Volume 16 Number 7 1988 Nucleic Acids Research

Cloning and characterization of a gene from *Rhizobium melilotii 2011* coding for ribosomal protein S1

Joachim Schnier[1]*, Sabine Thamm, Rudi Lurz, Atta Hussain[2], Gabriele Faist[3] and Beate Dobrinski

Max-Planck-Institut für Molekulare Genetik, D-1000 Berlin 33 (Dahlem), Ihnestrasse 73, FRG

Received November 11, 1987; Accepted March 2, 1988

ABSTRACT

A 7 kb chromosomal DNA fragment from R. melilotii was cloned, which complemented temperature-sensitivity of an E. coli amber mutant in rpsA, the gene for ribosomal protein S1 (ES1). From complementation and maxicell analysis a 58 kd protein was identified as the homolog of protein S1 (RS1). DNA sequence analysis of the R. melilotii rpsA gene identified a protein of 568 amino acids, which showed 47% identical amino acid homology to protein S1 from E. coli. The RS1 protein lacked the two Cys residues which had been reported to play an important role for the function of ES1. Two repeats containing Shine-Dalgarno sequences were identified upstream of the structural gene. Binding studies with RNA polymerase from E. coli and Pseudomonas putida located one RNA-polymerase binding site close to the RS1 gene and another one several hundred basepairs upstream. One possible promoter was also identified by DNA sequence comparison with the corresponding E. coli promoter.

INTRODUCTION

Ribosomal protein S1 from E. coli is the largest of all ribosomal proteins with a molecular weight of 61 kd. It was found to stimulate translation by facilitating mRNA binding to ribosomes (1). The gene rpsA and the protein have been sequenced (2,3,4). The sequence data and functional studies allowed a division of protein S1 into two domains of which the largest one is an RNA binding domain which consists of four repeats, with high internal amino acid and DNA sequence homology (5,6). The rpsA gene is preceded by an open reading frame coding for a 25 kd protein in which several promoters for the S1 gene were located (7).

The expression of protein S1 is translationally controlled by an autoregulatory mechanism (8). Recently, several rpsA gene fragments were expressed from plasmids. It was shown that overproduction of the 100 N-terminal amino acids of the protein re-

© IRL Press Limited, Oxford, England.

duced the amount of chromosomally encoded protein S1, which suggested that this fragment significantly contributes to the regulation of the synthesis of protein S1 (5).

Since protein S1 has an essential function in E. coli (9) and is one of the major host components of the replicase complex of bacterial RNA-phages like MS2 (10), we regarded it to be important to analyse S1 genes from other organisms. Hybridization experiments suggested that most or all Gram-negative bacteria have a gene which codes for a protein similar to protein S1 from E. coli (11). But we did not find evidence for an rpsA like gene in Bacillus or other Gram-positive bacteria which supported an earlier finding that antibodies against protein S1 did not react with protein extracts from Bacillus stearothermophilus (12).

In an attempt to continue this comparative study we decided to clone an rpsA gene from an organism which is not closely related to E. coli. We chose Rhizobium melilotii, because Southern-blot analysis had revealed that among all Gram-negative bacteria chromosomal DNA from R. melilotii gave the weakest signal after hybridization with a structural gene fragment from the E. coli gene for protein S1 (11). Therefore, our aim was to find out whether R. melilotii has a gene which corresponds to the rpsA gene from E. coli.

In this paper we describe the cloning, expression, promoter and DNA sequence analysis of the gene rpsA from Rhizobium melilotii and the identification of its gene product, protein RS1.

MATERIALS AND METHODS

Bacterial strains, plasmids and culture media.

Strain Rhizobium melilotii 2011 was used as a source for chromosomal DNA. The recipient E. coli strain for cloning was strain FS10 (F^-, leu_{am}, trp_{am}, $lac2_{am}$, $galK_{am}$, galE, tsx, relA, supD43, 74, rpsL, sueA, sueB, sueC, $rpsA_{am}$, hsdR17, recA, srl::Tn10). This strain was derived from mutant MB3001 (9) by transducing hsdR17 from strain MM294 (F^-, thi1, endA1, hsdR17, supE44). Then recA was introduced from strain UH748 (F506/UH69, aroD, man, pps, thi, rpsL, supF(?), trg::Tn5, recA, srl::Tn10). For maxicell analysis strain CSR603 (F^-, thrA, leuB6, proA2,

phr1, recA1, argE3, thi1, uvrA6, ara14, lacY1, galK2, xyl5, mtl1, rpsL31, tsx33, λ^-, supE44) was used. The cloning vector was plasmid pACYC184. The derivatives of plasmid pACYC184 are listed in Fig. 1. Bacteria were grown in L-Broth. Antibiotics, if necessary, were added: chloramphenicol (25 µg/ml) and tetracycline (10 µg/ml) final concentration.

Enzymes and chemicals

Restriction endonucleases, DNA ligase and 'Klenow' fragment of DNA polymerase I were purchased from Bethesda Research Laboratories, alkaline phosphatase, E. coli RNA-polymerase and exonuclease Bal31 from Boehringer Mannheim, deoxy- and dideoxynucleotides from Pharmacia. Pseudomonas putida RNA polymerase was a gift from Nicola Mermod (University of Geneva). [^{35}S-]methionine (800 Ci/mmol) and [^{35}S-]deoxy ATP (1000 Ci/mmol) were from Amersham Corp.

