Characterization of the *hemA*–*prs* region of the *Escherichia coli* and *Salmonella typhimurium* chromosomes: identification of two open reading frames and implications for *prs* expression

DAVID A. POST,^{1,2†} BJARNE HOVE-JENSEN² and ROBERT L. SWITZER^{1*}

¹Department of Biochemistry, University of Illinois, Urbana, Illinois 61801, USA

² University Institute of Biological Chemistry B, DK-1307 Copenhagen K, Denmark

(Received 24 June 1992; revised 26 October 1992; accepted 29 October 1992)

The *prs* gene, encoding phosphoribosylpyrophosphate synthetase, is preceded by a leader, which is 302 bp long in *Escherichia coli* and 417 bp in *Salmonella typhimurium*. A potential open reading frame (ORF) extends across the *prs* promoter and into the leader. The region between the *prs* coding region and an upstream gene (*hemA*) in *E. coli* and *S. typhimurium* was cloned, sequenced and shown to encode two ORFs of unknown function. ORF 1 encodes a 23 kDa protein and ORF 2 a 31 kDa protein, as observed by denaturing PAGE of extracts of cells bearing plasmids encoding the ORFs. Both ORFs are transcribed in the same direction as the *prs* gene with ORF 2 extending into the *prs* leader. Northern blot analysis showed that the *prs* message in *E. coli* was on 1.3 and 2.7 kb transcripts. The shorter transcript encoded the *prs* gene only, while the longer transcript also encoded the two ORFs. Thus, the *prs* gene is transcribed from two promoters, the first promoter (P_1) originating upstream of ORF 1, and expressing the *prs* gene in a tricistronic operon and a second promoter (P_2), located within the ORF 2 coding frame, which transcribes the *prs* gene only. The transcripts encoding *prs* only were 20 times as abundant as the tricistronic transcripts under all conditions examined. This was the case whether cells containing plasmid-encoded or only chromosomally encoded copies of the *hemA*–*prs* region were probed for these transcripts. Derepression of the *prs* gene upon pyrimidine starvation was shown to be due to an increase in the amount of message originating from the promoter P_2 .

Introduction

The gene encoding phosphoribosylpyrophosphate (PRPP) synthetase (*prs*) has been cloned and sequenced from both *Escherichia coli* (Hove-Jensen, 1985; Hove-Jensen *et al.*, 1986) and *Salmonella typhimurium* (Bower *et al.*, 1988); *prs* transcripts have a long 5' leader, 302 bp in *E. coli* (Hove-Jensen *et al.*, 1986) and 417 bp in *S. typhimurium* (Bower *et al.*, 1988). Roles for this leader have been suggested, but no clear function has been established. It was noted previously that a potential ORF originating 5' to the *prs* promoter extended through the *prs* promoter and terminated within the leader (Bower *et*

al., 1988). To study the possible effect of an upstream reading frame on the regulation of the *prs* gene, DNA upstream of the *prs* gene was cloned and sequenced, and transcripts encoded by this region were analysed. The results indicate that two ORFs upstream to *prs* are present, but that their transcription does not play a significant role in *prs* expression.

Methods

Growth conditions. Minimal medium was either AB (Clark & Maaløe, 1967) or M9 (Miller, 1972) supplemented with 0.1 % or 0.3 % glucose, respectively. Amino acid supplements were added at 50 µg ml⁻¹, thiamine at 0.5 µg ml⁻¹, uracil at 20 µg ml⁻¹, and for pyrimidine starvation, orotic acid (freshly dissolved at pH 12) was added at 100 µg ml⁻¹. Luria broth (Miller, 1972) was used for rich medium and was supplemented with 0.1 % glucose. Ampicillin was added to 50 µg ml⁻¹ in rich medium and 25 µg ml⁻¹ in minimal medium.

Cloning of the *hemA*–*prs* DNA. The DNA encompassing the *prs* and *hemA* genes from *S. typhimurium* was cloned by a plasmid rescue technique (Post & Switzer, 1991). A 1.8 kb *Pst*I/*Hind*III fragment directly upstream of the *prs* gene from plasmid pPS3, previously

* Author for correspondence. Tel. (217) 333 3940; fax (217) 244 5858.

† Present address: Abbott Laboratories, Strain Development Department, 1401 Sheridan Road, North Chicago, IL 60064, USA.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession numbers M77236 and M77237.

