

---

**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*, *rpIC*, *rpID*, *rpIW*, *rpIB*, *rpsS*, *rpIV*, *rpsC*, *rpIP*, *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*, *rpIC*, *rpID*, *rpIW*, *rpIB*, (*rpsS*, *rpIV*), *rpsC*, *rpIP*, *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 (*rpIC*), 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 *rpIC* and *rpsQ* and the intervening genes *rpID*, *rpIW*, *rpIB*, *rpsS*, *rpIV*, *rpsC*, *rpIP*, and *rpmC*.

**MATERIALS AND METHODS**

**Plasmids.** Plasmids pLF4.6 and pLF1.0, which bear the 4.6% and 1.0% *EcoRI* fragments of lambda *fusJ* (6), were obtained from J. Watson (Carnegie Institute of Washington, Stanford). Plasmid pN02003, which bears the 1620bp *EcoRI*-*HincII* fragment spanning the end of the S10 operon and the beginning 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

# Nucleic Acids Research

AAATACTBTBAABAABCAABAAACCCBBT1TAAGABACBCTGTCAAAAATBCACTBAACBAACCTTBABABATAACCCBAGBGGT1TTACTT1TACTAACAATACBCTTBCSTTCSBT88TT  
 -100

AAATGTATATAATBCBCC988TCTGTCTAGTACACAGBCTTCAATCTBAACCTCACTAGCCBAAT1TB9CTACCTAACAAATCTCCCAACTCB998ACBACTA8A9AC98TACACTCTCCDA  
 1

TCATCTGTAT889TCT88898TAATCATTTTC8TTTATAAAATTTG888CTCT88TCTC rps2 s10  
 ATG CAG AAC CAA AGA ATC CBT ACT CBT CBT AAA BCS TTT GAT CAT CBT  
 Met Glu Asn Glu Arg Ile Arg Ile Arg Leu Lys Ala Phe Asp His Arg  
 200

CTB ATC BAT CAA BCA ACC GCB BAA ATC BTC BAC ACT GCC AAG CBC ACT 88T GCB CAG GTC C8T 88T CCB ACT CCB CBT CCB ACA CBC AAA BAC CBC  
 Leu Ile Asp Glu Ala Thr Ala Glu Ile Val Glu Thr Ala Lys Arg Thr Gly Ala Glu Val Arg Gly Pro Ile Pro Leu Pro Thr Arg Lys Glu Arg  
 300

TTC ACT GTT BTT ATC TCC CCB CAC BTC AAA AAC BAC GCB CBC BAT CAG TAC BAA ATC CBT ACT CAC TTB CBT CBT BAC ATC GTT BAC CCA ACC  
 Phe Thr Val Leu Ile Ser Pro His Val Asn Lys Asp Ala Arg Asp Glu Tyr Glu Ile Arg Thr His Leu Arg Leu Val Asp Ile Val Glu Pro Thr  
 400

BAG AAA ACC BTT BAT GCT CTS ATB CBT CBT BAY CTS GCT SCC 88T 8TA BAC BGS CAB ATC AGC CBT 88T TAATCAB8TTCATBAC8BATTBAG88TBA8AAAC  
 rplC  
 Met Glu Thr Val Asp Ala Thr Met Arg Leu Asp Leu Ala Ala Gly Val Asp Val Glu Ile Ser Leu Gly  
 500

L3  
 Met Ile Gly Leu Val Gly Lys Lys Val Gly Met Thr Arg Ile Phe Thr Glu Asp Gly Val Ser Ile Pro Val Thr Val Ile Glu Val Glu Ala Asn  
 ATG ATT GGT TTA GTC GGT AAA AAA GTB 88T ATG ACC CBT ATC TTC ACA BAA GAC G8C BTT TCT ATC CCA BTA ACC GTA ATC BAA GTT BAA GCA AAC  
 600

Arg Val Thr Glu Val Lys Asp Leu Ala Asn Asp Gly Tyr Arg Ala Ile Glu Val Thr Gly Ala Lys Ala Asn Arg Val Thr Lys Pro Glu  
 CBT GTT ACT CAG GTT AAA BAC CTS GCT ACC BBT GCT TAC CBT GCT ACT ATT CAG BGS ACC ACC BBT GCT AAA AAA BCT AAC CBT BGS ACC BCT BAA  
 700

