J. Mol. Biol. (1987) **196**, 789–799

Positive Regulation of the *Escherichia coli* L-Rhamnose Operon is Mediated by the Products of Tandemly Repeated Regulatory Genes

J. F. Tobin and R. F. Schleif

Department of Biochemistry
Brandeis University
Waltham MA 02254, U.S.A.

(Received 10 February 1987, and in revised form 31 March 1987)

The *rhaC* gene, whose product is the positive activator of the genes required for L-rhamnose utilization, has been cloned along with the rhamnose structural genes. The *rhaC* sequence shows two partially overlapping reading frames, encoding two proteins of molecular weight 32,000 and 35,000 RhaS and RhaR. Both proteins show significant homology to AraC, the positive activator of the arabinose operon. S_1 mapping located transcriptional start points and showed that RhaR, and possibly RhaS, positively regulate transcription from the structural gene promoters as well as transcription from their own promoter. *In-vivo* dimethyl sulfate footprinting and DNase I footprinting indicate that the RhaR protein may bind to DNA elements upstream from its RNA polymerase binding site.

1. Introduction

The structural similarities between L-rhamnose, L-fucose and L-arabinose, as well as the similarities in their metabolic pathways (Power, 1967; Tecce & Digirolamo, 1956; Ajl, 1957; Chen & Lin, 1984; Schleif, 1985), raises the possibility that the regulatory mechanisms for each of these systems may also be similar. The regulation of the arabinose operon has been well studied (Schleif, 1985). It shows a complex pattern of positive and negative regulation mediated by an L-arabinose responsive regulatory protein. In this light we began a study of the rhamnose operon of *Escherichia coli* K12 and its regulatory protein(s).

Figure 1 shows the pathways for the metabolism of L-rhamnose, L-fucose and L-arabinose. The initial chemical steps in the catabolism of all three sugars are strikingly similar. The first step is an isomerization of the aldol forms of the sugars into the keto forms. In the second step the keto forms are phosphorylated. It is at the third step that the similarities between the L-arabinose pathway and the L-fucose and L-rhamnose pathways disappear. L-Ribulose 5-phosphate is converted by the epimerase to D-xylulose 5-phosphate, which enters the pentose phosphate shunt, while the phosphorylated keto derivatives of L-rhamnose and L-fucose are cleaved by their respective aldolases into L-lactaldehyde and dihydroxyacetone phosphate.

Using a genetic approach, Power (1967) began a study of the regulatory and structural genes of the rhamnose operon in *E. coli* K12. He identified four genes, which he named *rhaC*, *rhaB*, *rhaA* and *rhaD*, and determined their genetic order to be *CBAD*. Mutations in the *rhaB*, *rhaA* and *rhaD* genes identified them as the structural genes encoding the kinase, isomerase and aldolase, respectively. He identified the *rhaC* gene as a possible positive regulator of the three structural genes based on the behavior of *rhaC* mutations. Such mutations had a pleiotropic negative effect on the expression of the three structural gene products. He also observed two complementation groups within the *rhaC* cistron. These results were subsequently reproduced by Al-Zarban *et al.* (1984) in *Salmonella typhimurium* LT2.

In this paper we describe the cloning and structure of the rhamnose operon(s). We have identified four L-rhamnose inducible promoters. One of these is the promoter for the *rhaC* genes. The other three are apparently the promoters for the structural genes. We sequenced the *rhaC* gene and found that it encoded two proteins, RhaS and RhaR, both of which have extensive homology to the *araC* gene product, the positive activator of the arabinose operon. RhaR plays a major role in positively regulating transcription of *rhaSR* mRNA as well as the transcription originating from the three other L-rhamnose-inducible promoters, and

0022-2836/87/160789-11 $03.00/0

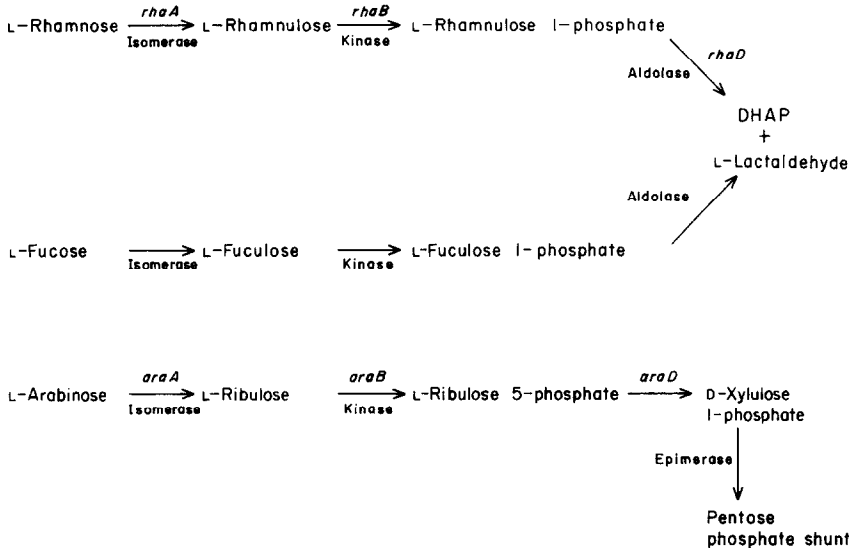

Figure 1. Pathway for the metabolism of L-rhamnose, L-fucose and L-arabinose dihydroxyacetone phosphate.

RhaS plays a minor role in such regulation. RhaR likely binds to a region of the DNA upstream from the *rhaSR* RNA polymerase binding site.

2. Materials and Methods

(a) *Media, plasmids, strains, chemicals and DNA manipulations*

General procedures and DNA manipulations were performed as outlined by Schleif & Wensink (1981) and Maniatis *et al.* (1982). The strains, phage and plasmids used for cloning and sequencing are described in Table 1.

Plasmid pJT5 was generated by digesting pLC5-5 at the *Pst*I sites. Approximately 5·0 kb† of DNA was removed from each end of the linear plasmid by digesting with nuclease *Bal*31. The DNA was made blunt-ended by filling out with the Klenow fragment of DNA polymerase I, digested with *Hin*dIII and separated by electrophoresis. DNA fragments between 7·0 and 8·0 kb were isolated. This procedure enriched for fragments which contained DNA to the left of the *Sma*I site and still retained the *Hin*dIII site. These fragments were then ligated into pUC12, which had been cut at the *Hin*dIII site and a blunt end site, and used to transform a RhaS$^-$/RhaR$^-$ (RhaC$^-$) strain.

