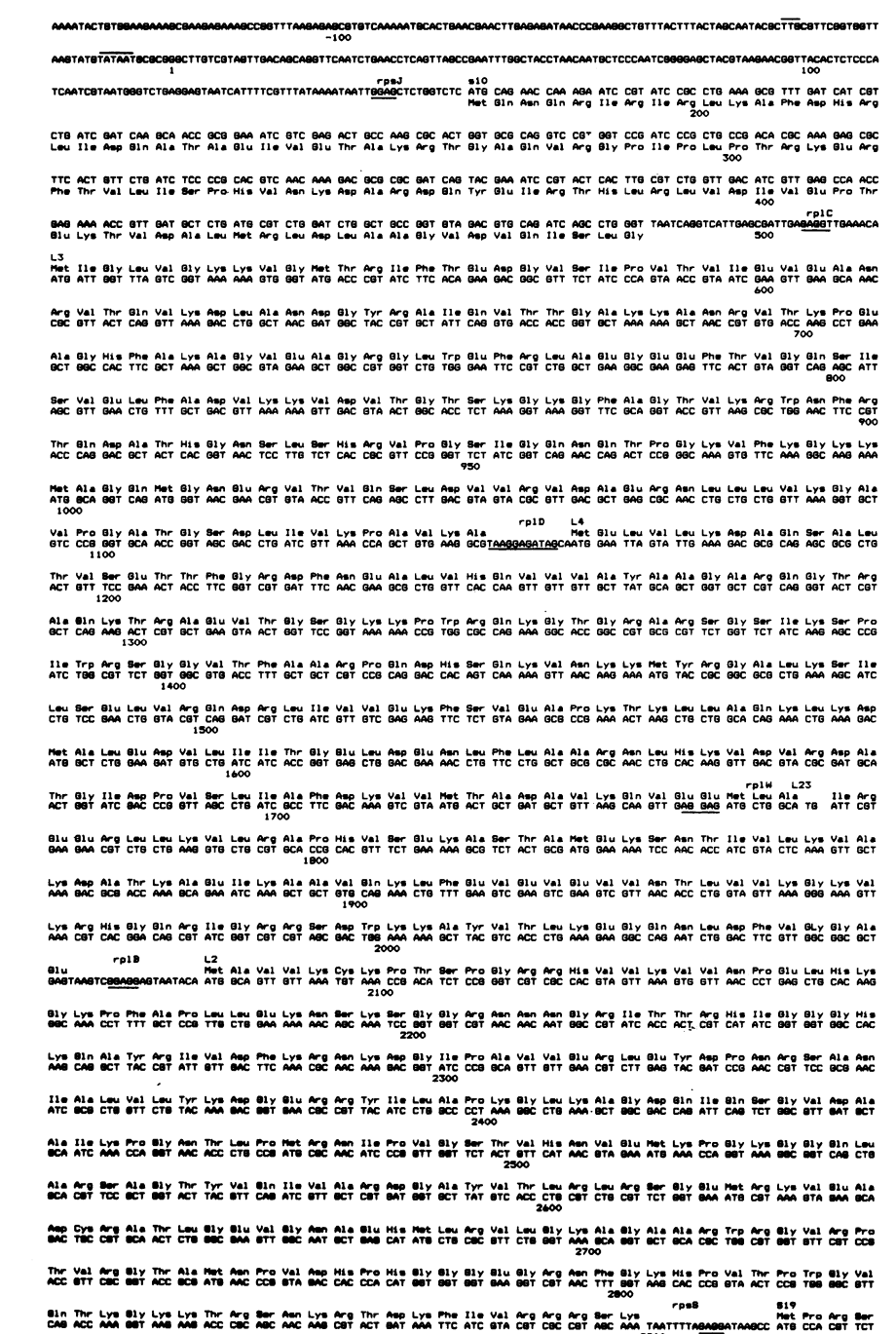
Sequencing and *in vitro* transcription studies have defined the structure of the promoter for the S10 operon and the structure of the first (*rpsJ*), and part of the second (*rplC*), structural genes of the operon (4). Similar studies have defined the sequence of part of the last gene of the operon (*rpsQ*) and the region between the S10 operon and the adjacent *spc* operon (5).