Preparative and analytical techniques

DNA techniques were generally carried out as described by Maniatis et al. (13). Chromosomal DNA was isolated with Sarcosyl LN97 according to Chow et al. (14). DNA sequence analysis was carried out according to Sanger et al. (15) using M13 phages (16). For autoradiography gels were dried and exposed to X-ray film (Fuji Film, Japan). RNA polymerase binding studies were carried out as described (17). Maxicell analysis to detect plasmid encoded gene products was carried out according to Sancar et al. (18). Proteins were separated by 15% SDS gel-electrophoresis according to Laemmli (19).

RESULTS

Cloning of the rpsA gene from R. melilotii

In order to clone a possible rpsA gene from R. melilotii we made use of the fact that there was weak hybridization of chromosomal DNA from R. melilotii with a 1.2 kb structural gene fragment from E. coli as a probe (11). The size of the DNA fragment after digesting with restriction endonuclease HindIII was estimated to be about 7 kb. We therefore digested chromosomal DNA from R. melilotii 2011 with HindIII and separated the DNA by agarose gel-electrophoresis. DNA of the size of 7 kb was eluted

from the gel and ligated into plasmid pACYC184 which was pre-treated with alkaline phosphatase.

If there is an rpsA gene in R. melilotii which codes for a protein which is functionally homologous to the E. coli protein S1, it might be possible to find complementation of a mutation in the E. coli rpsA gene. Therefore, we transformed the E. coli strain FS10, which is a derivative of MB3001 harbouring a temperature-sensitive amber suppressor and an amber mutation in rpsA (9). This strain is unable to grow at 42°C. We selected for temperature- and chloramphenicol-resistance. Six clones were obtained from which we isolated plasmids and retransformed mutant FS10. All plasmids were able to complement the amber mutation in rpsA. Analysis of the plasmid DNA showed that they all posessed an identical 7 kb DNA fragment which differed in its orientation on the plasmid. We concluded that the fragment contained a gene which was most probably homologous to the rpsA gene from E. coli.

Identification of the rpsA gene

In order to identify the location of the S1 gene on the 7 kb DNA fragment we subcloned fragments which expressed a functional protein RS1. We digested plasmids pFS300 and pFS301 with BamHI, since the fragments generated were quite large compared to the ones created by other endonucleases. Plasmid pFS300 (Fig. 1) was, therefore, digested with BamHI, religated and then strain FS10 was transformed. We screened the transformants for complementation of the rpsA amber mutation and isolated plasmids from these transformants and also from those which did not show complementation. We obtained plasmid pFS302 which had lost the 2.8 kb BamHI fragment and was still able to complement temperature-sensitive growth of strain FS10 to the same extent as the original plasmid pFS300 (Fig. 1, plasmid pFS302). Plasmid pFS303 which had also lost the 2.8 kb fragment but which contained the 3.8 kb BamHI fragment in the opposite orientation as the original plasmid was unable to support growth of strain FS10 at 42°C. A derivative of plasmid pFS302 which had maintained only a 1.7 kb fragment from the HindIII site to the first SalI site had also lost the ability to complement the S1 amber mutation (pFS304). We attempted deletion of the PstI fragment of plasmid pFS302. Some clones which we obtained were unable to complement the S1 amber mutation (e.g.

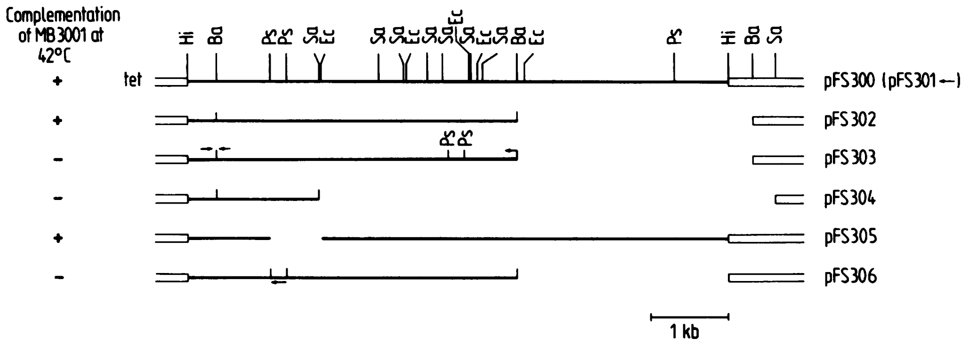

Fig. 1: Restriction map of plasmid pFS300 and its derivatives. The thin line is the DNA insert. The thick line is the plasmid part. The arrows indicate a reverse orientation of a fragment. The symbols are: Ba, BamHI; Ec, EcoRI; Hi, HindIII; Ps, PstI; Sa, SalI for restriction endonucleases and tet for tetracycline resistance gene.

pFS306). Restriction endonuclease analysis of such plasmids did not show any apparent difference to plasmid pFS302 itself. None of the plasmids had lost the PstI fragment. Our conclusion was, therefore, that plasmid pFS306 had the PstI fragment in the opposite orientation.