The RhaS$^-$/RhaR$^+$ plasmid was made by cutting plasmid pJTC9 at the *Bgl*II site located at position 560, filling in the site with the Klenow fragment of DNA polymerase I and religating the blunt-ended molecule. This procedure inserts 4 bp at the restriction site, producing a termination codon at position 598 within the RhaS coding sequence. Generation of a termination codon within the RhaS coding sequence could produce a translational polarity effect on RhaR. To avoid this problem, an in-frame deletion removing 63 bp of the RhaS coding sequence was constructed. This in-frame deletion mutant behaved identically with the *Bgl*II insertion mutant in transcription experiments. The in-frame deletion was constructed by cutting pJTC9 with *Pvu*II and *Bgl*II. A partial fill-out reaction was performed in which 3 of the 4 nucleotides of the *Bgl*II site were filled out. The remaining single-stranded nucleotide was removed with nuclease ExoVII, and the blunt-ended molecule was religated. The RhaS$^+$/RhaR$^-$ plasmid was made by filling in a unique *Nhe*I site and religating. This operation introduced a 4 bp insertion, which produced a termination codon at position 1595 within the RhaR coding sequence. The RhaS$^-$/RhaR$^-$ plasmid was constructed by filling out the *Bgl*II and *Nhe*I sites in plasmid pJTC9. All constructs were confirmed by DNA sequencing.

(b) *Complementation tests*

Colonies of cells containing mutations in any of the rhamnose genes appear white on a MacConkey/L-rhamnose indicating plate. If a DNA clone introduced into one of these mutant cell lines has the ability to complement the mutation the resulting colonies will appear deep red on the indicating plates. Partial complementation is characterized by a pink colony.

(c) *Sequencing*

The 2·2 kb *Bam*HI–*Eco*RI fragment of pJT5 was cloned into the *Bam*HI and *Eco*RI sites of plasmids pGC1 and pGC2 to produce plasmids pJTC8 and pJTC9, respectively. An ordered set of deletions was obtained using the method described by Hong (1982). Briefly, the plasmids were partially digested with DNase I and the linear DNA isolated by gel electrophoresis and electroelution. pJTC8 and pJTC9 were then digested with *Bam*HI and *Eco*RI, respectively, followed by filling out with the Klenow fragment of DNA polymerase I and religation. The DNA was introduced into JM101 by transformation. The approximate size of the deletions was determined by digesting the plasmids with restriction enzymes and sizing the excised inserts by gel electrophoresis. For deletion plasmids derived from pJTC8, restriction enzymes *Eco*RI and *Hin*dIII were used, and for deletions derived from pJTC9, restriction enzymes *Bam*HI and *Hin*dIII were used. Single-stranded DNA template was obtained by infecting cell lines containing the deletion plasmids with phage M13rv1, a mutant helper phage that increases the titer of transducing particles (Levinson *et al.*, 1984). The single-stranded DNA template was isolated from the cells by the polyethylene

†Abbreviations used: kb, 10^3 bases or base-pairs; bp, base-pair(s).

Table 1
Bacterial strains, plasmids and phage

Strain, plasmid or phage	Genotype	Source or reference
SH322	F⁻: *ara(CBAD)+leu* Δ *(lac74) galK str*	
SH387	F⁻: *rhaC lacZ gal*	CGSC4833
SH388	F⁻: *thr leu proA2 his argE3 metE xthA lacY galK mtl xyl ara rhaB str*	CGSC5483
SH391	F⁻: *ara* Δ *(arg F-lac) flb ptsF recA str* Δ *(glnG* and/or *glnL) rhaD deoC1*	CGSC6596
ECL116	F⁻: Δ *(lacU 169) thi endA hsdR*	Backman *et al.* (1981)
ECL343	F⁻: Δ *(rha-pfk)* Δ *(lacU 169) thi endA hsdR*	Y.-M. Chen
ECL514	F⁻: Δ *(lacU169) thi endA hsdR rhaB101*	Y.-M. Chen
ECL515	F⁻: Δ *(lacU169) thi endA hsdR rhaA502*	Y.-M. Chen
ECL516	F⁻: Δ *(lacU169) thi endA hsdR rhaD701*	Y.-M. Chen
ECL517	F⁻: Δ *(lacU169) thi endA hsdR rhaC702*	Y.-M. Chen
ECL518	F⁻: Δ *(lacU169) thi endA hsdR rhaA703*	Y.-M. Chen
pGC1	*amp*	Meyers *et al.* (1985)
pGC2	*amp*	Meyers *et al.* (1985)
pUC12	*amp*	Yanisch-Perron *et al.* (1985)
M13rv1		Levinson *et al.* (1984)

glycol precipitation method described by Dente *et al.* (1983). The deletion candidates were sequenced by the method described by Sanger (1977, 1980) using the GC primer (Meyers *et al.*, 1985) kindly provided by E. Kantrowitz. The reliability of the sequence determination appeared to be about 98% for each strand.

(d) *Sequence homology searches*

Amino acid sequence homologies were determined using a program developed by Lipman & Pearson (1985).

The algorithm described by Mulligan *et al.* (1984) was used to search for homology between a DNA sequence and the consensus sequence for an *E. coli* promoter. First a sequence is sought with homology to the highly conserved −35 and −10 consensus promoter elements with the limitation that the distance between the 2 elements be between 15 and 21 bp. These putative promoters are then given a score based on the weighting scheme described by Mulligan *et al.* (1984). A score in the 70% range indicates a strong promoter while a score in the 40% range indicates a weak promoter.

(e) *Preparation of total cellular RNA*

Total cellular RNA was extracted from cells by the procedure described by Stoner & Schleif (1983*b*), except that the cells were grown in 100 ml of M10 medium in the presence of 0·5% (v/v) Casamino acids, with or without 0·2% (w/v) L-rhamnose, to a density of 3×10^8 cells/ml.

(f) *DNA probes*

All the DNA probes described below were end-labeled on a single 5′ end by phage T4 polynucleotide kinase. Two probes were used to examine the start site of transcription from the p_{sr} promoter: a 161 bp *Eco*RI–*Mlu*I fragment labeled at the *Mlu*I end, and a 1589 bp *Eco*RI–*Nhe*I fragment labeled at the *Nhe*I end. Both of these probes were derived from pJTC9. Transcription from promoters p_1 and p_2 was monitored with a 2·0 kb *Eco*RI–*Bam*HI fragment labeled at the *Bam*HI end, which was derived from plasmid pJT5. The probe for transcription originating from promoter p_3 was the 3·5 kb *Hin*dIII–*Bam*HI fragment labeled at the *Bam*HI end and derived from pJT5. A 461 bp *Hin*dIII–*Mlu*I probe labeled at the *Mlu*I end, derived from plasmid pJTC8, was used for DNase I footprinting experiments.

A probe used to examine the 3′ end of the p_{sr} mRNA was labeled by filling in a *Bgl*II site in the presence of [α-^{32}P]dCTP with the Klenow fragment of DNA polymerase I.

(g) *S_1 nuclease mapping*

DNA/RNA hybridizations were performed as described by Berk & Sharp (1977). Each reaction contained 25 μg of total cellular RNA and 50 ng of the appropriate end-labeled probe containing 50.000 cts/min. The S_1 nuclease mapping procedure was performed as described by Hung *et al.* (1982), except that 100 units of S_1 nuclease was used per reaction. The probes used to examine transcription from p_1, p_2 and p_3, as well as the 1589 bp *Eco*RI–*Nhe*I probe, were hybridized with the RNA at 55°C. The small probe used to examine transcription from the *rhaS* and *rhaR* genes was hybridized with the RNA at 45°C. After treatment with S_1 nuclease the samples were loaded on a 6% (w/v) polyacrylamide denaturing gel (Sanger & Coulson, 1978) or a 1% (w/v) agarose denaturing gel (Schleif & Wensink, 1981) depending on the size of the probe. An A+G sequencing reaction (Maxam & Gilbert, 1980) was electrophoresed on the 6% polyacrylamide denaturing gel as a size standard.