E. coli		S. typhimurium		hema		CATGTCTCGCGGAAATATACCAACGTTGATAGCGTTAGTCTGCTTGCATCATACATACAGGATCGGTAAAGATCAATAAAAGAG		84
						GTAC- G T		83
						-35 (P1) -10 (P1)		
AGCGCCCCCTTTTGGAGTAATTGCCGAAAGCCGTTAG--ATTCTGGCAATTAAGACAACTTGAACATAGACGATAGCGGACGGTAAACGCTAGCATTAAGGGTTATAACTGCAAC-----	A	G ---	GC CGCTA	TGATG T	AG	GT ACCA	CGG A C T CGATAG	196
CAAAAAATTGCCGAATCGCGCGCGAACGCTTTTACAGCCTGCAGGTTTCAGTTTTCAGGCCTGATAAGCGTAACGCCATCAGGCAATTATTGCACC							GTATCTCAAGGACTTGTTCATCACT A T C	220
ORF1-								320
ATGCCCCCTGCCCGATTTCCTGCTTATCCGCTGCTACCGCTGGCTGCTCTTGTGCTCACTGCTGTTCCGTTACCAACGCCCAAAGGTCCTGGCAAAAGCCCGGATTCCGCCAATGGCGT	A	C G T T A G T A AAGC G T C	A GC C TGG AT C G G	C T G C				340
MetProLeuProAspPheArgLeuIleArgLeuLeuProLeuAlaAlaLeuValLeuThrAlaCysSerValThrThrProLysGlyProGlyLysSerProAspSerProGlnTrpArg	Thr		Ser		ThrLeuProGlyHis			440
CAGCATCAGCAAGACGTGCGCAATCTTAATCAGTATCAGACTCGCGCGCGCTTCGCTTATATTCTGACCAACAAAAGTGTAACGCCCGCTTTTCTGGCAGCAAAACCGCCAGGATCGC	G TC G A C G C T C C A A T G T C C G T C A G A							460
GlnHisGlnGlnAspValArgAsnLeuAsnGlnTyrGlnThrArgGlyAlaPheAlaTyrIleSerAspGlnGlnLysValTyrAlaArgPhePheTrpGlnGlnThrGlyGlnAspArg	Glu His				Asp			560
TACCGTCTGCTGCTCACTAACCCATTGGGCAGCACGGAACTGGAGCTGAATGTCTCAACCGGGTAACGTGACAGTATAGTCGACAATAAAGGTACAGCGTTATACCGCCGATGACGCCAAGAG	T C C T GC C G T C C G C T G G T C C							580
TyrArgLeuLeuLeuThrAsnProLeuGlySerThrGluLeuGluLeuAsnAlaGlnProGlyAsnValGlnLeuValAspAsnLysGlyGlnArgTyrThrAlaAspAspAlaGluGlu								680
ATGATTGGCAAATTGACCGGAATGCCAATTCGGCTCAACAGCTTGCGCCAGTGGATTTTAGGTTTACCGGGTGATGCAACCGACTACAAACTGGACGACCAGTACCGCTGAGCGAAATT	C C C G G T A C G CC C G T C C						T G G	700
MetIleGlyLysLeuThrGlyMetProIleProLeuAsnSerLeuArgGlnTrpIleLeuGlyLeuProGlyAspAlaThrAspTyrLysLeuAspAspGlnTyrArgLeuSerGluIle							Val	800
ACCTACAGCCAGAATGGCAAACTGGAAGGTGTTTATGTTGGTTATGACACCAAAACGCAACCTGCCGATGCCCAATATGGAACCTACCGACGGTGTCAACGCATCAAGTTAAAA	A TC G T A G C C G G C T T A G T T A C A G C A A C C G C A T T G G A A C T A C C G A C G G T G T C A A C G C A T C A A G T T A A A A							820
ThrTyrSerGlnAsnGlyLysAsnTrpLysValValTyrGlyGlyTyrAspThrLysThrGlnProAlaMetProAlaAsnMetGluLeuThrAspGlyGlyGlnArgIleLysLeuLys	Asn Arg Asp				Ser		Ser	920
ATGGATAACTGGATGAATAATGCGGACACAGTGGCCCTCTCCGGCAAACTTAATCTGTTTATACATTACCGGTACAGCGTTCGGGATGTTTACCACACGCTGCAAAACGCTGTTT	T --- AT C T T T A T C A A A C C G							939
MetAspAsnTrpIleValLys MetArgThrGlnTrpProSerProAlaLysLeuAsnLeuPheLeuTyrIleThrGlyGlnArgAlaAspGlyTyrHisThrLeuGlnThrLeuPhe	Met His							1036
CAGTTTCTTGATTACGGCGACACCATCAGCATTGAGCTTCGTGACGATGGGGATATTTCGTCTGTTAACGCCCGTTGAAGCGGTGGAACATGAAGATAACCTGATCGTTCGCGCAGCGCGA	G T AC CA C A C G C C A C A T A G A C T A A C C T A C T C T C G							1059
GlnPheLeuAspTyrGlyAspThrIleSerIleGluLeuArgAspAspGlyAspIleArgLeuLeuThrProValGluGlyValGluHisGluAspAsnLeuIleValArgAlaAlaArg	LeuHis Pro His Glu His Asn Asn							1156
TTGTTGATGAAAACTCGCGCAGACAGCGGGCTCTTCCGACGGGAAGCGGTGCGAATATCAGCATTGACAACGCTTTGCCGATGGGCGGCGGTCTCGGCGGTGGTTTCATCCAATGCCGCG	C GTC CT G G T C G CG C G C C T C G C C T C G C C T C							1179
LeuLeuMetLysThrAlaAlaAspSerGlyArgLeuProThrGlySerGlyAlaAsnIleSerIleAspLysArgLeuProMetGlyGlyGlyLeuGlyGlyGlySerSerAsnAlaAla	Val SerGlu Ala Asp Glu							1276
ACGGTCTCGTGGCATTAAATCATCTCTGGCAATGCGGGCTAAGCATGGATGAGCTGGCGGAAATGGGGCTGACGCTGGGCGCAGATGTTCTCTGTTTTCGGGGGCATGCCGCGTTT	C T GC C T TTC T A ACGC C C C C C G T C C							1299
ThrValLeuValAlaLeuAsnHisLeuTrpGlnCysGlyLeuSerMetAspGluLeuAlaGluMetGlyLeuThrLeuGlyAlaAspValPheValArgGlyHisAlaAlaPhe				Ile	ThrLeu			1396
GCCGAAGGCGTTGGTAATACTAACCGCGGTGGATCCGCCAGAGAAGTGGTATCTGGTGGCGCACCCCTGGTGTAAGTATTCGACTCCGGTGATTTTAAAGATCCTGAACTCCCGCGC	A C T A G A A C C A C T C A T C T G T							1419
AlaGluGlyValGlyGluIleLeuThrProValAspProProGluLysTrpTyrLeuValAlaHisProGlyValSerIleProThrProValIlePheLysAspProGluLeuProArg	Asn						Gln	1516
AATACGCCAAAAAGGTCAATAGAAACGTTGCTAAAAATGTGAATTGACCAATGATTGCGAGGTTATCGCAAGAAAAAGTTTTCGCGAGGTTGATGCGGTGCTTTCCTGGCTGTTAGAATAC	T A C A C C A A A C G T T T C G C A G G T T G A T C C G T T T C C T G G C T G T T A G A A T A C							1539
AsnThrProLysArgSerIleGluThrLeuLeuLysCysGluPheSerAsnAspCysGluValIleAlaArgLysArgPheArgGluValAspAlaValLeuSerTrpLeuLeuGluTyr	Asp						Ala	1636
GCCCCGTGCGCCCTGACTGGGACAGGGGCTGTGCTCTTGTGCTGAATTTGATACAGAGTCTGAAGCCCCCGCAGGTGCTAGAGCAAGCCCCGGAATGGCTCAATGGCTTTGTGGCGAAAGGC	G C C C C T G T G C T C T T G T G C T G A A T T G A T A C A G A G T C T G A A G C C C C G C A G G T G C T A G A G C A A G C C C C G G A A T G G C T C A A T G G C T T T G T G G C G A A A G G C							1659
AlaProSerArgLeuThrGlyThrGlyAlaCysValPheAlaGluPheAspThrGluSerGluAlaArgGlnValLeuGluGlnAlaProGluTrpLeuAsnGlyPheValAlaLysGly					Cys		Ala	1756
GCTAATCTTTCCCATTTGCACAGAGCCATGCTTTAA	TC C C TC AGT A C CGCATTCGGGTTTCAGGAAGATGCGGAAATGACGAGTCGCCGGTCATATATGCCGTATTTGACCGGTGTGAGGCATCGATACCA							1876