Ala Gly His Phe Ala Lys Ala Gly Val Glu Ala Gly Arg Gly Leu Trp Glu Phe Arg Leu Glu Ala Gly Glu Glu Phe Thr Val Gly Glu Ser Ile  
 GCT 88C BAC TTC BCT BAA GCT 88C GTT 88T GCT 88C CBT 88T CBT 88C BAA TTC CBT CBT G8A TCC CBT CBT G8A G8C BAA G8B TTC ACT BTA ACC AAC ATT  
 800

Ser Val Glu Leu Phe Ala Asp Val Lys Lys Val Asp Val Thr Gly Thr Ser Lys Gly Lys Gly Phe Ala Gly Thr Val Lys Arg Trp Asn Phe Arg  
 ABC GTT GAA C8B TTT GCT BAC GTT AAA AAA BTT BAC BTA ACT 88C ACC TCT AAA G8T AAA B8T TTC GCA B8T ACC GTT AA8 BCB TB8 AAC TTC CBT  
 900

Thr Glu Asp Ala Thr His Gly Asn Ser Leu Ser His Arg Val Pro Gly Ser Ile Gly Glu Asn Glu CBT ACT CCB 88T TCT ATC 88T CAG AAC C8B CBT Pro Gly Lys Val Phe Lys Gly Lys Lys  
 ACC CAG BAC CBT ACT CAC 88T AAC TCC TTB TCT CAC CCB GTT CCB 88T TCT ATC 88T CAG AAC C8B CBT CBT BTA CCA CCA CBT BTA TTB AAA G8C BCB C8B AAC AAA  
 950

Met Ala Gly Glu Met Gly Asn Glu Arg Val Thr Val Glu Ser Leu Asp Val Val Arg Val Asp Ala Glu Arg Asn Leu Leu Leu Val Lys Gly Ala  
 ATB BCA 88T CAG B8B ATB 88T AAC BAA CBT BTA ACC GTT CAG AGC CTT CAG BTA BTA C8B BTT CAG BCT GAC B8C C8C AAC CBT CBT CBT GTT AAA B8T GCT  
 1000

Val Pro Gly Ala Thr Gly Ser Asp Leu Ile Val Lys Pro Ala Val Lys Ala Met Glu Leu Val Lys Asp Ala Glu Ser Ala Leu  
 GTC CCB 88T BCA ACC 88T BAC CTS ATC BTT AAA CCA BCT B8B A8B GCBTAA8888T88AATB GAA TTA BTA TTB AAA G8C BCB C8B AAC AAA  
 1100

Thr Val Ser Glu Thr Thr Phe Gly Arg Asp Phe Asn Glu Ala Leu Val His Glu Val Val Val Ala Tyr Ala Ala Gly Ala Arg Glu Gly Thr Arg  
 ACT GTT TCC BAA ACT ACC TTC BBT CBT BAT TTC ACC GAA GCB C8B GTT CAC CAA GTT BTT GTT BTT TCC BTA ACT G8T CBT CBT C8B 88T ACT CBT  
 1200

Ala Glu Lys Thr Arg Ala Glu Val Thr Gly Ser Gly Lys Pro Trp Arg Glu Lys Gly Thr Gly Arg Ala Arg Ser Gly Ser Ile Lys Ser Pro  
 BCT CAG A8B ACT CBT GCT BAA BTA ACT 88T TCC 88T AAA AAA CCB TB8 G8C CAG AAA G8C ACC G8C CBT C8B CBT TCT 88T TCT ATC A8B ASC CCB  
 1300

Ile Trp Arg Ser Gly Gly Val Thr Phe Ala Ala Arg Pro Glu Asn His Ser Glu Lys Val Asn Lys Lys Met Tyr Arg Gly Ala Leu Lys Ser Ile  
 ATC TB8 CBT TCT 88T 88C 88B ATC TTT GCT 88T CBT C8B CAG CAC AGT CAA AAA GTT AAC A8B AAA GTT TAC C8B 88C BCB 88B AAA ACC BAC  
 1400