This work presents the structure of the internal part of the S10 operon including the rest of *rplC* and *rpsQ* and the intervening genes *rplD, rplW, rplB, rpsS, rplV, rpsC, rplP*, and *rpmC*.

MATERIALS AND METHODS

Plasmids. Plasmids pLF4.6 and pLF1.0, which bear the 4.6% and 1.0% *Eco*R1 fragments of lambda *fus3* (6), were obtained from J. Watson (Carnegie Institute of Washington, Stanford). Plasmid pNO2003, which bears the 1620bp *Eco*R1-*Hinc*II fragment spaning the end of the S10 operon and the begining of the *spc* operon (5), was provided by M. Nomura (University of California, Irvine).

Sequencing. Insert DNA's from pLF4.6 and pLF1.0 were sequenced by a combination of the dideoxy



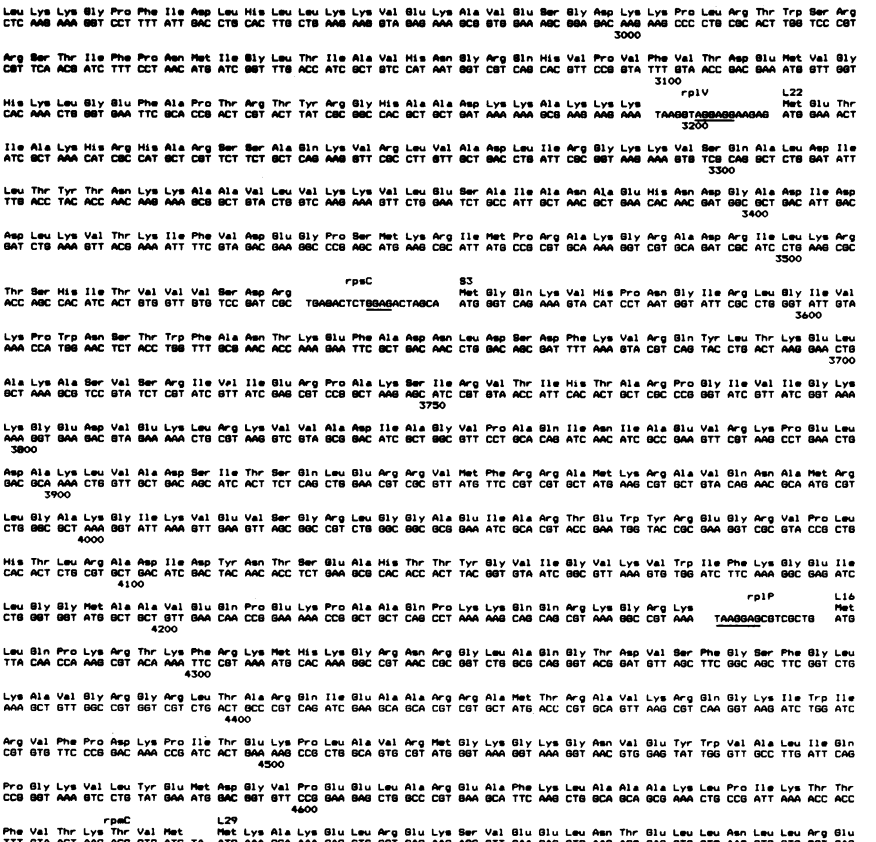


Fig. 1. The nucleotide sequence of the *E. coli* S10 operon commencing proximal to it's promoter region and ending at the *Hinc*II site in the promoter of the adjacent *spc* operon. The sequence of the promoter region, the first structural gene (*rpsJ*), and the initial 240 residues of *rplC*, have been determined previously (4) and are presented here for completeness. The distal 224 residues have been determined previously (5) and were resequenced in this work. The sequence is displayed as coding regions which are identified at the translation initiation codon of each gene. Note that for *rplW* the initiation codon overlaps with the last codon of *rplD*, thus the N terminal MET for L23 is not indicated. Regions proximal to translational initiation sites that are complementary to the 3' end of 16S ribosomal RNA are underlined. The overlined regions are, respectively, the '-35' and '-10' promoter sequences and the putative transcription termination site.

