

Structure of the *Escherichia coli pyrE* operon and control of *pyrE* expression by a UTP modulated intercistronic attenuation

Peter Poulsen, Fons Bonekamp and Kaj Frank Jensen

University of Copenhagen, Institute of Biological Chemistry B, Sølvgade 83, DK-1307 Copenhagen K, Denmark

Communicated by A.Munch-Petersen

Protein synthesis in 'minicells' showed that the DNA immediately preceding the *pyrE* gene of *Escherichia coli* directs the formation of considerable amounts of a polypeptide (mol. wt. ~ 30 000) of unknown function. The nucleotide sequence of this DNA revealed the existence of an open reading frame (ORF) of 238 codons that ends 68 nucleotide residues upstream to the structure start of *pyrE*, just prior to the GC-rich symmetry region of a sequence with features characteristic of a *rho*-independent transcription terminator. Deletion of the start of this 238 codons long ORF gene resulted in a dramatic fall in the level of *pyrE* expression, indicating that the two genes (ORF and *pyrE*) constitute an operon. S1-nuclease digestion of RNA-DNA hybrids revealed that both genes are transcribed from two promoters (P1 and P2) located in front of the ORF start. Furthermore, when the RNA used in these experiments was prepared from cells with different levels of *pyrE* expression, created by manipulations in their pyrimidine nucleotide supply, the frequency of transcription initiations at P1 and P2 was found to be constitutive, while only a pyrimidine regulated fraction of the mRNA chains reached into the *pyrE* gene. *In vitro* transcription of isolated DNA fragments showed that the mRNA chains are terminated between the ORF gene and *pyrE*. From these observations we conclude that *pyrE* expression is controlled by a UTP modulated intercistronic attenuation.

Key words: intercistronic attenuation/*pyrE* operon/pyrimidine biosynthesis/translational coupling to transcription/UTP modulated attenuation

Introduction

The *pyrE* gene encodes the enzyme orotate phosphoribosyltransferase [EC 2.4.2.10] which participates in the *de novo* biosynthesis of UMP, the precursor of all other pyrimidine nucleotides. The gene is located at 81 minutes on the linkage map of the *Escherichia coli* chromosome, between *dut* and *spoT* (Bachman and Brooks Low, 1980), and is transcribed in a counter-clockwise direction towards *dut* (Poulsen *et al.*, 1983). The expression of the *pyrE* gene, as well as of *pyrB* and *pyrF*, is controlled primarily by the intracellular concentration of UTP and is high when the UTP pool is low, while it is low when the concentration of UTP is high (Schwartz and Neuhard, 1975; Kelln *et al.*, 1975; Piérard *et al.*, 1976). In addition, guanine nucleotides seem to participate in this regulation (Jensen, 1979). So far the only mutations identified as causing high levels of *pyrB* and *pyrE* expression in the presence of high intracellular concentrations of repressing nucleotides are RNA polymerase (*rpoBC*) mutants, indicating that the expression of these *pyr* genes is controlled by the degree of

saturation of RNA polymerase with UTP (Jensen *et al.*, 1982).

The nucleotide sequence of the *pyrB* gene (Roof *et al.*, 1982; Hoover *et al.*, 1983; Navre and Schachman, 1983; Turnbough *et al.*, 1983) and of the *pyrE* gene (Poulsen *et al.*, 1983) has been published. Immediately preceding the protein coding region of both genes there is a structure with features of a *rho*-independent transcription terminator, suggesting that the two genes are regulated by attenuation (Roof *et al.*, 1982; Navre and Schachman, 1983; Turnbough *et al.*, 1983; Poulsen *et al.*, 1983). In the *pyrE* gene an open reading frame (ORF) with the capacity to encode a hypothetical leader peptide 44 amino acid residues long was found to straddle the terminator, and it was suggested that the frequency of termination at the *pyrB* attenuator is determined by UTP-induced variations in the tightness of coupling between the transcribing RNA polymerase and the elongating ribosome during the synthesis of this leader peptide (Turnbough, 1983; Turnbough *et al.*, 1983).

We previously determined the nucleotide sequence of the structural *pyrE* gene of *E. coli*, including ~400 nucleotide residues upstream of the coding frame for orotate phosphoribosyltransferase (Poulsen *et al.*, 1983). In this sequence we were unable to find ORFs for putative leader peptides with functions similar to that of the *pyrB* gene. Instead we observed that the entire DNA segment of 324 bp upstream of the *pyrE* attenuator structure constituted an ORF that might well represent the end of a preceding gene. Furthermore, the only sequences of any similarity to the consensus promoter were embedded in this ORF (Poulsen *et al.*, 1983). Here we show that *pyrE* is expressed as the second gene in an operon after the gene for a polypeptide (mol. wt. 25 497) of unknown function, and that the regulation of *pyrE* expression occurs by a modulated attenuation between the two cistrons of this operon.