Leu Ser Glu Leu Val Arg Glu Asp Arg Leu Ile Val Val Glu Lys Phe Ser Val Glu Ala Pro Lys Thr Lys Leu Leu Ala Glu Lys Leu Lys Asp  
 C8B TCC BAA C8B BTA CBT CAG B8T CBT C8B ATC GTT BTC B8B A8B TTC TCT BTA BAA CCB CCB AAA ACT A8B C8B C8B CCA CAA C8B AAA G8C  
 1500

Met Ala Glu Leu Asp Val Leu Ile Thr Gly Glu Leu Asp Glu Asn Leu Phe Leu Ala Ala Arg Asn Leu His Lys Val Asp Val Arg Asp Ala  
 ATB BCT C8B BAA GTB 88T C8B ATC ACC B8T 88B C8B BAA AAC C8B TTC C8B CBT CCB C8C AAC C8B C8B C8B C8B AAC BTT BAC BTA C8B BAC GCA  
 1600

Thr Gly Ile Asp Pro Val Ser Leu Ile Ala Phe Asp Lys Val Val Met Thr Ala Asp Ala Val Lys Glu Val Glu Glu Met Leu Ile Arg  
 ACT 88T BCT BAC CCB GTT AAC C8B C8B ATC TCC G8C AAA B8C BTA ATB ACT CBT B8T 88T AAC C8B BTT G8B B8B ATB C8B G8B T8B ATT CBT  
 1700

Glu Glu Arg Leu Leu Lys Val Leu Arg Pro His Val Ser Glu Lys Ala Ser Thr Ala Met Glu Lys Val Asn Thr Leu Val Val Asn Pro Glu Lys Val Ala  
 BAA BAA CBT C8B C8B A8B GTB C8B CBT BCA CCB C8B BTT TCT BAA AAA C8B TCT ACT C8B ATB BAA AAA TCC AAC ACC ATC BTA C8B Lys Ala BTT GCT  
 1800

Lys Asp Ala Thr Lys Ala Glu Ile Lys Ala Ala Val Glu Lys Leu Phe Glu Val Glu Val Glu Val Asn Thr Leu Val Val Lys Gly Lys Val  
 AAA BAC C8B ACC AAA BCA BAA ATC AAA BCT 88T GTB C8B AAA C8B TTT BAA B8C BAA B8C BAA GTC BAA B8C BAA GTT AAC ACC C8B BTA BTT AAA 888 AAA GTT  
 1900

Lys Arg His Gly Glu Arg Ile Gly Arg Arg Ser Asp Trp Lys Lys Ala Tyr Val Thr Leu Lys Glu Gly Glu Asn Leu Asp Phe Val Gly Gly Ala  
 AAA CBT CAG B8A C8B CBT ATC 88T CBT CBT ABC BAC TB8 AAA AAA CBT TAC B8C ACC C8B AAA G8C B8C AAT C8B BAC TCC GTT 88C 88C GCT  
 2000

rplB L2  
 Glu Met Ala Val Val Lys Cys Lys Pro Thr Ser Pro Gly Arg His Val Lys Val Asn Pro C8B Glu Leu His Lys  
 B8BTA8BCT888888TAATACA ATB BCA BTT BTT AAA T8T AAA CCB ACA TCT CCB 88T CBT C8C C8B BTA GTT AAA B8B BTT AAC C8B C8B C8B CAC A8B  
 2100

Gly Lys Pro Phe Ala Pro Leu Leu Glu Lys Asn Ser Lys Ser Gly Gly Arg Asn Asn Asn Gly Arg Ile Thr Thr Arg His Ile Gly Gly Gly His  
 88C AAA CCT TTT BCT CCB TTB C8B BAA AAA AAC ABC AAA TCC 88T 88T CBT AAC AAC AAT B8C CBT ACT ACC ACT CBT CAT ACT 88T 88T 88C CAC  
 2200