Fig. 1. Nucleotide sequences and deduced amino acid sequences for ORF1 and ORF2 from the *hemA* to *prs* segment of the *E. coli* and *S. typhimurium* chromosome. The sequence from *E. coli* is the top line in each case and differences only in the *S. typhimurium* sequences are shown below. The putative -35 and -10 sequences for P_1 (Verkamp & Chelm, 1989) and P_2 (Hove-Jensen *et al.*, 1986) are underlined. ORFs are designated above their initiator codons. Stop codons are boxed. Dashes denote deleted nucleotides. Both sequences were determined on both strands with complete overlaps of all cloning junctions. The nucleotide sequence determined in the present work was 921 bp of the *E. coli* chromosome from the *Sall* site at nucleotide number 538 to the *EcoRI* site at nucleotide number 1459, and 822 bp of the *S. typhimurium* chromosome from the *Bam*HI site at nucleotide number 734 to the *EcoRI* site at nucleotide

obtained from a *prs* mutant strain (Post & Switzer, 1991), was isolated and inserted into the polylinker of pUC18, yielding plasmid pPE4 (Fig. 2). Since the *S. typhimurium* chromosomal DNA contained on plasmid pPS3 was isolated from a mutant strain, the DNA sequence from the plasmid pTE201, isolated from wild type *S. typhimurium* (Elliott, 1989), was also determined on one strand from the *Bam*HI to *Eco*RI sites. The sequences derived from pPS3 and from pTE201 were identical.