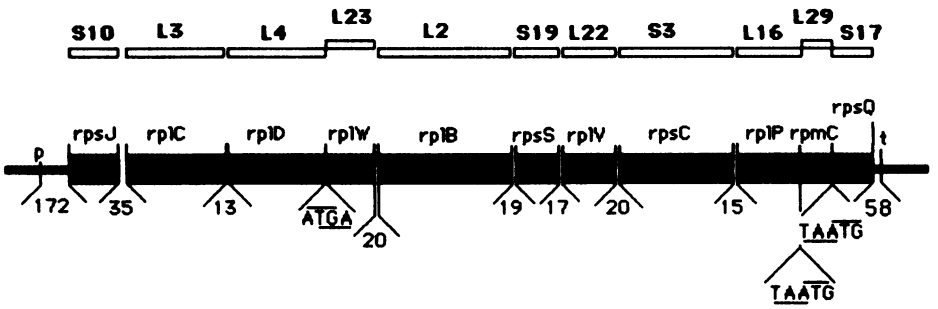


Fig. 2. A map of the S10 operon indicating the order of the genes. Under the map are shown intergenic distances in base pairs, or as a sequence in the three cases where the translation stop codon (underlined) and the translation start codon (overlined) overlap. The 5' leader region and the presumed 3' untranslated region are also indicated. Above the genetic map are represented the 11 ribosomal proteins which this operon encodes.

TABLE 1
S10 Operon Ribosome Binding Sites

Gene	S.D.*	Space δ	Coordinates†
rpsJ	GGAG	10	59-62
rplC	GAGGT	7	505-509
rplD	TAAGGAG	6	1144-1150
rplW	AGGAG	8	1746-1750
rplB	GGAGG	9	2065-2069
rpsS	GAGG	7	2906-2909
rplV	AGGAGG	5	3199-3204
rpsC	GGAG	7	3549-3552
rplP	TAAGGAG	8	4259-4265
rpmC	TAAG--GGTGA	4	4669-4679#
rpsQ	AAGG	10	4861-4864

* Nucleotides complementary to the 3' end of 16S RNA

δ The distance between the last nucleotide shown for the S.D. and the beginning of the initiator ATG

† Refers to nucleotide numbers in Fig. 1

rpmC has two S.D. sequences i.e. TAAG with a space of 11 and GGTGA with a space of 4. Either, or both, may have a potential role in ribosome binding.

TABLE 2
Protein Sequence Discrepencies

Protein	Change	Residue †	Location ¶
L23	TRP inserted	80	1996-1998
L2	HIS -> GLY	230	2766-2768
	GLY -> HIS	233	2775-2777
S19	ARG inserted	36	3022-3024
	ARG deleted	43*	
	ASN -> ASP	43	3043-3045
	ASP -> ASN	86	3172-3174
S17	CYS inserted	53	5031-5033
	CYS deleted	59*	

† Refers to residues starting from the initiator MET, note that the N terminal MET is absent in mature L2, S3, S17 and S19. Changes are relative to the DNA-derived sequence.
¶ Refers to nucleotide numbers given in Fig. 1
* Refers to residue numbers in published protein sequences

(7) and chemical degradation (8) methods. Insert DNA from pNO2003 was sequenced entirely by the dideoxy method. All regions were sequenced by two methods, or on two strands. All restriction sites were overlapped except the *Eco*R1 sites in *rplC*, *rpsS*, *rpsC* and the *Bam*H1 site in *rplD*. The sequences at these sites were, however, consistent with the relevant protein sequences.

RESULTS AND DISCUSSION

The Structure and Organisation of the Operon. Figure 1 shows the nucleotide sequence of the S10 operon commencing proximal to the operon's promoter (4) and ending at the *Hinc* II site in the vicinity of the promoter for the adjacent *spc* operon (5). The figure shows the sequence translated into the open reading frames which encode the 11 ribosomal proteins specified by this operon. The total transcribed length of the operon (assuming the transcription termination point suggested by Post *et. al.*, 5) is 5181bp.