When plasmid pFS301 (Fig. 1) was digested with BamHI, religated and the rpsA amber mutant was transformed, we did not obtain any plasmid which complemented the amber mutation. This result indicated that the BamHI fragment alone was not sufficient for complementation which was in accordance with the above result. By chance we obtained a plasmid (pFS305) derived from plasmid pFS300 containing a rearrangement of fragments, with a small deletion which altered the flanking HindIII/SalI 1.7 kb fragment (Fig. 1). This plasmid could still complement the amber mutation. These results suggested that DNA from a part of the HindIII site in plasmid pFS302 continuing to the second BamHI site is required for the complementation of temperature-sensitive growth of mutant FS10.

Another possibility to identify the location of the RS1 gene is Southern Blot analysis. We digested plasmids pFS300 and pFS301 with various restriction endonucleases, separated the fragments by agarose electrophoresis and used a 1.2 kb structural gene fragment from the E. coli S1 gene as a probe (11). The 3.8 kb

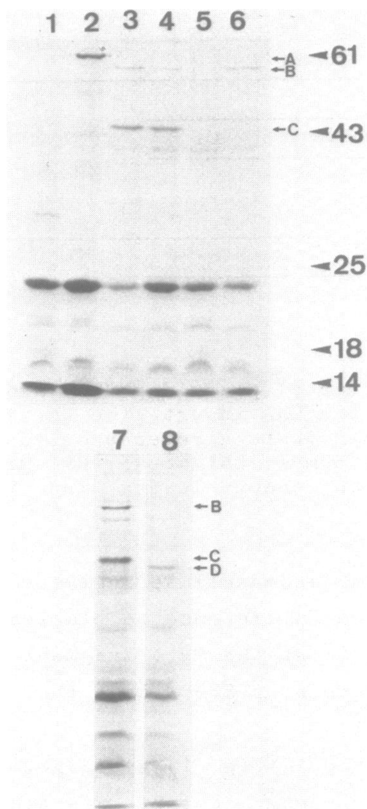

Fig. 2: Maxicell analysis of various plasmids. Samples were applied on a 15% SDS-gel and proteins separated. The numbers next Ato the arrowheads refer to the molecular weight in kilodalton. The symbols are (A) protein S1 from E. coli, (B) 58 kd polypeptide, (C) 43 kd polypeptide and (D) 40 kd polypeptide. Extracts were derived from cells containing plasmids pACYC184 (lane 1), pJS200 (lane 2), pFS300 (lane 3), pFS301 (lane 4), pFS303 (lane 5), pFS304 (lane 6), pFS302 (lane 7) and pFS305 (lane 8).

BamHI fragment, the 0.9 kb EcoRI fragment as well as several smaller (0.2 to 0.3 kb) SalI fragments gave a signal (data not shown). This showed the location of the RS1 structural gene as indicated in Fig. 3.

Maxicell analysis of various plasmids

In order to identify the polypeptides encoded by the various plasmids we performed maxicell analysis. The proteins were separated by SDS gel-electrophoresis (Fig. 2). Both plasmids, pFS300

and pFS301, containing the 7 kb HindIII fragment in different orientations on the plasmid expressed at least two proteins (-lanes 3 and 4) which were not expressed by plasmids pACYC184 or pJS200, a plasmid containing the rpsA gene from E. coli (lanes 1 and 2). These proteins had a molecular weight of about 43 kd and 58 kd. Since both proteins were also expressed from plasmid pFS302 (lane 7) the corresponding genes were localized to the 4.2 kb HindIII / BamHI DNA fragment (Fig. 1). Plasmid pFS303 which had the same BamHI fragment as pFS302 but in the opposite orientation gave rise to a very weak band at 58 kd, but the 43 kd protein was completely missing (lane 5). Plasmid pFS305 expressed one protein band at 58 kd (lane 6). Plasmid pFS304 with a DNA fragment extending from the HindIII site of pFS302 to the first SalI site of 1.7 kb length gave rise to a new protein of about 40 kd molecular weight. From these results we concluded that only plasmids expressing this protein were able to complement temperature-sensitive growth caused by the amber mutation in the rpsA gene in strain FS10.

Identification of RNA-polymerase binding sites

The E. coli rpsA gene is preceded by four promoters clustered in two regions. Two promoters are comparable in strength with the promoters of the tet gene (7). For comparison, we wanted to localize possible promoters on the fragment containing the RS1 gene. One of the possibilities is to determine the number of RNApolymerase molecules which are bound specifically to certain regions of the DNA by the aid of electron microscopy as it is shown in Fig. 3A. We could identify the plasmid vector promoters among which P1 to P4 have been previously described (23). However, the number of RNA-polymerase molecules bound to the inserted DNA fragment from R. melilotii under the same conditions was several fold lower indicating that RNA-polymerase from E. coli has a low affinity for insert promoters. But we could localize at least three binding regions for RNA-polymerase which are called P6, P7 and P8.