For cloning from *E. coli*, a gene library was constructed in the λ D69 vector (Misuzawa & Ward, 1982). Bacterial DNA, isolated from strain HO241 (Hove-Jensen, 1985), was digested with *Hind*III and ligated to *Hind*III digested λ D69 vector DNA. Mature λ particles were generated by *in vitro* packaging (Sternberg *et al.*, 1977) and the resulting lysate was amplified by growth in strain HO340 (Nilsson & Hove-Jensen, 1987). From the gene library several *hemA* transducing phages were isolated by complementation of the mutant *hemA* allele harboured in strain HO371 (Bower *et al.*, 1989). The DNA was isolated from one transducing λ strain and shown to contain a 6.1 kb *Hind*III DNA fragment. This DNA fragment was subcloned in pBR322 and further into appropriate phagemid vectors. Restriction endonuclease and complementation analyses showed that the 6.1 kb *Hind*III DNA fragment contained the *hemA*, ORF1, ORF2 and a small part of the *prs* coding sequence.

Nucleic acid methods. Techniques for manipulation of DNA and RNA were described previously (Davis *et al.*, 1980; Sambrook *et al.*, 1989; Schleif & Wensink, 1981; Silhavy *et al.*, 1984). DNA nucleotide sequence was determined by the dideoxy-chain termination method of Sanger *et al.* (1980) for double-stranded plasmid templates (pBluescript or pUC), which had been purified on Qiagen columns, or for M13 bacteriophage templates, using the Sequenase kit (US Biochemicals). Sequence data were analysed with software from DNASTar.

Northern hybridization. RNA was purified as described by Sambrook *et al.* (1989). Total cellular RNA (15 μ g) from *E. coli* JM83 (Yanisch-Perron *et al.*, 1985) bearing the plasmids of interest or RNA (5 μ g) from *S. typhimurium* JL1002 (Bower *et al.*, 1988) were used per lane and electrophoresed in 1% (w/v) formaldehyde agarose gel. RNA was then blotted to a nylon membrane (Gene Screen or Nytran) by using a vacuum blotter (Pharmacia) and the RNA was cross-linked to the membrane using a UV cross-linker (Bio-Rad). When more than one probe was used, the lanes were run in duplicate. After cross-linking, the membrane was cut into strips, and the strips were hybridized separately with the desired probes. Prehybridization and hybridization were carried out in Seal-a-Meal bags (Sears) as described by the supplier of the membrane.

Probes against ORF1, ORF2 and *prs*, illustrated in Fig. 4, were inserted into the polylinker of pTZ18R or pTZ19R. The probe for the *bla* gene was prepared as described by Sorenson *et al.* (1989). The pTZ-probe-containing plasmids were purified by CsCl gradient centrifugation, digested with the appropriate restriction enzyme to linearize the plasmid, phenol extracted, precipitated and brought up to a final concentration of 1 mg ml⁻¹. RNA probes were made from the plasmid template as described by Promega (1991).

Methods of protein analysis. Electrophoresis of proteins on polyacrylamide gels under denaturing conditions was performed as described by Laemmli (1970). Samples (100 μ l) of overnight cultures were centrifuged in a microfuge, 20 μ l of SDS containing sample buffer was added to the cell pellet, the suspension was vigorously mixed and incubated at 100 °C for 5 min, and the entire sample was analysed. Amino acid sequencing was performed by electroblotting 50 μ g of crude extract containing the overproduced proteins onto a PVDF membrane as described by Matsudaira (1987). The bands corresponding to the two ORF-encoded proteins were excised and sequenced by the University of Illinois Biotechnology Center in an Applied Biosystems 470A gas-phase sequencer.

Results

Nucleotide sequence of the *hemA* to *prs* region

The nucleotide sequence of the approximately 2 kb segment of the *E. coli* and *S. typhimurium* chromosome that lies between the *hemA* and *prs* genes is presented in Fig. 1. Portions of the sequence of this region were previously reported by other investigators; a number of corrections to these published sequences are listed in the legend in Fig. 1. Inspection of the sequenced region revealed that in both *E. coli* and *S. typhimurium* two ORFs are found between the *hemA* and *prs* genes (Fig. 2). The sequence of this region from the two species are very similar: 80 and 84% of the bases are identical for ORF1 and ORF2, respectively, and the deduced amino acid sequences are 91% identical for both ORFs. However, there are two short inserts found in the *S. typhimurium* sequence that are absent in *E. coli* (dashed lines in Figs 1 and 2). The first of these (102 bp) is found in the leader before ORF1 and the second (115 bp) immediately follows the termination codon of ORF2, but precedes the *prs* gene. This latter insert was previously described by Bower *et al.* (1988). Both ORF1 and ORF2 are transcribed in a direction opposite of *hemA*, apparently from the promoter previously suggested by Verkamp & Chelm (1989), labelled P₁ in Figs 1 and 2. The promoter sequences are identical in *E. coli* and *S. typhimurium*. In *E. coli* the termination codon for ORF1 (UAA) overlaps by one nucleotide with ORF2, whereas in *S. typhimurium* the stop codon (UGA) overlaps ORF2 by 4 bases, since there is a 3 bp deletion in the *S. typhimurium* sequence. In *S. typhimurium* there