(h) In-vivo *dimethyl sulfate footprinting*

In-vivo dimethyl sulfate footprinting was performed using the scaled-down procedure described by Martin *et al.* (1986). The plasmid DNA isolated from the cells was digested with *Mlu*I and 5′ end-labeled by T4 polynucleotide kinase. The linear DNA was next digested with *Eco*RI and the excized 161 bp fragment was isolated by gel electrophoresis followed by electroelution. The DNA was cleaved at the methylated bases by 2 different reactions. Reaction with piperidine (Maxam & Gilbert, 1980) gave a G > A reaction, and a reaction with gentle acid followed by base (Maxam & Gilbert, 1977) gave an A > G reaction.

(i) *DNase I footprinting*

DNase I footprinting experiments were performed according to the procedure described by Galas & Schmitz (1978) with the following modifications. The DNA fragment was incubated for 10 min with the appropriate cell lysate in 50 μl of binding buffer (10 mM-Tris-acetate

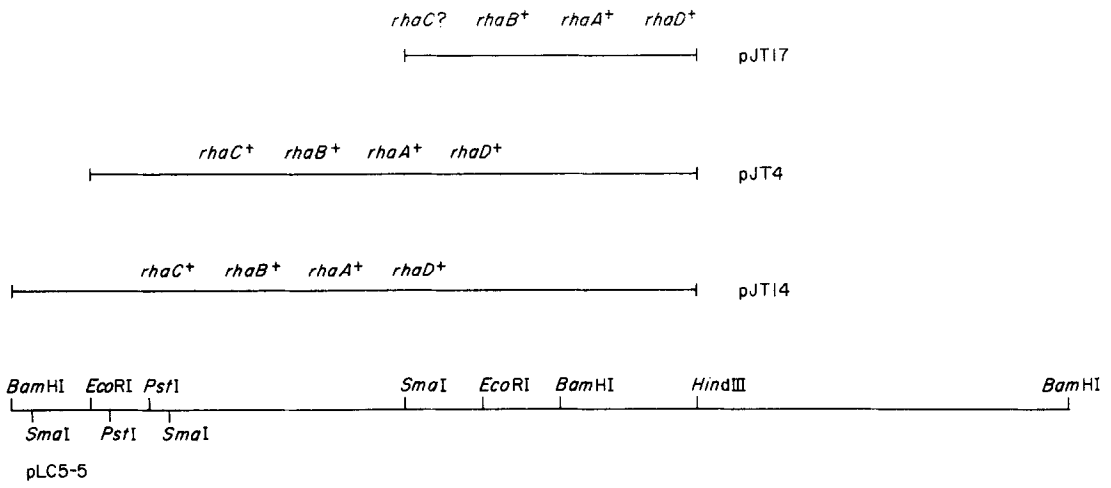

Figure 2. Restriction map of pLC5-5 and the phenotypes of the DNA fragments cloned into pUC12. *rhaC*$^+$ indicates that the clone was able to fully complement the RhaC$^-$ mutation, while *racC*? indicates only partial complementation. Plasmid pLC5-5 is 28 kb in length, pJT14 17·5 kb, pJT4 15·5 kb and pJT17 are 7·5 kb.

(pH 7·4), 50 mM-KCl, 50 mM-L-rhamnose, 1 mM-EDTA, 5% (v/v) glycerol, 50 μg bovine serum albumin/ml, 1 mM-dithioerythritol, 1 mM-$CaCl_2$, 2·5 mM-$MgCl_2$ and 0·05% (v/v) NP40) at 21 °C. Then DNase I was added to a final concentration of 0·08 μg/ml and incubated for 30 s at 21 °C. The reaction was stopped by the addition of an equal volume of 3 M-ammonium acetate, 10 mM-EDTA. The samples were precipitated with ethanol and electrophoresed on a 6% polyacrylamide denaturing gel (Sanger & Coulson, 1978).

(j) *Lysate preparation*

Plasmids containing the wild-type and various nonsense mutants of the *rhaS* and *rhaR* genes under the control of the *lac* promoter (Amann *et al.*, 1983) were transformed into strain JM101. The strains were grown in YT broth (Schleif & Wensink, 1981) to an A_{550} of 0·6, at which time isopropyl-β-D-thiogalactoside was added to a final concentration of 1 mM. After incubation for 2 h, 5 ml of cells were pelleted by centrifugation and resuspended in 0·5 ml of lysate buffer (0·1 M-potassium phosphate (pH 7·4) 50 mM-KCl, 1 mM-EDTA, 10% glycerol, 1 mM-dithioerythritol and 160 μg phenylmethylsulfonate/ml). The cells were lysed by sonication (3 × 6 s pulses with a microprobe) and the cell debris removed by centrifugation at 4 °C.

3. Results

(a) *Cloning of the rhamnose genes*

Clarke & Carbon (1976) reported two plasmids, pLC5-5 and pLC3-24, that were capable of complementing a Rha$^-$ strain to Rha$^+$. We obtained these two plasmids and assayed their ability to complement a RhaB$^-$ strain. One of the plasmids, pLC5-5, was capable of complementing the RhaB$^-$ strain as well as RhaS$^-$/RhaR$^-$ (RhaC$^-$), RhaA$^-$ and RhaD$^-$ strains.

Rha$^+$ subclones containing restriction fragments from pLC5-5 were used to locate the rhamnose genes on this plasmid. The ability of these clones to complement the RhaS$^-$/RhaR$^-$ (RhaC$^-$), RhaB$^-$, RhaA$^-$ and RhaD$^-$ mutations is shown in Figure 2. The restriction sites at the end of each fragment indicate the sites in pUC12 (Yanisch-Perron *et al.*, 1985) into which the fragments were cloned.

Although all the *rha* genes could be confined to a 7·5 kb *Sma*I–*Hin*dIII fragment, this fragment could only partially complement the RhaS$^-$/RhaR$^-$ (RhaC$^-$) mutation. Since there were no convenient restriction sites close to the left-hand side of the *Sma*I site that could be used to make a larger subclone, DNA was deleted from the large clone, pLC5-5, to generate a clone that could fully complement the RhaC$^-$ mutation. A clone, pJT5 (Fig. 3), was isolated that was capable of fully complementing the RhaS$^-$/RhaR$^-$ (RhaC$^-$) mutation as well as the mutations in the *rhaB*, *rhaA* and *rhaD* genes. When we sequenced the *rhaS* and *rhaR* (*rhaC*) gene we found that the *Sma*I site was located within the second open reading frame. Thus, the partial complementation occurred because the 7·5 kb *Eco*RI–*Sma*I fragment did not contain an intact *rhaR* gene.