We repeated the experiment with plasmid pFS302, digested with BamHI and selected only BamHI fragments which had bound RNA-polymerase molecules neglecting the vector part (Fig. 3B). We confirmed the binding sites P6 and P7, and we found another weak

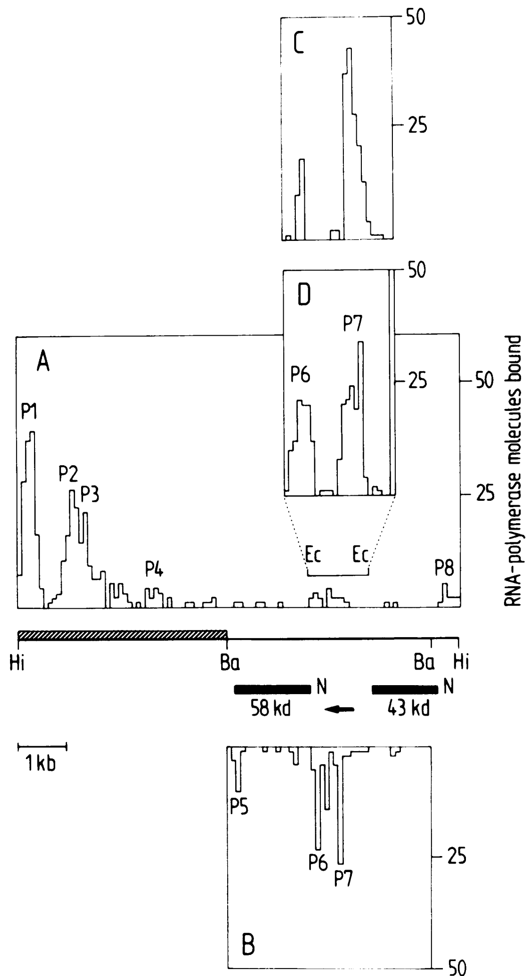

Fig. 3: RNA-polymerase binding sites on plasmid pFS302. A: plasmid pFS302 was linearized with HindIII after the binding of RNA-polymerase. B: plasmid pFS302 was digested with BamHI after the binding of RNA-polymerase and only DNA fragments with RNA-polymerase bound were selected; C: the binding assay was done with purified 1.1 kb EcoRI fragment and RNA-polymerase from *Pseudomonas putida*; D: as C but with RNA-polymerase from *E. coli*. P1 to P4 are plasmid promoters as reported (23). The dashed line is the plasmid part and the thin line the insert DNA. The arrow shows the direction of transcription for the 43 kd and the 58 kd protein genes as concluded from various data. The symbols for restriction endonucleases are those described in Fig. 1.

binding site (P5) which, however, may be an artefact. With the aid of purified DNA fragments the two sites P6 and P7 were located more precisely on a 1.1 kb EcoRI fragment with *E. coli* RNA-polymerase (Fig. 3D). We also observed a strong binding peak at one end of the fragment which, however, is most probably an artefact. We carried out the binding studies with RNA-polymerase from *Pseudomonas putida* and obtained a similar result (Fig. 3C). It should be noted, however, that there is a clear difference in the heights of the peaks in Fig. 3C but much less in Fig. 3D. From our previous analysis we concluded that P6 and P7 are located upstream from the 58 kd coding region which we identified as the structural gene for protein S1 but several hundred nucleotides away from each other so that most probably only one of the two binding sites belongs to the RS1 gene.

DNA sequence of the RS1 gene

In order to elucidate the fine structure of the RS1 gene we determined the DNA sequence of the relevant regions. We detected an open reading frame whose corresponding protein sequence showed homology to the ES1 gene (Fig. 4). The number of amino acids deduced from the DNA sequence was 567 compared to the *E. coli* S1 gene with 556 amino acids. This was surprising, since the RS1 protein showed a smaller molecular weight than the ES1 protein. Differences in the number of residues were found at the N-terminal end, which is longer in case of the RS1 protein, and at the C-terminal end (see also Fig. 5). A comparison between the amino acid sequences showed an overall 47% identical homology, with a higher local homology from amino acid position 195 to about 450 of the RS1 protein. This region which forms the RNA binding domain of protein S1 with several repeats of internal homology (20) had 58% identical amino acids in the proteins from both organisms. It is also interesting to note that the RS1 lacks the two Cys residues.

Behind the stop codon of the RS1 gene we found a palindromic structure which is followed by a row of T bases. This is typical for a rho-independent terminator (22). We also investigated the region preceding the RS1 gene for possible regulatory sequences. We found two repeats of 9 basepair identity both of which contained Shine-Dalgarno sequences (Fig. 4). Furthermore, the

GCACTCGATCGACGACCGTGCTTTCCGAGCATTCGATCGGCGAGGACGGCGTCGAAGATCGCGTCATGCCGGCGGTGCGCCG

GCGCTGGTCGAGGCGCACGGAGCCTTCGCGCGAAGGAGCCGGGTACCGTCCTCGACGGCGCGACATCGGCACGGTCGTCTGC

CCTGATGCGGCGGTGAAGCTCTATGTGACCGCTTCCCCCGAAGTGCGGGCGAAGCGGCGCCATGACGAGATCGTGCCGGCGG

CGGAAAGGCTGACTACACCGCGATTTTCGAGGACGTGAAGAAGGCCGACGGGCGCGACATGGGGCGTGCCGACAGTCCTTTG

AGACCTGCCGAGGACGCGCACTTGCTTGATACTTCCGAAATGAGTATAGAGGCGGCATTCCAGGCGGCGAAGACGCTGGTCG

ACGCCGCCCTGAAAGAGAAAATTTGAAGAAGGCCTGATCCCGGCCCAGCCGGTCTCCTTCTTGAAAATAGCCTGAAATTCCG

-10

TCCACGCGCCGGATTGCTCCATGACAGGAGCGGACTGGGTTCAGGCCTGTCTAACACCAACCCCGGCGCATATGCGTCTCCG

GCAGGAGATGCAGCAGGAGATTAT ATG TCT GCA ACC AAC CCC ACC CGT GAT GAT TTC GCA GCG CTT

Met Ser Ala Thr Asn Pro Thr Arg Asp Asp Phe Ala Ala Leu

CTG GAA GAG TCC TTC GCC AAG ACG GAC CTC GCC GAA GGC TAT GTC GCC AAG GGC ATC GTC

Leu Glu Glu Ser Phe Ala Lys Thr Asp Leu Ala Glu Gly Tyr Val Ala Lys Gly Ile Val

ACG GCG ATC GAA AAA GAC GTC GCA ATC GTC GAC GTC GGC CTG AAG GTC GAA GGC CGC GTA CCG

Thr Ala Ile Glu Lys Asp Val Ala Ile Val Asp Val Gly Leu Lys Val Glu Gly Arg Val Pro