The Ribosome Binding Sites of the S10 Operon. Figure 2 is a map of the operon which details the sizes of the intergenic regions. Table 1 shows the extent and location of regions proximal to translation start sites having complementarity to the 3' end of *E. coli* 16S ribosomal RNA. Such sequences are known to have an important function in ribosome binding (9). The extent, location, and nature of all the regions of complementarity shown fall within the ranges encountered in other *E. coli* ribosome binding sequences (9).

Three of the intergenic regions have overlapping translation stop / translation start codons (Fig.

2). In at least one operon, the *trp* operon (10), this arrangement is typical for genes encoding subunits of protein complexes whose components are required in equimolar amounts.

Protein Sequence Corrections Based on the DNA Sequence. All the proteins encoded by the S10 operon have been sequenced (11-20). Differences between these protein sequences and protein sequences derived from the DNA sequence are shown in Table 2. The DNA sequence data at the points of difference was exhaustively rechecked and all the discrepencies (except an undetected TRP residue and a GLN -> GLU change) can be accounted for by a misassignment of position in the protein sequence.

CONCLUSIONS

The data presented above complete the full physical description of the structure of the *E. coli* S10 operon. This information should serve as a firm basis for the further study of the expression and regulation of the operon. Furthermore, the data will add to compilations of sequence banks for the study of the nature of codon usage and ribosome binding sites and correct a number of errors in previously published protein sequences.

ACKNOWLEDGEMENTS

We would like to thank B. Wittmann-Liebold for helping us reconcile DNA and protein sequence data, and L. Schauer and M.Nomura for encouraging us to complete this work.

REFERENCES

1. Bachmann, B.J. (1983) Microbiol. Rev., 47, 180-230.
2. Lindahl, L., Post, L., Zengel, J., Gilbert, S.F., Strycharz, and Nomura, M. (1977) J. Biol. Chem. 252, 7365-7383.
3. Nomura, M. and Post, L. (1980) Ribosomes: Structure, Function and Genetics. Chambliss, G., Craven, G.R., Davis, J., Davis, K., Kahan, L. and Nomura, M. Eds., pp. 671-691, Baltimore: University Park Press.
4. Olins, P.O. and Nomura, M. (1978) Cell, 26, 205-211.
5. Post, L.E., Arfsten, A.E., Reusser, F. and Nomura, M. (1978) Cell 15, 215-229.
6. Watson, J. and Surzycki, S.J. (1983) Curr. Gen. 7, 201-210.
7. Senger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
8. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
9. Stormo, G.D., Schneider, T.D. and Gold, L.D. (1982) Nucl. Acids Res. 10, 2971-2996.
10. Oppenheim, D.S. and Yanofsky, C. (1980) Genetics 95, 785-795.
11. Maranova, T.A., Muranov, A.V., Markova, L.F. and Ovchinnikov, Y.A. (1978) FEBS Lett. 96, 301-305.
12. Kimura, M. and Wittmann-Liebold, B. (1980) FEBS Lett. 121, 317-322.
13. Wittmann-Liebold, B. and Greuer, B. (1979) FEBS Lett. 108, 69-74.
14. Kimura, M., Minde, L. and Wittmann-Liebold, B. (1982) FEBS Lett. 149, 304-312.
15. Yaguchi, M.and Wittmann, H.G. (1978) FEBS Lett. 88, 227-230.
16. Wittmann-Liebold, B. and Greuer, B. (1980) FEBS Lett. 121, 105-112.
17. Bauer, D. and Roming, R. (1979) FEBS Lett. (1979) 106, 352-357.
18. Brosius, J. and Chen, R. (1976) FEBS Lett. 68, 105-109.
19. Bitar, K.G. (1975) Biochim. Biophys. Acta 386, 99-106.
20. Yaguchi, M. and Wittman, H.G. (1978) FEBS Lett. 87, 37-40.