Lys Glu Ala Tyr Arg Ile Val Asp Phe Asn Lys Asp Ile Pro Ala Val Val Glu Arg Leu Glu Tyr Asp Pro Asn Arg Ser Ala Asn  
 A8B C8B BCT TAC CBT ATT BTT BAC TTC AAA C8C AAC AAA B8C 88T CBT CCB B8A BTT BTT BAA CBT CBT B8C AAC C8B C8B C8B C8B C8B  
 2300

Ile Ala Leu Val Leu Tyr Lys Asp Gly Glu Arg Arg Tyr Ile Leu Ala Pro Lys Gly Leu Ala Gly Ala Ser Ile Glu Ser Gly Val Asp Ala  
 ATC BCB C8B GTT C8B TAC AAA 88T BAA C8C CBT TAC ATC C8B C8C CCT AAA 88C C8B Lys AAA B8C BAC C8B ATT C8B BCT 88C BTT BCT  
 2400

Ala Ile Lys Pro Gly Asn Thr Leu Pro Met Arg Asn Ile Pro Val Gly Ser Thr Val His Asn Val Glu Met Lys Pro Gly Lys Gly Glu Glu  
 B8A ATC AAA CCA 88T AAC ACC C8B C8B ATB C8C AAC ATC CCB BTT 88T TCT ACT BTT CAT AAC BTA BAA ATB AAA CCA 88T AAA 88C 88T C8B C8B  
 2500

Ala Arg Ser Ala Gly Thr Tyr Val Glu Ile Val Ala Arg Asp Gly Ala Tyr Val Thr Leu Arg Leu Arg Ser Gly Glu Met Arg Lys Val Glu Ala  
 C8B CBT TCC 88T 88T ACT TAC BTT BAC ATT 88T BCT CBT 88T CBT 88T CBT TAT B8C ACC C8B C8B C8B C8B C8B C8B C8B C8B C8B C8B C8B  
 2600

Asp Cys Arg Ala Thr Leu Gly Glu Val Gly Asn Ala Glu His Met Leu Arg Val Leu Gly Lys Ala Gly Ala Ala Arg Trp Arg Gly Val Arg Pro  
 B8C TB8 CBT BCA ACT C8B 88C BAA GTT 88C AAT BCT B8B CAT ATB C8B C8B BTT C8B 88T AAA B8A 88T BCT B8A C8C TB8 CBT 88T BTT CBT C8B  
 2700

Thr Val Arg Thr Ala Met Ser Val Asp His Pro His Gly Gly Glu Gly Arg Asn Phe Gly Lys His Pro Val Thr Pro Trp Gly Val  
 ACC BTT CBT 88T ACC BCB ATB AAC C8B BTA BAC CAC CCA CAT 88T 88T BAA B8T CBT 88T BAA B8T CBT AAC TTT 88T A8B C8C C8B BTA ACT C8B TB8 88C BTT  
 2800

Glu Thr Lys Ile Lys Lys Thr Arg Ser Asn Lys Arg Thr Asp Lys Phe Ile Val Arg Arg Arg Ser Lys  
 CAB ACC AAA 88T A8B A8B ACC C8C ABC AAC A8B CBT ACT AAA TTC ATC BTA CBT C8C CBT AAC AAA TAATTTTAA88BATA88C ATB CCA CBT TCT  
 2900