CTG AAG GAA TTC GGC GCC AAG GCC AAG GAC GGC ACG CTG AAG GTC GGC GAC GAA GTC GAA GTC

Leu Lys Glu Phe Gly Ala Lys Ala Lys Asp Gly Thr Leu Lys Val Gly Asp Glu Val Glu Val

TAT GTC GAG CGC ATT GAA AAC GCA CTC GGC GAA CGA GTC CTC TCG CGC GAG AAG GCC CGC CGC

Tyr Val Glu Arg Ile Glu Asn Ala Leu Gly Glu Arg Val Leu Ser Arg Glu Lys Ala Arg Arg

GAA GAG AGC TGG CAG CGC CTG GAA GTC AAG TTC GAA GCC GGC GAA CGC GTC GAA GGC ATC ATC

Glu Glu Ser Trp Gln Arg Leu Glu Val Lys Phe Glu Ala Gly Glu Arg Val Glu Gly Ile Ile

TTC AAC CAG GTC AAG GGC GGC TTC ACC GTC GAT CTC GAC GGC GCC GTA GCC TTC CTG CCG CGT

Phe Asn Gln Val Lys Gly Gly Phe Thr Val Asp Leu Asp Gly Ala Val Ala Phe Leu Pro Arg

TCG CAG GTC GAC ATC CGT CCG ATC CGC GAC GTG ACG CCT GCT GAT GCA CAA CCC GCA GCC CTT

Ser Gln Val Asp Ile Arg Pro Ile Arg Asp Val Thr Pro Ala Asp Ala Gln Pro Ala Ala Leu

CGA AAT CTC AAG ATG GAC AAG CGC CGC GGC AAC ATC GTC GTT TCG CGC CGC ACG GTT CTC GAA

Arg Asn Leu Lys Met Asp Lys Arg Arg Gly Asn Ile Val Val Ser Arg Arg Thr Val Leu Glu

GAG TCC CGC GCC GAG CAG CGT TCT GAA ATC GTT CAG AAC CTC GAG GAA GGC CAG GTT GTC GAG

Glu Ser Arg Ala Glu Gln Arg Ser Glu Ile Val Gln Asn Leu Glu Glu Gly Gln Val Val Glu

GGT GTC GTC AAG AAC ATC ACC GAT TAC GGT GCG TTC GTC GAC CTC GGC GGC ATC GAC GGC CTG

Gly Val Val Lys Asn Ile Thr Asp Tyr Gly Ala Phe Val Asp Leu Gly Gly Ile Asp Gly Leu

CTG CAC GTC ACC GAC ATG GCA TGG CGC CGC GTC AAA CAT CCG TCG GAG ATC CAG AAC ATC GGC

Leu His Val Thr Asp Met Ala Trp Arg Arg Val Lys His Pro Ser Glu Ile Gln Asn Ile Gly

CAG CAG GTC AAG GTT CAG ATC ATC CGC ATC AAC CAG GAA ACC CAC CGC ATC TCG CTC GGC ATG

Gln Gln Val Lys Val Gln Ile Ile Arg Ile Asn Gln Glu Thr His Arg Ile Ser Leu Gly Met

AAG CAG CTC GAG TCC GAT CCG TGG GAT GGC ATC GGT GCG AAG TAT CCG GTC GGC AAG AAG ATC

Lys Gln Leu Glu Ser Asp Pro Trp Asp Gly Ile Gly Ala Lys Tyr Pro Val Gly Lys Lys Ile

TCC GGT ACC GTC ACG AAC ATC ACC GAC TAC GGT GCG TTC GTC GAG CTG GAG CCG GGC ATC GAA

Ser Gly Thr Val Thr Asn Ile Thr Asp Tyr Gly Ala Phe Val Glu Leu Glu Pro Gly Ile Glu

GGC CTG ATA CAC ATC TCC GAA ATG TCC TGG ACG AAG AAG AAC GTT CAC CCC GGC AAG ATC CTG

Gly Leu Ile His Ile Ser Glu Met Ser Trp Thr Lys Lys Asn Val His Pro Gly Lys Ile Leu

TCC ACG AGC CAG GAA GTC GAC GTG GTC GTT CTC GAA GTC GAC CCG ACC AAG CGC CGC ATC TCG

Ser Thr Ser Gln Glu Val Asp Val Val Val Leu Glu Val Asp Pro Thr Lys Arg Arg Ile Ser

CTC GGC CTC AAG CAG ACG CTC GAG AAC CCG TGG CAG GCC TTC GCG CAT AGC CAT CCG GCT GGC

Leu Gly Leu Lys Gln Thr Leu Glu Asn Pro Trp Gln Ala Phe Ala His Ser His Pro Ala Gly