Further subcloning showed that the 2·2 kb *Eco*RI–*Bam*HI fragment of pJT5, cloned into pUC12, complemented only a RhaC$^-$ mutation. Therefore, we conclude that this fragment contains the *rhaC* gene.

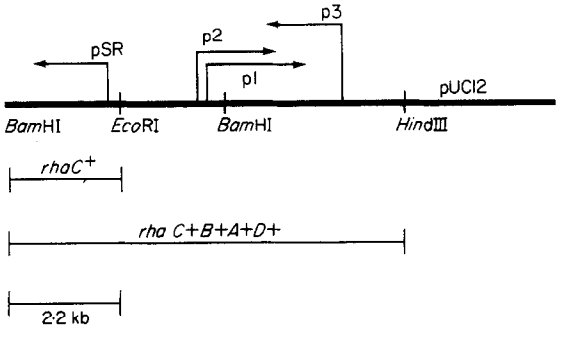

Figure 3. Plasmid pJT5. The arrows indicate the 5′ start sites of transcription for the 4 inducible mRNAs.

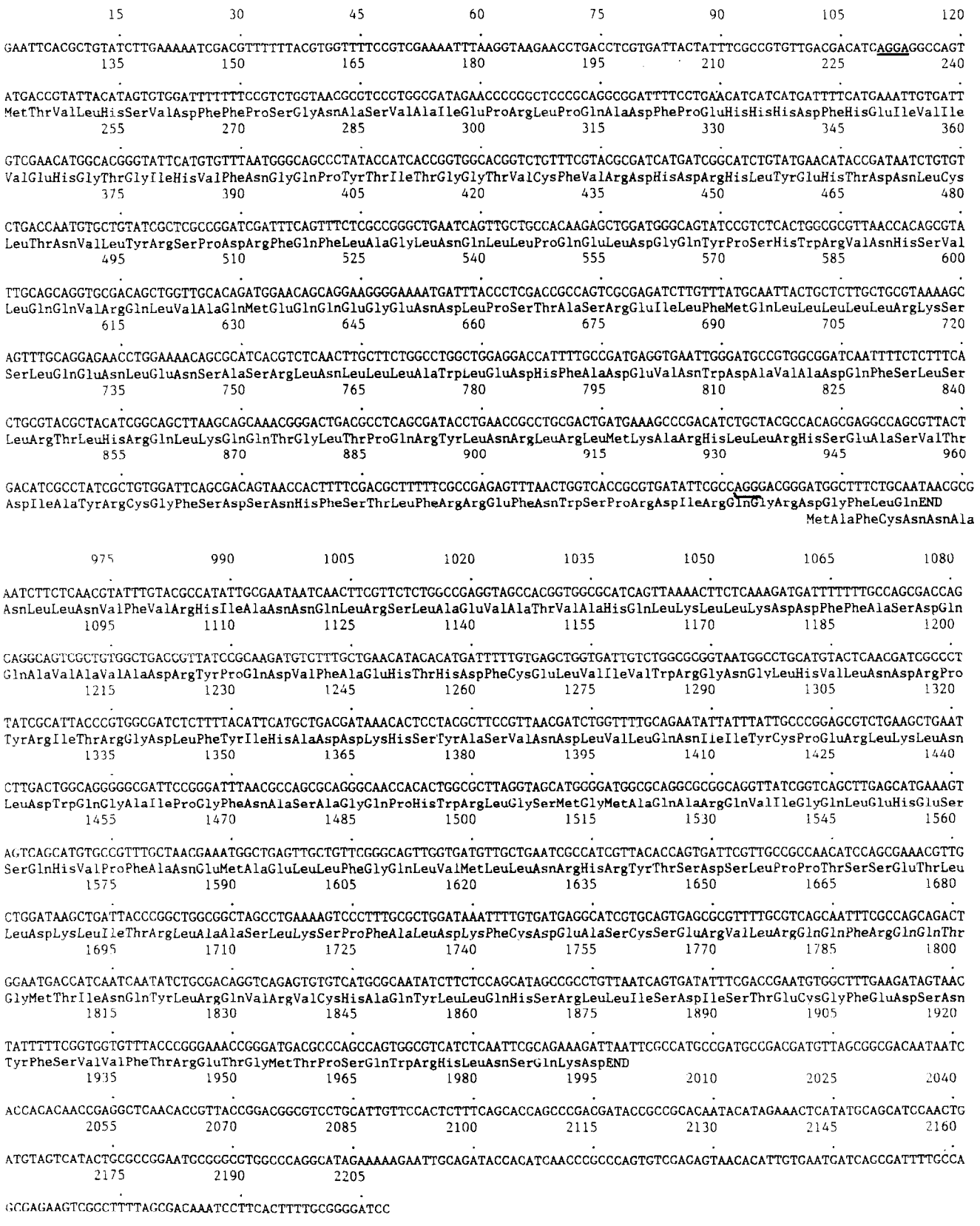

Figure 4. Sequence of the *rhaS/rhaR* (*rhaC*) genes. The Shine-Dalgarno sequences are underlined.

(b) *Coding region of the* rhaS *and* rhaR *(*rhaC*) genes*

The DNA sequence of the *rhaS* and *rhaR* (*rhaC*) genes revealed two partially overlapping reading frames. The first open reading frame extends from position 121 to 954 and the second open reading frame from position 941 to 1876 (Fig. 4). The first open reading frame begins with an ATG codon and ends with a TAA codon. It is preceded by a Shine–Dalgarno sequence, AGGA, located 7 bp before the initiation codon. This open reading frame can code

for a protein with 278 amino acids with a predicted molecular weight of 32,254. We have named this protein RhaS. The second open reading frame begins with an ATG codon and terminates with a TAA codon. The best Shine–Dalgarno sequence located 6 bp before the initiation codon is AGG. It can code for a protein of 312 amino acids, with a predicted molecular weight of 35,605. This protein has been named RhaR.

(c) *Homology between RhaS, RhaR and other proteins*

A protein homology program, developed by Lipman & Pearson (1985), was used to search for homology between the RhaS and RhaR proteins and an extensive bank of protein sequences. The algorithm screens sequences for similarity by looking for aligned identical amino acids and assigns a score based on identity as well as conservative amino acid replacements. The search revealed that RhaS and RhaR were highly homologous to each other over the entire length of RhaS. The 33 "extra" N-terminal amino acids of RhaR are undistinguished. There is no unambiguous signal sequence located within the extra 33 amino acids. RhaR and RhaS are also homologous to portions of the AraC protein from *E. coli* K12 (Fig. 5). No other protein sequences showed statistically significant homology to either RhaR or RhaS.

RhaS compared with the positive activator of the arabinose operon, AraC, had an aligned score of 148. The protein with the next closest homology had a score of only 57. Most of the homology between the two proteins occurred in the C terminus. RhaR was even more homologous to AraC having a score of 202; its next closest competitor having a score of only 54. Most of the homology between RhaR and AraC was also confined to the C terminus of the two proteins. When RhaS and RhaR were compared to each other they received a score of 486. The nucleic acid sequences encoding the three proteins reveal little homology.