number 1556. Portions of this sequence previously reported by others are: in *E. coli*, nucleotides 1 to 175 (Drolet *et al.*, 1989), nucleotides 1 to 227 (Verkamp & Chelm, 1989) and nucleotides 1 to 1337 (Li *et al.*, 1989) and, nucleotides 1337 to 1848 (Hove-Jensen *et al.*, 1986), and in *S. typhimurium*, nucleotides 1 to 738 (Elliott, 1989) and nucleotides 1497 to 2060 (Bower *et al.*, 1988). Corrections to the previously published sequences are as follows. In *E. coli*, nucleotides 651, 745 and 1179 were deleted (double underlining); had inserted nucleotides at positions 802 and 822 (*); and reported different nucleotides at positions 621–2, and 1274 (bold). In *S. typhimurium* nucleotide 1741 was deleted (double underlined) and different nucleotides were reported at positions 1498 and 1694–5 (bold). A consequence of the sequencing errors in *E. coli* is that both ORF1 and ORF2 were not detected, and the sequencing errors in *S. typhimurium* shifted the C-terminal reading frame for ORF2. The *E. coli* and *S. typhimurium* nucleotide sequences reported in Fig. 1 have been assigned GenBank accession numbers M77237 and M77236, respectively.

Fig. 2. Schematic representation of the ORFs between *hemA* and *prs*, and the plasmids used in characterizing them. Boxes indicate the ORFs, with arrows indicating the direction of their translation; broken boxes are not drawn to scale. Dotted lines indicate segments of DNA found in *S. typhimurium*, but not the *E. coli*. Promoters (P_1 and P_2) and their directions of transcription are indicated by short arrows. Plasmids used for studying expression of ORF1 and ORF2 are shown by lines below the diagram with restriction sites used for insertion of the segments into the multiple cloning site of pUC18 indicated at their ends.

Fig. 3. SDS-PAGE of extracts of *S. typhimurium* SB139 bearing plasmids encoding ORF1 and ORF2. Lanes 1 and 10: molecular mass standards. Plasmids in the cells are described in Fig. 2 and were: lane 2, pUC18; lane 3, pPE1; lane 4, pPE3; lane 5, pPE4; lane 6, pDPH12; lane 7, pDPH22; lane 8, pDPH13; lane 9, pUC18. Overexpressed ORFs are indicated with labelled pointers.

are two possible AUG start codons for ORF2. Microsequencing of the overproduced plasmid-encoded *S. typhimurium* ORF2-encoded protein (see below) revealed that 75% of the protein had the sequence of Met-Met-Thr-His-Trp-?-Ser-Pro-Ala-Lys and 25% had the sequence Met-Thr-His-Trp-Pro-?-Pro-Ala-Lys-Leu. Thus,

it appears that both start codons are used, with the first AUG being chosen 3 times as often as the second. In both species, ORF2 is terminated at a UAA codon, which is immediately followed by the 115 bp insert in *S. typhimurium* (Figs 1 and 2).

Overexpression of the proteins encoded by ORF1 and ORF2

Plasmids were constructed with one or both of the ORFs from *E. coli* or *S. typhimurium* in either pUC18 or pUC19, so that they would be transcribed from the plasmid-borne *lac* promoter (Fig. 2). The plasmids were expressed in *S. typhimurium* strain SB139 (Bower *et al.*, 1988), and crude extracts were analysed by electrophoresis on denaturing polyacrylamide gels (Fig. 3). Both ORF1- and ORF2-encoded proteins, which were predicted to have molecular masses of 23600 and 30900 Da, respectively, were well overproduced and migrated at their expected molecular masses in extracts of cells bearing the appropriate plasmids (Fig. 3). Since the ORF2-encoded protein has the same molecular mass as the unprocessed form of β -lactamase, the β -lactamase catalytic activity of all the extracts was determined and found to be essentially constant. Thus, it is very unlikely that the large amount of 31 kDa protein seen in Fig. 3 was β -lactamase instead of the ORF2-encoded protein. N-terminal sequence analysis proved the identity of the 31 kDa protein as coded by ORF2. Applications of the same procedure to the ORF1- encoded protein did not yield any useful sequence data. Since the proteins encoded by ORF1 and ORF2 were detected only by overproduction in cells bearing high-copy plasmids

Fig. 4. Plasmids and hybridization probes used in analysis of transcripts. Symbols are as in Fig. 2. Plasmids used for overproduction of mRNA are indicated with restriction sites for introduction of chromosomal DNA into pUC18 (pDPH13) or pUC19 (pDPH31). Single stranded RNA probes used in the Northern blots in Figs 5 and 6 are indicated below their corresponding complementary sequences in each ORF with a dash-dot line, and the lengths of the probes, in nucleotides, are given.