ACG GAA GTC GAA GGC GAA GTC AAG AAC AAG ACC GAA TTC GGT CTG TTC ATC GGC CTC GAT GGC

Thr Glu Val Glu Gly Glu Val Lys Asn Lys Thr Glu Phe Gly Leu Phe Ile Gly Leu Asp Gly

GAT GTC GAC GGC ATG GTC CAC CTC TCC GAC CTC GAC TGG AAC CGC CCG GGC GAG CAG GTC ATC

Asp Val Asp Gly Met Val His Leu Ser Asp Leu Asp Trp Asn Arg Pro Gly Glu Gln Val Ile

GAA GAA TTC AAC AAG GGC GAC GTC GTC CGT GCT GTC GTT CTC GAT GTG GAC GTC GAC AAG GAG

Glu Glu Phe Asn Lys Gly Asp Val Val Arg Ala Val Val Leu Asp Val Asp Val Asp Lys Glu

CGC ATC TCG CTC GGC ATC AAG CAG CTC GGC CGT GAT GCG GTC GGT GAA GCT GCC GCT TCC GGC

Arg Ile Ser Leu Gly Ile Lys Gln Leu Gly Arg Asp Ala Val Gly Glu Ala Ala Ala Ser Gly

GAA CTG CGC AAG AAC GCC GTC GTT TCG GCT GAA GTC ATC GGC GTC AAC GAT GGC GGC ATC GAA

Glu Leu Arg Lys Asn Ala Val Val Ser Ala Glu Val Ile Gly Val Asn Asp Gly Gly Ile Glu

GTG CGG CTC GTC AAC CAC GAG GAC GTC ACC GCC TTC ATC CGT CGC GCC GAC CTC TCG CGC GAT

Val Arg Leu Val Asn His Glu Asp Val Thr Ala Phe Ile Arg Arg Ala Asp Leu Ser Arg Asp

CGC GAC GAA CAG CGT CCG GAG CGT TTC TCG GTC GGC CAG ACC GTC GAT GCG CGG GTC ACC AAC

Arg Asp Glu Gln Arg Pro Glu Arg Phe Ser Val Gly Gln Thr Val Asp Ala Arg Val Thr Asn

TTC TCC AAG AAG GAC CGC AAG ATC CAG CTG TCG ATC AAG GCT CTG GAA ATC GCG GAA GAG AAG

Phe Ser Lys Lys Asp Arg Lys Ile Gln Leu Ser Ile Lys Ala Leu Glu Ile Ala Glu Glu Lys

GAA GCC GTC GCT CAG TTC GGC TCT TCC GAC TCG GGT CGT TCG CTC GGC GAC ATT CTG GGC GCT

Glu Ala Val Ala Gln Phe Gly Ser Ser Asp Ser Gly Arg Ser Leu Gly Asp Ile Leu Gly Ala

GCG CTC AAG AAC CGC CAG AAC AAC GAG TAA TCGTTGACGGACTTGAAATGATAAGCCCGCGGAGAGCGATCTC

Ala Leu Lys Asn Arg Gln Asn Asn Glu •••

CGCGGGCTTTTTCGTGGCAGGGCGAAAACCGTTCCGCACTTTTCCTGAAAGTGCTCCAAGGCCTGTCGAGATTCGGCATAGC

CGCCGCGGCTGGCTTGGCTTGATCTCCTCATCCCTGTGCTCGTCATCAGGGATC

repeats were partly contained within a secondary structure (see also Fig. 6). Since we mapped RNA-polymerase binding sites in the neighbourhood of the RS1 gene, we looked for promoter sequences by comparing the RS1 sequence with the corresponding sequence in the E. coli rpsA gene. We found a sequence showing homology to a region in the ES1 gene which is identical to promoter P3 (7). We did not detect any other promoter-like sequences or other striking homologies to the ES1 gene.

DISCUSSION

We have cloned and identified a gene from R. melilotii corresponding to ribosomal protein S1 from E. coli. Not only is there sequence homology between the two genes but the protein products are functionally interchangeable *in vivo*. There are, however, a number of differences between the proteins and the genes. One of the differences is the molecular weight. Although the migration of the RS1 protein suggests a smaller molecular weight than ES1, the deduced amino acid sequence would suggest a larger molecular weight. It could either be that the RS1 protein is proteolytically digested in E. coli or that its synthesis is prematurely terminated. In both cases the C-terminal portion of S1 should be affected, since the N-terminal part is essential for the function (5). A third possibility could be a migration artefact. It is known that also the ES1 protein does not migrate at the molecular weight position where it is expected. Instead its molecular weight was estimated to be about 10000 dalton larger than other proteins with a similar molecular weight (e.g. 24).