(d) *S_1 nuclease mapping*

(i) *Transcriptional start site of the* rhaS *and* rhaR *genes*

S_1 analysis indicated that one mRNA encoded both the RhaS and RhaR proteins. Analysis of the 5′ end of the mRNA, using a probe complementary to the entire RhaS open reading frame and more than half of the RhaR open reading frame, revealed the same start site as a probe complementary to only a small part of the RhaS open reading frame. In addition analysis of the 3′ end revealed one termination site. Therefore, there can be only one major mRNA encoding both proteins.

Fine structure S_1 analysis showed that the polycistronic mRNA encoding both proteins possesses multiple transcription start sites within a 7 bp region, with most of the initiations beginning at a guanine, position +1, or at a thymine, position +2, but with appreciable initiations deriving from several other minor starts as indicated in Figure 6.

The region upstream from the transcription start site of the *rhaS* and *rhaR* gene possesses a poor RNA polymerase consensus binding sequence. The best −10 and −35 sequences are TActAT and TTaAgg, respectively (Fig. 4). Not only is the −35 sequence a poor fit to the −35 consensus sequence but its distance of 20 bp from the −10 sequence is significantly greater than the optimal spacing of 16 or 17. The promoter has a predicted strength of 41·1% (Mulligan *et al.*, 1984). We have named this

Figure 5. Amino acid homology between (a) AraC and RhaS, (b) AraC and RhaR and (c) RhaS and RhaR. A (:) indicates an amino acid identity while a (•) indicates a conservative amino acid change.

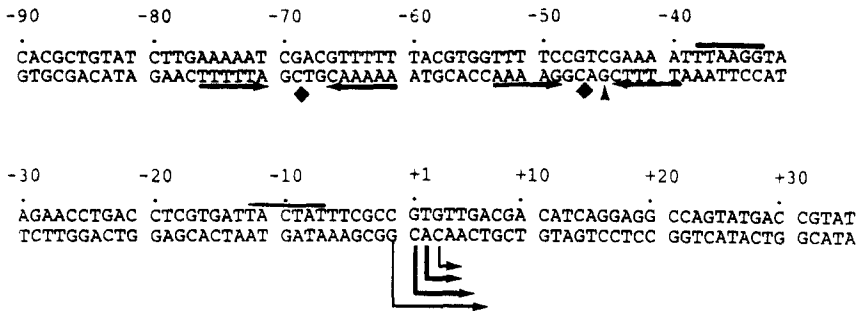

Figure 6. Nucleotide sequence of the upstream region of the *rhaS* and *rhaR* genes. The (⟶) indicate the major transcription start sites and (↳) the minor start sites. The (→←) indicate the location of the inverted repeat elements and (◆) indicates the center of symmetry within the elements. The overlined regions correspond to the consensus −10 and −35 regions of the RNA polymerase binding sites. The (▲) indicates the site of an enhancement in the *in-vivo* dimethyl sulfate footprinting studies.

promoter p_{sr}. Strong promoters that require no auxiliary factors receive high scores. The hybrid *tac16* promoter or the phage T7 *A1* promoter have scores of 69·1% and 68·1%, respectively, while the *araBAD* and *galP1* promoters that require positive-acting factors have scores of 45·7% and 43·8%, respectively. The DNA sequence upstream from the −35 region also contains two inverted repeat elements. These elements are often part of the binding site to which regulatory proteins bind. The first element, (5′–3′) TTTTCCGTGGAAAA, begins at position −53, and the second element, (5′–3′) AAAAATCGACGTTTTT, begins at position −76 (Fig. 6). We will present evidence later in the paper which shows that both of these elements are protected from DNase I digestion in $RhaR^+$ extracts, but not in $RhaR^-$ extracts.

(ii) *Other* L-*rhamnose-inducible promoters*

S_1 analysis has revealed three other rhamnose-inducible messages. The exact nucleotide at which each of these messages start or what gene products they encode has not been determined, although they likely encode the rhamnose structural proteins, since they lie within clones that complement various rhamnose structural gene mutations. Two of these messages, made from promoters p_1 and p_2 (Fig. 3), overlap each other and their start sites are separated by no more than 20 bp. The start sites are located approximately 500 bp from the *Bam*HI site in plasmid pJT5. A third inducible message, made from promoter p_3, starts 1500 bp from the *Hin*dIII site of pJT5 (Fig. 3). This mRNA is oriented in opposite transcriptional polarity with respect to the messages made from p_1 and p_2, but in the same polarity as the message made from the p_{sr} promoter.

(iii) *Kinetics of mRNA synthesis*

To examine the kinetics of message synthesis, total RNA was extracted from wild-type cells that had been grown in the presence of L-rhamnose for various times. At the first time point, five minutes after the addition of L-rhamnose to the cells, mRNA from p_{sr} is two-thirds of the fully induced level.

The kinetics of mRNA accumulation from promoters p_1, p_2 and p_3 are different from that of the p_{sr} promoter. The time span between the addition of L-rhamnose and the appearance of mRNA is longer for these three promoters. Cells grown in the absence of L-rhamnose show no detectable transcription from p_1, p_2 and p_3. Five minutes after the addition of L-rhamnose transcription is less than a tenth of the fully induced level. It is not until the second time point, 30 minutes later, that the fully induced level of mRNA can be detected from these three promoters.

(iv) *Effects of mutations in the* rhaS *and* rhaR *genes on* L-*rhamnose-induced transcription*

To determine if RhaS and RhaR are positive regulators of L-rhamnose-induced gene transcription, the effects of mutations in *rhaS* and *rhaR* on transcription from the four L-rhamnose-inducible promoters was examined. Cells containing the mutations were grown in the absence or presence of L-rhamnose. Total cellular RNA was extracted and used in S_1 mapping experiments to examine transcription from each of the promoters. Figure 7(a) shows the results for the p_{sr} promoter. In a cell line containing the $RhaS^+/RhaR^+$ plasmid, grown in the presence of L-rhamnose, the start sites and the amount of transcription from the p_{sr} promoter are the same as in a cell line containing a single chromosomal copy of *rhaS* and *rhaR*. In a cell line containing a mutation in RhaS there is a twofold decrease in transcription from p_{sr}. A mutation in RhaR or a double mutation abolishes transcription. In the absence of L-rhamnose there is no transcription from the p_{sr} promoter in any of the cell lines.

The preceding results indicate that RhaR in the presence of L-rhamnose strongly stimulates p_{sr}. RhaS may also stimulate, but its twofold effect is so small as to keep this question open. We did not see the twofold stimulation effected by RhaS in $RhaR^-$ situations, indicating that the activity of RhaS is dependent on RhaR or that the level of transcription is so low in $RhaR^-$ cells that a twofold stimulation is undetectable.