rps8 s19  
 Met Pro Arg Ser

```

Leu Lys Lys Gly Pro Phe Ile Asp Leu His Leu Leu Lys Lys Val Glu Lys Ala Val Glu Ser Gly Asp Lys Lys Pro Leu Arg Thr Trp Ser Arg
CTC AAG AAA GGT CCT TTT ATT BAC CTS CAC TTS CTS AAG AAG GTA BAA AAA SCB STS BAA ABC BBA BAC AAG AAG CCC CTS CSC ACT TGG TCC CBT
3000

Arg Ser Thr Ile Phe Pro An Met Ile Gly Leu Thr Ile Ala Val His An Gly Arg Gln His Val Pro Val Phe Val Thr Asp Ala Met Val Gly
CST TCA ACG ATC TTT CCT AAG ATC BCT AAT GGT BTT TTS ACC ATC GCT BTC CAT AAT GGT CBT CAG CAC GTT CDS STA TTT GTA ACC BAA ATG GTT GGT
3100

His Lys Leu Gly Glu Phe Ala Pro Thr Arg Thr Tyr Arg Gly His Ala Ala Asp Lys Lys Ala Lys Lys Lys rplV L22 Met Glu Thr
CAC AAA CTS BGT BAA TTC GCA CCB ACT CBT ACT TAT CAC BGC CAC BCT GCT BAT AAA AAA GCB AAG AAG AAA TAAGTABBAAAGAG ATB BAA ACT
3200

Ile Ala Lys His Arg His Ala Arg Ser Ser Ala Gln Lys Val Arg Leu Val Ala Asp Leu Ile Arg Gly Lys Lys Val Ser Gln Ala Leu Asp Ile
ATC CTA AAA CAT CBC CAT BCT TCT TCT ACC ACT CAG AAG BTT CBC CTT GTT BCT BAC CTS BGT AAG AAA GGT CBT GCA BCT AAG AAA BAA ATG BAA ATT
3300

Leu Thr Tyr Thr An Lys Lys Ala Ala Val Cys Leu Val Lys Lys Val Leu Glu Ser Ala Ile Ala An Ala Glu His An Asp Gly Ala Asp Ile Asp
TTS ACC TAC ACC AAC AAG AAA GCB BCT STA CTS BTC AAG AAA BTT CTS GAA TCT GCC ATT BCT AAC BCT BAA CAC AAC BAA GBT BGC BCT BAC ATT BAC
3400

Asp Leu Lys Val Thr Lys Ile Phe Val Asp Glu Gly Pro Ser Met Lys Arg Ile Met Pro Arg Ala Lys Gly Arg Ala Asp Ile Leu Lys Arg
BAT CTS AAA GTT ACG AAA ATT TTC GTA BAC BAA BGC CCB AOC ATB AAG CCB ATT ATG CCB CBT GCA AAA GGT CBT GCA BCT AAG AAA BAA ATG BAA GBT
3500

Thr Ser His Ile Thr Val Val Val Ser Asp Arg rpsC S3 Met Gly Gln Lys Val His Pro An Gly Ile Arg Leu Gly Ile Val
ACC ABC CAC ATC ACT BTA GTT BTA TCC BAT CBC TBAACTCTBBAAGACTABCA ATG BGT CAG AAA GTA CAT CCT AAT GBT ATT CCB CTS BGT ATT GTA
3600

Lys Pro Trp An Ser Thr Trp Phe Ala An Thr Lys Glu Phe Ala Asp An Leu Asp Ser Asp Phe Lys Val Arg Gln Tyr Leu Thr Lys Glu Leu
AAA CCA TGG AAC TCT ACC TGG TTT CCB AAC ACT CCA AAA BAA TTT CAC BAC AAC CTS BAC ABC BAA TTT AAA GTA CBT AAC TAC CTS ACT AAG BAA CTS
3700

Ala Lys Ala Ser Val Ser Arg Ile Val Ile Glu Arg Pro Ala Lys Ser Ile Arg Val Thr Ile His Thr Ala Arg Pro Gly Ile Val Ile Gly Lys
BCT AAA GCB TCC BTA TCT CBT ATC BTT ACT AAG CBT CCB BCT AAG ABC ATC CBT BTA ACC ATT CAC ACT BCT CCB CCB BGT ATC BTT ACT BGT AAA
3790

Lys Gly Glu Asp Val Glu Lys Leu Arg Lys Val Val Ala Asp Ile Ala Gly Val Pro Ala Gln Ile An Ile Ala Glu Val Arg Lys Pro Glu Leu