A comparison between both proteins revealed that like the ES1 protein (6), the RS1 protein also can be subdivided into regions of internal homologies. Those are from amino acid position 196-279 (R1), 280-366 (R2) and 367-453 (R3, for positions see Fig. 5). In contrast to the ES1 protein the fourth repeat at the C-terminal end shows low homology to the other repeats. This part

Fig. 4: DNA sequence of the rpsA gene from R. melilotii. The boxed sequences contain possible promoters. The dots above the bases show identical residues to the corresponding E. coli rpsA gene sequence (3,7). The thin lines are two repeats. The arrows point to palindromic structures.

```
RS1   1 MSATNPTRDD FA A L L EES FAKTDLAE G YVAK G I V T AI E KDV AI VD V GLK V E GRV P LKE F GAKAKD 65
ES1   1      MTES FA Q L F EES LKEIETRP G SIVR G V V V AI D KDV VL VD A GLK S E SAI P AEQ F KNAQGE 59

RS1  66 GTLK VGDEV E V YVERI E NAL GE RV LSREKA R R E E S W QR LE VKF E AG E R V E G I I FNQ VKGGFT 128
ES1  60 LEIQ VGDEV D V ALDAV E DGF GE TL LSREKA K R H E A W IT LE KAY E DA E T V T G V I NGK VKGGFT 122

RS1 129 V D LDG AV AFLP R S Q VD I RP I RD VTPADAQPAALRNL K M D KR R G N I VVSRR T V L E ESRAEQ R S 190
ES1 123 V E LDG IR AFLP G S L VD V RP V RD TLHLEGKELEFKVI K L D QK R N N V VVSRR A V I E SENSAE R D 184

RS1 191 EIVQ NL E EG QV V E G V VKN I TDYGAFVDLGG I DGLLH V TDMAW R RVKHPSEI Q N I G QQVK V QIIRINQ 256
ES1 185 QLLE NL Q EG ME V K G I VKN L TDYGAFVDLGG V DGLLH I TDMAW K RVKHPSEI V N V G DEIT V KVLKFDR 250

RS1 257 E TH R I SLG M KQL ES DPW DG I GAK YP V G K K IS G T VTN I TDYG A FVE L E P G I EGL I H I S 314
ES1 251 E RT R V SLG L KQL GE DPW VA I AKR YP E G T K LT G R VTN L TDYG C FVE I E E G V EGL V H V S 308

RS1 315 EM S WT K KN V HP G K ILSTSQE V D V V VL EV D PTK RRISLGLKQ TLE NPWQ A FA HS H PA G TE VEG 376
ES1 309 EM D WT N KN I HP S K VVNVGDV V E V M VL DI D EER RRISLGLKQ CKA NPWQ Q FA ET H NK G DR VEG 370

RS1 377 EV K NK T E FG L FIGLDG DV DG M VHLSD LD WN RP GE QVIE E FN KGD VVR AVVL D VD VDK ERISLG I 441
ES1 371 KI K SI T D FG I FIGLDG GI DG L VHLSD IS WN VA GE EAVR E YK KGD EIA AVVL Q VD AER ERISLG V 435

RS1 442 KQL GR D AVGEAA A SGELR K N A V V SAE V IG V NDG G IE V R L VNHED V TAFI R RADL SRDR D E QR 501
ES1 436 KQL AE D PFNNWV A --LNK K G A I V TGK V TA V DAK G AT V E L --ADG V EGYL R ASEA SRDR V E DA 491

RS1 502 PERF SVG QT V D A RV T NFSK K D R K I Q LS IK A LEI A E EK E A V A QFGSSDSGRSLGDILGA A L K 562
ES1 492 TLVL SVG DE V E A KF T GVDR K N R A I S LS VR A KDE A D EK D A I A TVNKQEDANFSNNAMAE A F K 552

RS1 563 NRENN E•
ES1 553 AAKG- E•
```

Fig. 5: Comparison between the protein sequences of the ES1 and RS1 proteins. The amino acids are shown as one-letter code. Identical amino acids are boxed.

of S1 is dispensable both in vitro and in vivo (21,5) which might explain its divergency. The three internal repeats R1–R3 show the highest number of conserved amino acid residues in comparison with the ES1 protein. This part is the major RNA binding domain (20) and it is essential (5), which seems to be confirmed by the natural conservation of amino acids.

Another difference between both proteins is the absence of the two Cys residues in the RS1 protein. Based on chemical studies it had been suggested that they might play an important

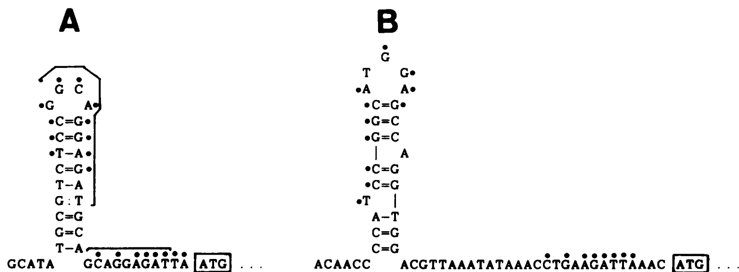

Fig. 6: Comparison of the RS1 (A) and ES1 (B) gene ribosome initiation sites. The dots show homologous regions between the two sequences. The lines in A point to repeats. The start codons are boxed.

role in the RNA unwinding property of protein S1 (25) as well as in the binding of fMet-tRNA to a 30S-MS2 RNA complex (26). On the other hand, ribosomes containing the same modified S1 could translate QβRNA as well as polyU (27,28). Whatever the reason for these discrepancies might be, the lack of the Cys-residues in RS1 suggests that they themselves are of minor importance. These results do not exclude the possibility that amino acids other than Cys, e.g. Ala as found here, influence essential properties of the S1 protein.