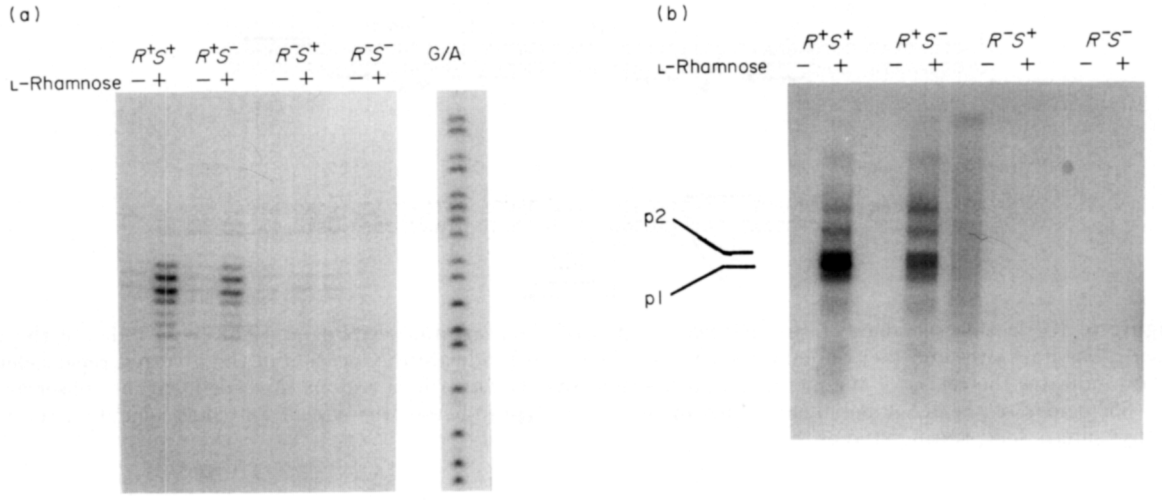

Figure 7. S_1 analysis of (a) the p_{sr} promoter and (b) the p_1 and p_2 promoters in a cell line (ECL343) containing plasmids with the wild-type, *rhaS* and *rhaR* genes, or plasmids with mutations in the *rhaS* and/or *rhaR* genes. A (+) indicates that cells were grown in the presence, a (−) in the absence, of 0·2% (w/v) L-rhamnose.

The results for the other three L-rhamnose-inducible promoters are similar to that obtained with the p_{sr} promoter. Figure 7(b) shows that in cells containing the $RhaS^+/RhaR^+$ plasmid transcription from p_1 and p_2 is only observed in the presence of L-rhamnose. The approximate start sites and levels of transcription are the same as in a cell line containing a single chromosomal copy of *rhaS* and *rhaR*. A mutation in RhaS decreases transcription from p_1 by twofold but has no effect on transcription originating from p_2. The RhaR mutant and the double mutant completely abolish transcription from both promoters. The results for p_3 (data not shown) are identical with the results obtained from p_1. We conclude that RhaR positively regulates transcription from the promoters p_1, p_2 and p_3, and that RhaS may regulate p_1 and p_3 weakly.

(e) *RhaR, but not RhaS, binds to the inverted repeats upstream from its gene's RNA polymerase binding site*

It is possible that RhaS and RhaR positively regulate their gene expression by binding to the two upstream inverted repeat elements in the p_{sr} promoter (Fig. 6). To test this possibility we used the *in-vivo* dimethyl sulfate footprinting technique (Martin *et al.*, 1986) to examine the occupancy of these two elements in cell lines containing mutations in *rhaS* and *rhaR*. Figure 8 shows that, compared to $RhaS^-/RhaR^-$ cells, methylation of guanine −45 increased twofold upon the addition of L-rhamnose to cells containing the $RhaS^+/RhaR^+$ plasmid. This hypermethylation is dependent on RhaR but independent of RhaS. The presence of the RhaS protein did not dramatically effect the methylation rate of any of the guanines in either region.

Figure 9 shows the result of an *in-vitro* DNase I footprinting experiment on the upstream region of

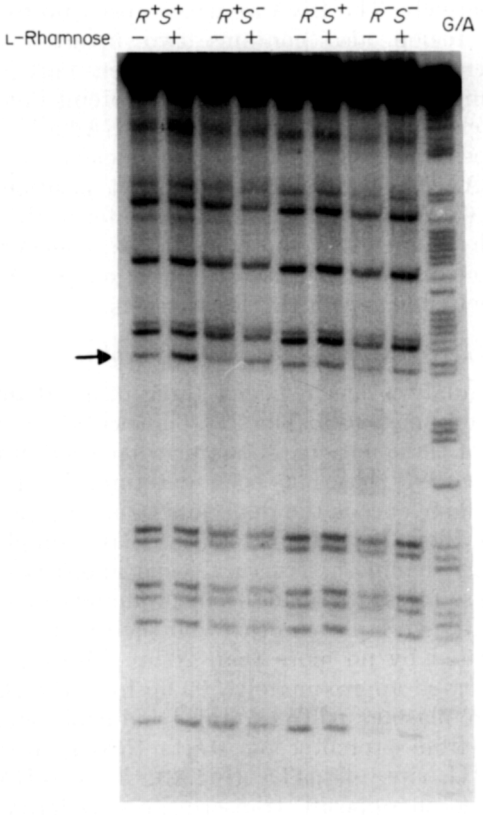

Figure 8. *In-vivo* dimethyl sulfate footprinting of the upstream region of the *rhaS* and *rhaR* genes. A (+) indicates that cells were grown in the presence, a (−) in the absence of 0·2% L-rhamnose. The arrow indicates the location of enhancement. The gel was scanned using a Joyce–Loebl densitometer and the areas under the peaks determined. The area of the peak corresponding to the enhancement was normalized to account for differences in the amount of material loaded per lane. The normalization was performed by dividing the area of the enhanced peak by the area of a peak well outside the region of interest that was unaffected by the *rhaS* and *rhaR* mutations.