AAA BGT BAA BAC CBT ABA AAA ATT CAG CTS AAG BCT AOC BCT BGC BTT CCT GCA CAC BCT AAC ATC GCC BAA GTT CBT AAG CCA BAA CTS
3800

Asp Ala Lys Leu Val Ala Asp Ser Ile Thr Ser Gln Leu Glu Arg Arg Val Met Phe Arg Arg Ala Met Lys Arg Ala Val Gln An Ala Met Arg
GCA GCA AAA CTS BTT BCT BAC AOC ATC ACT TCT CAC CTS BAA CBT CCB GTT ATT TTC CBT CBT GCT ATB AAG CBT BCT GTA CAG AAC GCA ATC CBT
3900

Leu Gly Ala Lys Gly Ile Lys Val Glu Val Ser Gly Arg Glu Gly Ile Ala Glu Ile Ala Arg Thr Gly Trp Tyr Arg Glu Thr Val Pro Leu
CTS BGC BCT AAA BGT ATT AAA GTT BAA GTT AAG BGC CBT CTS BGC BGC BAA ATC GCA CBT ACC BAA TGG TAC CCB BAA BGT CCB GTA CCB CTS
4000

His Thr Leu Arg Ala Asp Ile Asp Tyr An Thr Ser Glu Ala His Thr Thr Tyr Gly Val Ile Bly Val Lys Val Trp Ile Phe Lys Gly Glu Ile
CAC ACT CTS CBT BCT BAC ATC BAC TAC AAC ACC TCT BAA GCB CAC ACC ACT TAC BGT GTA CBT BCT ATT AAA BTA TGG TAC TTC AAA BGC BAC ATC
4100

Leu Gly Gly Met Ala Ala Val Glu Gln Pro Glu Lys Pro Ala Ala Gln Pro Lys Lys Gln Gln Arg Lys Gly Arg Lys rplP L14 Met
CTS BGT BGT ATB BCT BCT GTT BAA CAA CCB BAA AAA CCB BCT BCT CAG CCT AAA AAG CAG CAG CBT AAA GCB CBT AAA TAAGSACBCTCCTB ATG
4200

Leu Gln Pro Lys Arg Thr Lys Phe Arg Lys Met His Lys Gly Arg An Arg Gly Leu Ala Gln Gly Thr Ser Val Ser Gly Ser Phe Gly Leu
TTA CAA CCA AAG CBT ACR AAA TTC CBT AAA ATB CAC AAA BGC CBT AAC CCB BGT CTS BCB CAG BGT ACG BAT BTT AAG TTC GGC AAC TTC BGT CTS
4300

Lys Ala Val Gly Arg Gly Arg Leu Thr Ala Arg Gln Ile Glu Ala Ala Arg Arg Ala Met Thr Arg Ala Val Lys Arg Gln Gly Lys Ile Trp Ile
AAA BCT GTT BGC CBT BGT CBT CTS ACT BCC CBT CAG ATC BAA GCA GCA CBT CBT GCT ATG ACC CBT GCA GTT AAG CBT CAA GGT AAG ATC TGG ATC
4400

Arg Val Phe Pro Asp Lys Pro Ile Thr Lys Glu Pro Leu Ala Val Arg Met Gly Lys Gly Lys Gly An Val Glu Tyr Trp Val Ala Leu Ile Gln
CST BTA TTC CCB BAA CCB AAG CCB ATC ACT BAA AAG CCB CTS GCA BTA CBT ATG BGT AAG BGT AAA BGT AAG CBT GAG TAT TGG GTT CCB TTG ATT CAG
4500

Pro Gly Lys Val Leu Tyr Glu Met Asp Gly Val Pro Glu Glu Leu Ala Arg Glu Ala Phe Lys Leu Ala Ala Lys Leu Pro Ile Lys Thr
CTS BGT AAA BTC CTS TAT BAA ATC BAC BGT GTT CCB BAA BAC CTS BCC CBT BAA BCA TTC AAG CTS BCA GCA GCB AAA CCB CBT ATT AAA ACC ACC
4600

Phe Val Thr Lys Thr Val Met L29 Met Lys Ala Lys Glu Leu Arg Glu Lys Ser Val Glu Glu Leu An Thr Glu Leu Leu An Leu Leu Arg Glu
TTT GTA ACT AAG ACG BTA ATB TA ATB AAA GCA AAA BAC CBT CBT BAC AAG ABC BTT BAA GAG CTS AAC ACC GAG CTS CTS AAC CTS CTS CBT BAC
4700