Fig. 5. Northern blot hybridization analysis of *prs* and ORF-encoding transcripts. RNA was extracted as in Methods from *E. coli* JM83/pUC18 (lanes 1, 3, 5, 7) and from JM83/pDPH13 (lanes 2, 4, 6, 8), and probed with probes described in Fig. 4. Pointers indicate transcripts and their sizes in kb.

bearing these ORFs, it is possible that these proteins are not actually formed from the chromosomal copy of ORF1 and ORF2 in cells lacking plasmids. However, it will be shown in the next section that transcripts encoding ORF1 and ORF2 are formed from the chromosomal copy of these genes. Thus, it seems likely that these proteins are actually synthesized from such transcripts, although no doubt at much lower levels than seen in Fig. 3.

Northern blot analysis of ORF1-, ORF2- and *prs*-encoding transcripts in *E. coli*

prs-encoding transcripts have been previously mapped (Hove-Jensen *et al.*, 1986) to a promoter (designated P₂

in Figs 2 and 4) located within ORF2. ORF1 and ORF2 are presumably transcribed from P₁ (Verkamp & Chelm, 1989) (Fig. 4). Do transcripts encoding ORF1 and ORF2 read through and include the *prs* gene? Alternatively, are such transcripts terminated before the *prs* gene? If so, would transcription termination interfere with *prs* transcription? To investigate these questions, transcripts from *E. coli* JM83 cells bearing the pDPH13 plasmid, which encodes the *E. coli* ORF1, ORF2 and *prs* genes (Fig. 4), were analysed by Northern blotting. As a control, transcripts from JM83 bearing pUC18 were also analysed. Radiolabelled single-stranded RNA probes against ORF1, ORF2 and the *prs* sequences shown in Fig. 4 were prepared and used for hybridization analysis as described in Methods. Fig. 5 shows clearly that the

prs-encoding transcripts of the expected lengths (1.3 and 2.7 kb) were found. The relative amounts of these two transcripts was the same in cells grown on minimal and rich medium. Both the 1.3 and 2.7 kb transcripts were also detectable in the same relative amounts in cells that did not contain the pDPH13 plasmid, i.e. which contained only chromosomally encoded transcripts, although they were present at much lower levels. Thus, we do not believe that the production of the less abundant 2.7 kb transcript encoding ORF1 and ORF2 is an artifact of using high-copy plasmids to detect it.

Excision of the transcript bands and determination of their radioactivity by liquid scintillation counting showed about 95% of the total in the smaller (1.3 kb) species. When ORF1 and ORF2 probes were used, both were found to be encoded on the larger (2.7 kb) transcript, as well as smaller bands (1.4 and 1.7 kb), which may be degradation products or premature termination products (Fig. 5). *bla* transcripts (1.0 and 1.1 kb) (Sorenson *et al.*, 1989), from JM83/pUC18 (Fig. 5, lane 7) were readily detectable using a probe to the *bla* gene (Sorenson *et al.*, 1989), showing that the control contained abundant hybridizable RNA. *bla* transcripts were also detectable in JM83/pDPH13 cells (Fig. 5, lane 8) but were much less abundant than with pUC18, presumably because the larger pDPH13 plasmid was maintained at a lower copy number.

Similar Northern blot analysis was performed with *S. typhimurium* SB139/pDPH31, which carried the corresponding region from the *S. typhimurium* chromosome (Fig. 4), using an RNA probe derived from the *S. typhimurium prs* sequence. *prs*-encoding transcripts were detected just as with *E. coli*, except that they were of slightly larger sizes, 1.4 and 2.9 kb, as expected from the DNA sequence (data not shown).

Pyrimidine starvation increases transcription at P₂ only

Derepression of PRPP synthetase is observed upon growth of a pyrimidine auxotrophic strain in the presence of orotic acid (White *et al.*, 1971), so the *S. typhimurium* strain JL1002 (Δ pyrA81) (Bower *et al.*, 1988) was grown in the presence of orotic acid on minimal medium. Cells grown under these conditions typically had a 3- to 5-fold higher specific activity of PRPP synthetase than those grown with uracil. It had been suggested earlier that derepression was the result of increased transcription at promoter P₂ from a plasmid-borne *prs* gene (Hove-Jensen, 1985). Thus, to determine whether derepression resulted from increased transcription from one or both of the promoters identified above, RNA was extracted from starved and uracil-grown cells and analysed on Northern blots. The strain used did not contain plasmid,

Fig. 6. Northern blot hybridization analysis of transcripts from pyrimidine-starved cells and cells grown with 20 µg uracil ml⁻¹. *S. typhimurium* JL 1002 (Δ pyrA81) was grown and RNA extracted as in Methods. The membrane was probed with the 832 bp *S. typhimurium prs* probe depicted in Fig. 4. Pointer indicates transcript from promoter P₂ and its size in kb. The 2.9 kb tricistronic transcript was not detected under the exposure conditions used in this experiment.