We also tried to find RNA-polymerase binding sites preceding the structural gene for RS1. A sequence comparison with the E. coli gene showed sequence homology of one promoter (P3, see 7), although parts of the -35 and -10 region are different. In this region we have mapped one of the RNA-polymerase binding sites. It is therefore likely that this is a promoter for the RS1 gene. We failed to detect another RNA polymerase binding site or sequence homology to the P1/P2-promoters which are located about 100 basepairs upstream from P3 in E. coli. Either these promoters are absent or they are highly specific to RNA-polymerase from R. meliloti. In this connection the apparent low expression of RS1 in E. coli should be noted, which was obvious in the maxicell analysis as well as by the fact that a fragment containing two RNA-polymerase binding sites did not complement the defect of the rpsA amber mutant. Apart from the possibility of proteolytic degradation, it could be that either transcription of the RS1 gene is inefficient in E. coli due to differences in the RNA polymerases or that translation is impaired due to different codon usages.

We further compared the initiation sites for ribosome binding between both the ES1 and the RS1 genes. The RS1 gene has two 9 basepairs consecutive stretches before the start codon which are identical and contain Shine-Dalgarno sequences. One sequence is located within a possible stem-loop structure (Fig. 6). When we compared the initiation sites from both the ES1 and the RS1 genes we detected sequence homology. This suggested a possible secondary structure with a second Shine-Dalgarno sequence preceding also the ES1 gene (Fig. 6). The two regions with Shine-Dalgarno

sequences might represent different ribosome binding sites important for the regulation of the rpsA gene.

ACKNOWLEDGEMENT
We thank Dr. H.G. Wittmann for support.

ABBREVIATIONS
ES1: Protein S1 from *E. coli*

RS1: Protein S1 from *R. melilotii*

Present addresses: [1]Department of Microbiology and Immunology, University of California, Berkeley, CA 94720, USA, [2]Umea University, Unit for Applied Cell and Molecular Biology, Umea S-90187, Sweden and [3]Serono, Taunusstrasse 24, D-1000 Berlin 41, FRG

*To whom correspondence should be addressed

REFERENCES
1. Van Dieijen, G., Van Knippenberg, P.H. and Van Duin, J. (1976) Eur. J. Biochem. 64, 511-518.
2. Schnier, J., Kimura, M., Foulaki, K., Subramanian, A.R., Isono, K. and Wittmann-Liebold, B. (1982) Proc. Natl. Acad. Sci. USA 79, 1008-1011.
3. Schnier, J. and Isono, K. (1982) Nucl. Acids Res. 10, 1857-1865.
4. Kimura, M., Foulaki, K., Subramanian, A.R. and Wittmann-Liebold, B. (1982) Eur. J. Biochem. 123, 37-53.
5. Schnier, J., Stöffler, G. and Nishi, K. (1986) J. Biol. Chem. 261, 11866-11871.
6. Subramanian, A.R. (1983) Progr. in Nucl. Acid Res. 28, 101-142.
7. Pedersen, S., Skouv, J., Kajitani, M. and Ishihama, A. (1984) Mol. Gen. Genet. 196, 135-140.
8. Christiansen, L. and Pedersen, S. (1981) Mol. Gen. Genet. 181, 548-551.
9. Kitakawa, M. and Isono, K. (1982) Mol. Gen. Genet. 185, 445-447.
10. Kamen, R., Kondo, M., Romer, W. and Weissmann, C. (1972) Eur. J. Biochem. 31, 44-51.
11. Schnier, J. and Faist, G. (1985) Mol. Gen. Genet. 200, 476-481.
12. Isono, K. and Isono, S. (1976) Proc. Natl. Acad. Sci. USA 73, 767-770.
13. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Cold Spring Harbor Laboratory, New York.
14. Chow, L.T., Kahmann, R. and Kamp, D. (1977) J. Mol. Biol. 113, 591-609.
15. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
16. Messing, J., Crea, R., Seeburg, P. (1981) Nucl. Acid Res. 9, 309-321.
17. Bagdasarian, M., Lurz, R., Rückert, B., Franklin, F.C.H., Bagdasarian, M.M., Frey, J. and Timmis, K.N. (1981) Gene 16, 237-247.
18. Sancar, A., Hack, A.M. and Rupp, W. (1979) J. Bacteriol. 137, 692-693.
19. Laemmli, U.K. (1970) Nature 227, 680-685.
20. Subramanian, A.R., Rienhardt, P., Kimura, M., Suryanarayana, T. (1981) Eur. J. Biochem. 119, 245-249.

21. Subramanian, A.R. and Mizushima, S. (1979) J. Biol. Chem. 254, 4309-4312.
22. Adhya, S. and Gottesmann, M. (1978) Ann. Rev. Biochem. 47, 967-996.
23. Stüber, D. and Bujard, H. (1981) Proc. Natl. Acad. Sci. USA 78, 167-171.
24. Yokota, T., Arai, K., Kaziro, Y. (1979) J. Biochem. (Tokyo) 86, 1725-1737.
25. Thomas, J.O., Kolb, A., Szer, W. (1978) J. Mol. Biol. 123, 163-176.
26. Chu, Y.G., Cantor, C.R. (1979) Nucl. Acid Res. 6, 2363-2379.
27. Cole, P.E., Sawchyn, I., Guerrier-Takada, C. (1982) J. Biol. Chem. 257, 12929-12934.
28. Yokota, T., Arai, K., Kaziro, Y. (1979) J. Biochem. (Tokyo) 86, 1739-1745.