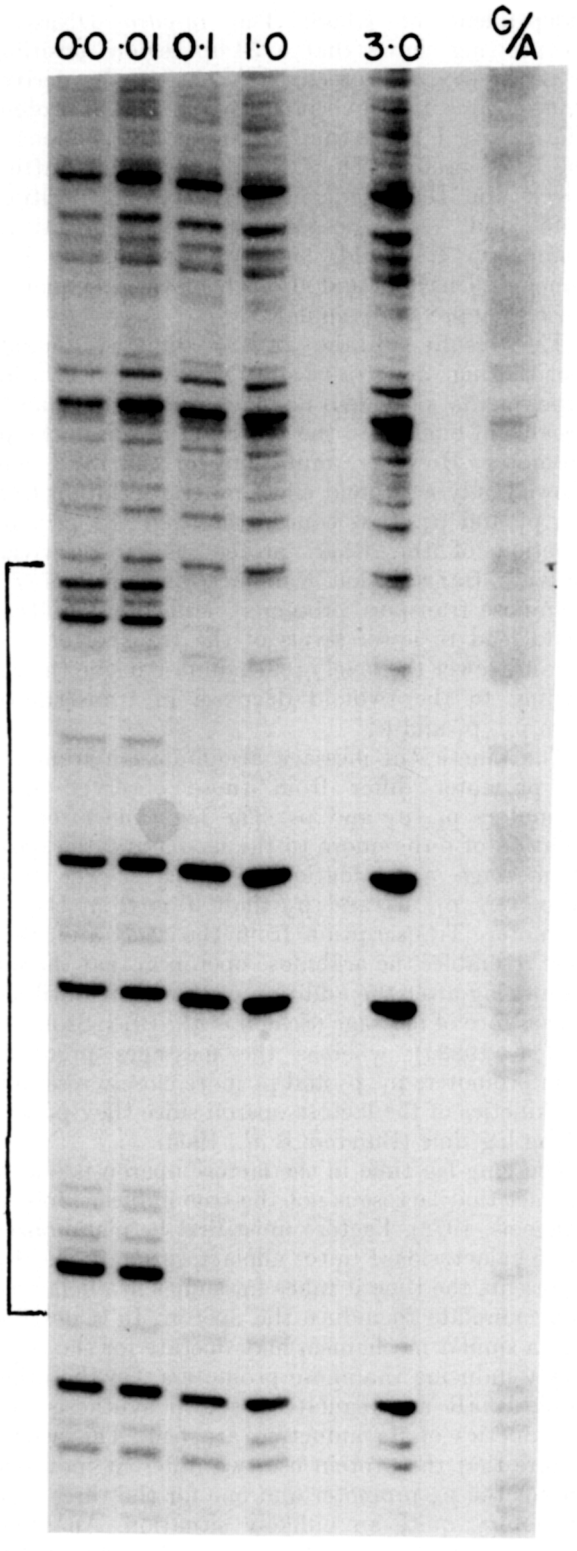

Figure 9. *In-vitro* DNase I footprinting of the upstream region of the *rhaS* and *rhaR* genes. The numbers at the top of each lane refer to the amount of cell lysate (in μl) added to each 50 μl DNase I reaction. The bracket shows the extent of the protected region.

the p_{sr} promoter. The experiment was performed in the presence of L-rhamnose with a cell extract derived from a strain containing a plasmid that overproduces both the RhaS and RhaR proteins. A 54 bp region from position −34 to −87 is protected. This footprint covers the region containing both inverted repeat elements. The footprint is also generated by cell extracts derived from strains containing an $RhaS^-/RhaR^+$ overproducing plasmid but is not generated by cell extracts from strains containing either $RhaS^+/RhaR^-$ or $RhaS^-/RhaR^-$ overproducing plasmids.

4. Discussion

The chemical similarities between L-rhamnose, L-fucose and L-arabinose as well as the similarities in their catabolism prompted us to begin a study of the rhamnose operon and its regulatory proteins. Despite the rather striking homologies between the amino acid sequences of AraC, RhaS and RhaR, the two operons appear to be quite different.

Cell lines containing mutations in several of the structural genes as well as the putative regulatory gene (Power, 1967) were used to clone the rhamnose operon. The regulatory genes, *rhaS* and *rhaR*, were located to a 2·2 kb fragment (Fig. 3), which defines by itself a separate cistron. The *rhaB*, *rhaA* and *rhaD* genes are also contained within the 7·7 kb fragment in plasmid pJT5 (Fig. 3) but their exact location has not been determined.

Analysis of nuclease S_1-digested complexes between known labeled DNA probes and RNA extracted from cells was used to determine the location of L-rhamnose-inducible transcripts within the fragment known to contain the rhamnose structural and regulatory genes. Four inducible mRNAs were identified. One of these, p_{sr}, encodes the *rhaS* and *rhaR* gene products and is divergently oriented from the mRNAs produced from promoters p_1 and p_2 (Fig. 3), which themselves likely encode the structural gene products. The fourth message, produced from promoter p_3, initiates approximately 2·5 kb downstream from the messages produced from promoters p_1 and p_2 and is oriented in opposite transcriptional polarity with respect to them.

Although the 3′ ends of the messages originating from promoters p_1 and p_2 has not been mapped, it is likely that the messages encode the same product(s). The locations of their 5′ ends differ by no more than 20 nucleotides. Tandemly repeated promoters are common in *E. coli* (McClure, 1985). In the galactose operon the tandemly repeated promoters p_1 and p_2 are subject to differential control. p_1 is dependent on cAMP receptor protein (CRP) and cAMP for full activity while p_2 is CRP independent (Musso *et al.*, 1977). Mutations in RhaS decrease transcription from p_1 by twofold but do not affect transcription from p_2. Thus, the rhamnose-inducible promoters, p_1 and p_2, may also be slightly differentially regulated. It is believed that differential control of the galactose operon allows a higher basal level of expression independent of the cAMP concentration inside the cell (deCrombrugghe & Pastan, 1978). We do not

yet see a need for differential control in the *rha* operon.

Sequence analysis of the *rhaS* and *rhaR* (*rhaC*) genes revealed two partially overlapping open reading frames that encode two proteins, RhaS and RhaR, of molecular weight 32,254 and 35,605, respectively. Both of these proteins are made from a single polycistronic message. Despite the fact that Power (1967) found two separate complementation groups among uninducible mutations within the *rhaC* (*rhaS* and *rhaR*) cistron, our finding of two proteins encoded by the *rhaC* gene does not seem to explain his results. Mutations in RhaR make cells unable to grow on L-rhamnose and such strains appear strongly Rha^- on MacConkey indicating plates. Mutations in RhaS only have a twofold effect on transcription, and a $Rha^-/RhaR^+$ strain has essentially the same phenotype on an indicating plate as a wild-type strain. Power's (1967) procedure for isolating *rhaC* (*rhaS* and *rhaR*) mutations would have been incapable of identifying mutations in RhaS. We note, however, that our complementation studies were performed with multicopy plasmids, whereas Power used single copy episomes. Therefore, at this time the source of the two complementation groups remains unclear.

Significant amino acid sequence homology exists between RhaS, RhaR and AraC. However, there is little homology between the three genes at the nucleotide level. Several mutations have been isolated in AraC that decrease its ability to bind to its corresponding DNA binding sites (A. Brunelle, personal communication; Cass & Wilcox, 1986). These mutations are clustered in the C terminus of the protein. Pabo & Sauer (1984) have shown that some DNA binding proteins have similar amino acid sequences in their DNA-contacting domains. Such a sequence can be found in the AraC protein and the mutations affecting binding lie within it (A. Brunelle, personal communication; Cass & Wilcox, 1986). It is interesting to note that most of the homology between AraC, RhaS and RhaR is found in the C terminus, perhaps the two *rha* proteins bind their target sites using a similar sequence located in their C termini.

The *rhaS* and *rhaR* genes' RNA polymerase binding site is a very poor one as indicated by its low homology to consensus promoters (Hawley & McClure, 1983). Its −10 region contains four of six consensus bases while the −35 region only contains three of six consensus bases. In addition, the spacing (20 bp) is more than the optimal spacing of 16 or 17. Promoters that contain an RNA polymerase binding site with a poor −35 sequence are often positively regulated by dissociable factors; therefore it is not surprising that the *rhaS* and *rhaR* genes are under autogenous positive control.