so only transcription from the chromosome was examined. Pyrimidine starvation clearly increased the amount of *prs*-encoding transcript from P₂ (1.4 kb in *S. typhimurium*) (Fig. 6). Excision and counting of the transcript bands showed that these transcripts were 3 times as abundant in starved cells as in uracil-grown cells. The specific activity of PRPP synthetase was 4 times as great in the same pyrimidine-starved cells, and the amount of protein immunochemically cross-reactive with PRPP synthetase on immunoblots was 3 times as abundant as in uracil-grown cells. Thus, pyrimidine starvation causes derepression of PRPP synthetase largely or entirely as a consequence of increased transcription at P₂.

Discussion

In this research we sought to clarify the role of an upstream ORF that overlaps the *prs* promoter in *S. typhimurium* and *E. coli* (Bower *et al.*, 1988; Hove-Jensen *et al.*, 1986). While the work was in progress, several laboratories published portions of the sequence upstream of the *prs* gene and established that *prs* was separated from *hemA* by only about 2 kb (Drolet *et al.*, 1989; Elliott, 1989; Li *et al.*, 1989; Verkamp & Chelm, 1989).

However, the sequence of this region was not fully determined in *S. typhimurium* and the *E. coli* sequence contained several errors which obscured the existence of two ORFs upstream of the *prs* gene. The results of this paper clarify the structure of the *hemA* to *prs* region and its transcription in *E. coli* and *S. typhimurium*.

Our experiments established clearly that the *prs* gene is transcribed from two promoters: P₁, which yields a tricistronic transcript encoding ORF1, ORF2, and *prs* genes, and P₂, which yields a monocistronic transcript encoding *prs* only. The transcripts originating from P₁ constitute only 5% or less of the total *prs*-encoding mRNA under all conditions examined. Since the monocistronic P₂ transcripts make up 95% of *prs* message and P₂ appears to be the sole site of regulation of *prs* expression by pyrimidines, it is doubtful that P₁ transcripts are important in the physiological regulation of PRPP synthetase activity.

Although it has been known for some time that pyrimidine starvation results in a modest (usually 3- to 5-fold) derepression of PRPP synthetase in *S. typhimurium* (Olszowy & Switzer, 1972; White *et al.*, 1971) and *E. coli* (Hove-Jensen, 1985; Neuhaard & Nygaard, 1987), the present results provide the first direct proof that such derepression results from increased transcription at the P₂ promoter. The biochemical mechanisms governing pyrimidine regulation of transcription are unknown. No regulatory protein or transcriptional attenuation mechanism involving regulation of the *prs* gene has yet been identified.

The physiological function of ORF1 and ORF2 are unknown, although the failure of attempts to obtain disruption mutations in ORF1 (T. Elliott, personal communication) and ORF2 (unpublished results) suggests that they may be essential for cell viability. None of the protein sequences in the databases we searched had sufficient relatedness to the deduced sequence of ORF1 to allow us to suggest a possible function for this protein. A number of proteins known to bind nucleotides were shown to have limited sequence similarity to the deduced sequence of ORF2 in a glycine-rich segment that is often associated with nucleotide binding sites, so it may be speculated that ORF2 encodes a nucleotide-binding enzyme. Overexpression of either or both of the ORFs had no effect on PRPP synthetase levels, so it is unlikely that they act in *trans* to regulate *prs* expression. We have observed that the overproduced, plasmid-encoded ORF2 protein is released from *E. coli* cells by osmotic shock, whereas the ORF1 protein is not. The Northern blots in Fig. 5 indicated the presence of at least two other species encoding ORF1 and ORF2, besides the intact 2.7 kb tricistronic transcript. The length of these transcripts (1.4 and 1.7 kb) and the fact that they hybridize to both ORF1 and ORF2 probes, but not to the *prs* probe,

indicate that they may result from termination or cleavage of the P₁ transcripts after or at P₂ but preceding the *prs* gene. This raises the interesting possibility that they arise from processes, such as attenuation that could be involved in the regulation of *prs* expression.

We thank Dr T. Elliott for the donation of the plasmid pTE201, Dr M. Kilstrup for the pTZ and pMA31 plasmids and T. Dedenroth Hansen for excellent technical assistance. D.A.P. conducted part of this research at the University of Copenhagen with support from the Danish Research Academy. He expresses his thanks to B.H.-J. and many others at the University of Copenhagen for their help and support during his stay at that institution.

This research was supported by the Public Health Service grant DK13488 from the National Institute of Diabetes, Digestive, and Kidney Diseases to R. L. S. and from the Danish Center of Microbiology and the Danish Research Academy to B.H.-J.