Gln Phe An Leu Arg Met Gln Ala Ser Gly Gln Leu Gln Gln Ser His Leu Leu Lys Gln Val Arg Arg Asp Val Ala Arg Val Lys Thr Leu
CAG TTC AAC CTS BCT ATB CAG BCT BCA AAT BGC CAG CTS CAA CAG TCT CAC CTS TTS AAG CTA BCC CBT BAA BCA TTC AAG CTS BCA GCA GCB AAA CCB CBT ATT AAA ACT TTA
4800

Leu An Glu Lys Ala Gly Ala Met Thr Asp Lys Ile Arg Thr Leu Leu Gln Gly Arg Val Val Ser Asp Lys Met Glu Lys Ser Ile Val Val Ala
CTS AAC BAA AAG BCB BGT CCB TA ATB ACC BAT AAA ATC CBT ACT CTS CAA BGT CCB GTT BTT ABC CAG AAA ATB BAA AAA TCC ATT GTT BTT CBT
4900

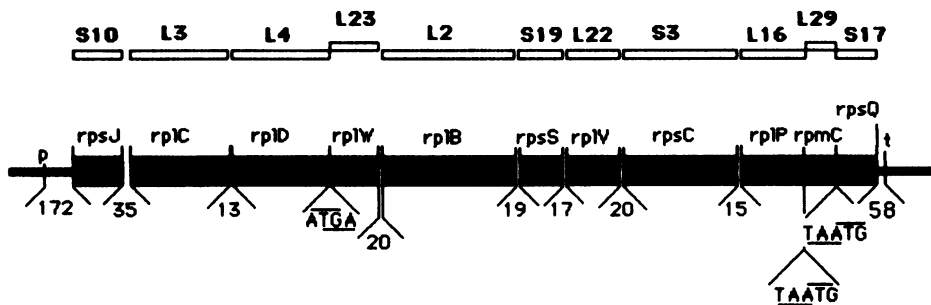
Ile Glu Arg Phe Val Lys His Pro Ile Tyr Gly Lys Phe Ile Lys Arg Thr Thr Lys Leu His Val His Asp Glu An An Glu Cys Gly Ile Gly
ATC BAA CBT TTT BTA AAA CAC CCB ATC TAC BGT AAA TTC ATC AAG CBT ACS ACC AAA CTS CAC GTA CAT BAC BAC AAG AAC BAA TBC BGT ATC BGT
5000

Asp Val Val Glu Ile Arg Glu Cys Arg Pro Leu Ser Lys Thr Lys Ser Trp Thr Leu Val Arg Val Val Glu His Ala Val Leu
BAC BTA BTT BAA ATC CCB BAA TBC CBT CCB TCC AAG ACT AAA TCC TGG ACG CTS BTT CBT BTA BAA AAA CCB GTT CTS TAA TACBACTACT
5100

CTCTCAATACBAAATACBCBCTCAGAAATBAGCBTTTATTTTTCTACCCATATCCTTBAAGCBGTBTATAATGCCBCCCTCBATATBGBBATTTTTACBACCTGATTTTCBGBCTCAGTA
5200

```

**Fig. 1.** The nucleotide sequence of the *E. coli* S10 operon commencing proximal to its promoter region and ending at the *HincII* 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 $\delta$	Coordinates <sup>†</sup>
<i>rpsJ</i>	GGAG	10	59-62
<i>rpIC</i>	GAGGT	7	505-509
<i>rpID</i>	TAAGGAG	6	1144-1150
<i>rplW</i>	AGGAG	8	1746-1750
<i>rpIB</i>	GGAGG	9	2065-2069
<i>rpsS</i>	GAGG	7	2906-2909
<i>rplV</i>	AGGAGG	5	3199-3204
<i>rpsC</i>	GGAG	7	3549-3552
<i>rplP</i>	TAAGGAG	8	4259-4265
<i>rpmC</i>	TAAG--GGTGA	4	4669-4679 <sup>**</sup>
<i>rpsQ</i>	AAGG	10	4861-4864

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

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

<sup>†</sup> 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 Discrepancies

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 pN02003 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 *rp1C*, *rpsS*, *rpsC* and the *Bam*H1 site in *rp1D*. 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 discrepancies (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. Scheuer 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., Strycharcz, 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. Sanger, 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., Münde, 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.