In-vivo dimethyl sulfate footprinting shows that the presence of RhaR within a cell increases the methylation rate of a guanine residue, position −45 (Fig. 6), upstream from the *rhaS* and *rhaR* genes' RNA polymerase binding site. This enhancement is independent of RhaS. The *in-vitro* DNase I footprinting shows that a 54 bp region, positions −34 to −87, is protected by cell extracts derived from strains that overproduce the RhaR protein. The DNase I protection pattern is independent of the RhaS protein. The DNase I protection pattern covers the DNA sequences centered at positions −46 and −68 possessing twofold rotational symmetry. It would, however, be amazing if a dimer of RhaR bound to each of these sequences since they are so dissimilar.

The results of the *in-vivo* dimethyl sulfate footprinting and *in-vitro* DNase I footprinting experiments provided no evidence that RhaS is capable of binding to the upstream region of the p_{sr} promoter. However, mutations in the *rhaS* gene showed only a twofold effect on transcription from p_{sr}, p_1 and p_3. It is conceivable that the primary function of the RhaS protein is to positively activate transcription from a gene(s) encoding a rhamnose transport protein(s). Mutations in RhaS could lead to lower levels of the transporter, and thereby lower the level of L-rhamnose inside the cell leading to the twofold decrease in transcription from p_{sr}, p_1 and p_3.

The kinetics of message accumulation from the p_{sr} promoter differ from those observed from promoters p_1, p_2 and p_3. The lag time after the addition of L-rhamnose to the media and the onset of message accumulation is much slower from promoters p_1, p_2 and p_3 than it is from the p_{sr} promoter. Transcription from the *rhaS* and *rhaR* gene resembles the arabinose operon in that there is a short lag after the addition of the sugar until the appearance of message (Schleif *et al.*, 1973; Stoner & Schleif, 1983*a*); whereas the messages produced from promoters p_1, p_2 and p_3 more closely resemble the kinetics of the lactose operon since they possess a long lag time (Burstein *et al.*, 1965).

The long lag time in the lactose operon is due to the fact that lactose is not the true inducer (Jobe & Bougeois, 1972). Lactose must first be metabolized by β-galactosidase into allolactose and the lag represents the time it takes for sufficient allolactose to accumulate to induce the operon. It is possible that a similar mechanism may operate for the three slowly inducing rhamnose promoters. On the other hand, RhaR autoregulates its own synthesis and the kinetics of its induction are fast. This would require that the protein use two different inducers, one for the p_{sr} promoter and one for the three slow promoters, quite an unlikely situation. An alternative explanation for the different kinetics is that the protein activates transcription once it binds to the DNA, and that the affinity of RhaR for its DNA binding sites is lower at promoters p_1, p_2 and p_3 than it is at the p_{sr} promoter.

We thank S. Hahn for the strains used in the initial part of this work as well as for help with the initial cloning experiments. We also thank Y.-M. Chen and E. C. C. Linn for their generous gift of strains.

References

Al-Zarban, S., Hefferernan, L., Nishitani, J., Ransone, L. & Wilcox, G. (1984). *J. Bacteriol.* **158**, 603–608.

Amann, E., Brosius, J. & Ptashne, M. (1983). *Gene*, **25**, 167–178.

Backman, K., Chen, Y.-M. & Magasanik, B. (1981). *Proc. Nat. Acad. Sci., U.S.A.* **78**, 3743–3747.

Berk, A. J. & Sharp, P. A. (1977). *Cell*, **12**, 721–732.

Burstein, C., Cohn, M., Kepes, A. & Monod, J. (1965). *Biochim. Biophys. Acta*, **95**, 634–639.

Cass, L. & Wilcox, G. (1986). *J. Bacteriol.* **166**, 892–900.

Chen, Y.-M. & Lin, E. C. C. (1984). *J. Bacteriol.* **157**, 828–832.

Chiu, T. H. & Feingold, D. S. (1965). *Biochem. Biophys. Res. Commun.* **19**, 511–516.

Clarke, L. & Carbon, L. (1976). *Cell*, **9**, 91–99.

deCrombrugghe, B. & Pastan, I. (1978). In *The Operon*, pp. 303–324. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Dente, L., Cesareni, G. & Cortese, R. (1983). *Nucl. Acids Res.* **11**, 1645–1655.

Galas, D. J. & Schmitz, A. (1978). *Nucl. Acids Res.* **9**, 3157–3170.

Hawley, D. K. & McClure, W. R. (1983). *Nucl. Acids Res.* **11**, 2237–2255.

Hong, G. F. (1982). *J. Mol. Biol.* **158**, 539–549.

Hung, M.-C., Barnett, T., Woolford, C. & Wensink, P. C. (1982). *J. Mol. Biol.* **154**, 581–602.

Jobe, A. & Bourgeois, S. (1972). *J. Mol. Biol.* **69**, 397–408.

Levinson, A., Silver, D. & Seed, B. (1984). *J. Mol. Appl. Genet.* **2**, 507–517.

Lipman, D. J. & Pearson, W. R. (1985). *Science*, **227**, 1435–1441.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Martin, K., Huo, L. & Schleif, R. (1986). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 3654–3658.

Maxam, A. & Gilbert, W. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 560–564.

Maxam, A. & Gilbert, W. (1980). *Methods Enzymol.*, **65**, 499–560.

McClure, W. R. (1985). *Annu. Rev. Biochem.* **54**, 171–204.

Meyers, R. M., Lerman, L. S. & Maniatis, T. (1985). *Science*, **229**, 242–247.

Mulligan, M. E., Hawley, D. K., Entriken, R. & McClure, W. R. (1984). *Nucl. Acids Res.* **12**, 789–800.

Musso, R. E., DiLauro, R., Adhya, S. & deCrombrugghe, B. (1977). *Cell*, **12**, 847–854.

Pabo, C. O. & Sauer, R. T. (1984). *Annu. Rev. Biochem.* **53**, 293–321.

Power, J. (1967). *Genetics*, **55**, 557–568.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 5463–5467.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980). *J. Mol. Biol.* **143**, 161–178.

Sanger, F. & Coulson, A. R. (1978). *FEBS Letters*, **87**, 107–110.

Schleif, R. F. (1985). In *The Molecular Biology of the Bacterial Cell*, pp. 194–201, Jones and Bartlett, Boston.

Schleif, R. F. & Wensink, P. (1981). *Practical Methods in Molecular Biology*, Springer-Verlag, New York.

Schleif, R. F., Hess, W., Finkelstein, S. & Ellis, D. (1973). *J. Bacteriol.* **115**, 9–14.

Stoner, C. M. & Schleif, R. F. (1983*a*). *J. Mol. Biol.* **170**, 1049–1053.

Stoner, C. M. & Schleif, R. F. (1983*b*). *J. Mol. Biol.* **171**, 369–381.

Tecce, G. & Digirolamo, M. (1956). *Gen. Microbiol.* **1**, 286–295.

Wilson, D. M. & Ajl, S. (1957). *J. Bacteriol.* **73**, 410–414.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). *Gene*, **33**, 103–119.

Edited by S. Brenner