References

- BOWER, S. G., HOVE-JENSEN, B. & SWITZER, R. L. (1988). Structure of the gene encoding phosphoribosylpyrophosphate synthetase (*prsA*) in *Salmonella typhimurium*. *Journal of Bacteriology* **170**, 3243–3248.
- BOWER, S. G., HARLOW, K. W., SWITZER, R. L. & HOVE-JENSEN, B. (1989). Characterization of the *Escherichia coli* *prsA1*-encoded mutant phosphoribosylpyrophosphate synthetase identifies a divalent cation-nucleotide binding site. *Journal of Biological Chemistry* **264**, 10287–10291.
- CLARK, D. J. & MAALØE, O. (1967). DNA replication and the division cycle in *Escherichia coli*. *Journal of Molecular Biology* **23**, 99–112.
- DAVIS, R. W., BOTSTEIN, D. & ROTH, J. R. (1980). *Advanced Bacterial Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- DROLET, M., PELOQUIN, L., ECHELARD, Y., COUSINEAU, L. & SASARMAN, A. (1989). Isolation and nucleotide sequence of the *hemA* gene of *Escherichia coli* K 12. *Molecular and General Genetics* **216**, 347–352.
- ELLIOTT, T. (1989). Cloning, genetic characterization, and nucleotide sequence of the *hemA-prfA* operon of *Salmonella typhimurium*. *Journal of Bacteriology* **171**, 3948–3960.
- HOVE-JENSEN, B. (1985). Cloning and characterization of the *prs* gene encoding phosphoribosylpyrophosphate synthetase of *Escherichia coli*. *Molecular and General Genetics* **201**, 269–276.
- HOVE-JENSEN, B., HARLOW, K. H., KING, C. J. & SWITZER, R. L. (1986). Phosphoribosylpyrophosphate synthetase of *Escherichia coli*: Properties of the purified enzyme and primary structure of the *prs* gene. *Journal of Biological Chemistry* **261**, 6765–6771.
- LAEMMLI, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, London* **227**, 209–217.
- LI, J.-M., RUSSELL, C. S. & COSLOY, S. D. (1989). Cloning and structure of the *hemA* gene of *Escherichia coli* K-12. *Gene* **82**, 209–217.
- MATSUDAIRA, P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *Journal of Biological Chemistry* **262**, 10035–10038.
- MILLER, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MISUZAWA, S. & WARD, D. F. (1982). A bacteriophage lambda vector for cloning with *Bam*HI and *Sau*3A. *Gene* **20**, 317–322.
- NEUHAARD, J. & NYGAARD, P. (1987). Purines and pyrimidines. In *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*, vol. 1, pp. 445–473. Edited by J. L. Ingraham, K. B. Low, F. C. Neidhardt, B. Magasanik, M. Schaechter & H. E. Umbarger. Washington DC: American Society for Microbiology.
- NILSSON, D. & HOVE-JENSEN, B. (1987). Phosphoribosylpyrophosphate synthetase of *Bacillus subtilis*. Cloning, characterization and chromosomal mapping of the *prs* gene. *Gene* **53**, 247–255.
- OLSZOWY, J. & SWITZER, R. L. (1972). Specific repression of

- phosphoribosylpyrophosphate synthetase by uridine compounds in *Salmonella typhimurium*. *Journal of Bacteriology* **110**, 450–451.
- POST, D. & SWITZER, R. L. (1991). *prsB* is an allele of the *Salmonella typhimurium prsA* gene: characterization of a mutant phosphoribosylpyrophosphate synthetase. *Journal of Bacteriology* **173**, 1978–1986.
- PROMEGA CORPORATION. (1991). *Catalogue and Reference Guide – Biological Research Products*, pp. 58–61. Madison, WI.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SANGER, F., COULSEN, A. R., BARRELL, B. G., SMITH, A. J. H. & ROE, B. A. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology* **143**, 161–178.
- SCHLEIF, R. F. & WENSINK, P. C. (1981). *Practical Methods in Molecular Biology*. New York: Springer-Verlag.
- SILHAVY, T. J., BERMAN, M. L. & ENQUIST, L. W. (1984). *Experiments with Gene Fusions*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- SORENSEN, M. A., KURLAND, C. G. & PEDERSEN, S. (1989). Codon usage determines translation rate in *Escherichia coli*. *Journal of Molecular Biology* **207**, 365–377.
- STERNBERG, N., TIEMEIER, D. & ENQUIST, L. (1977). *In vitro* packaging of a lambdaDam vector containing *EcoRI* DNA fragments of *Escherichia coli* and phage P1. *Gene* **1**, 255–280.
- VERKAMP, E. & CHELM, B. K. (1989). Isolation, nucleotide sequence, and preliminary characterization of the *Escherichia coli* K-12 *hemA* gene. *Journal of Bacteriology* **171**, 4728–4735.
- WHITE, M. N., OLSZOWY, J. & SWITZER, R. L. (1971). Regulation and mechanism of phosphoribosylpyrophosphate synthetase: repression by end products. *Journal of Bacteriology* **108**, 122–131.
- YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.