Results

*Identification of a protein coding segment of the DNA in front of the *pyrE* gene*

Plasmid pPP1 contains an intact and normally regulated *pyrE* gene (Table I) cloned as a 2.0-kb *PvuII-PvuII* DNA fragment in the vector pBR322 (Figure 1). When this plasmid was introduced in a 'minicell' strain and the [³⁵S]methionine-labeled proteins, made by the 'minicells', were analysed by SDS-polyacrylamide gel electrophoresis three polypeptide bands (A, B, and C) were observed (Figure 2, lane 1) in addition to the bands seen with the vector pBR322 (Figure 2, lanes 5 and 8). The peptide band C (mol. wt. 24 000) corresponds to the subunit of orotate phosphoribosyltransferase, encoded by *pyrE* (Poulsen *et al.*, 1983), while the two polypeptides A and B (mol. wts. 32 000 and 30 000) arise from unknown genes in the plasmid. An inversion of the orientation of the *PvuII-PvuII* fragment relative to the vector did

not alter this protein pattern (plasmid pPP2, Figure 2, lane 2). The polypeptides were also made with plasmid pPP4 (Figure 2, lane 3), which is a *ClaI*-generated deletion derivative of pPP1 (Figure 1). However, plasmid pPP5, which has deleted the *SphI-SphI* segment of pPP1 (Figure 1), has lost the ability to form the polypeptides A and B, but does still direct the synthesis of the orotate phosphoribosyltransferase subunit, band C (Figure 2, lanes 4 and 7). Finally, as seen from lane 6 of Figure 2, a deletion of the *BamHI-BamHI* segment of plasmid pPP2 (i.e., plasmid pPP6, Figure 1) results in loss of ability to form the *pyrE* gene product (band C), while the capacity to encode the polypeptides A and B is retained by pPP6. These results show that the ~1100-bp segment of *E. coli* DNA between the *ClaI* site and the *BamHI* site in the start of *pyrE* directs the synthesis of both polypeptides A and B, and that the coding frame for both chains starts between

the *ClaI* and the *SphI* sites (Figure 1). Provided that overlapping genes do not exist in *E. coli* the two polypeptide bands must therefore represent two forms of the same protein arising either by post-translational processing, or by alternative sites for initiation (or termination) of translation, or more likely, perhaps, they may be caused by a partial denaturation of the protein by dodecyl sulfate.

Nucleotide sequence of an ORF encoding the polypeptides A and B

The 342 bp in front of the *pyrE* attenuator constitute an ORF in the same orientation as the *pyrE* gene (Poulsen *et al.*, 1983). Since there are only ~450 additional base pairs up to the *ClaI* site, this ORF must represent the promoter-distal end of the structural gene for the polypeptides A and B. Therefore, we continued the sequence determinations beyond the *MluI* site almost to the *PvuII* site at the end of the cloned fragment in pPP1. Figure 3 shows the strategy used. The nucleotide sequences have been determined in both strands of the DNA at least once, and all restriction sites have been straddled by sequence determinations from different sites. The results are given in Figure 4. The ORF extends from position 369 to position 1082 (Figure 4), i.e., it is 238 codons long. Figure 4 also shows the deduced sequence of the 238 amino acid residues of the protein product, mol. wt. 25 497. An analysis of the codon usage (Grosjean and Fiers, 1982) indicates that frame-shifts have not been introduced during sequencing. Furthermore, the hydrophobic profile of the deduced amino acid sequence, calculated as described by Kyte and Doolittle (1983) (not shown), reveals the existence of alternating hydrophobic and hydrophilic segments in the se-

Table I. Regulation of the *pyrE* gene cloned in plasmid pPP1

Strain	Uracil supplement (20 µg/ml)	Orotate phosphoribosyltransferase (units/mg)
SΦ1240 (<i>pyrH</i>)	absent	10.5
SΦ1258 (<i>pyrE</i>)	absent	1.1
SΦ1258 (<i>pyrE</i>)	present	0.4

The strains, transformed with plasmid pPP1, were grown at 37°C in a glucose-casamino acids medium supplemented with uracil as indicated and with ampicillin (25 µg/ml). At OD₄₃₆ ~0.8, the cultures were harvested for determination of orotate phosphoribosyltransferase activity. (When OD₄₃₆ in the cultures was ~0.5, aliquots were harvested for isolation of RNA.)

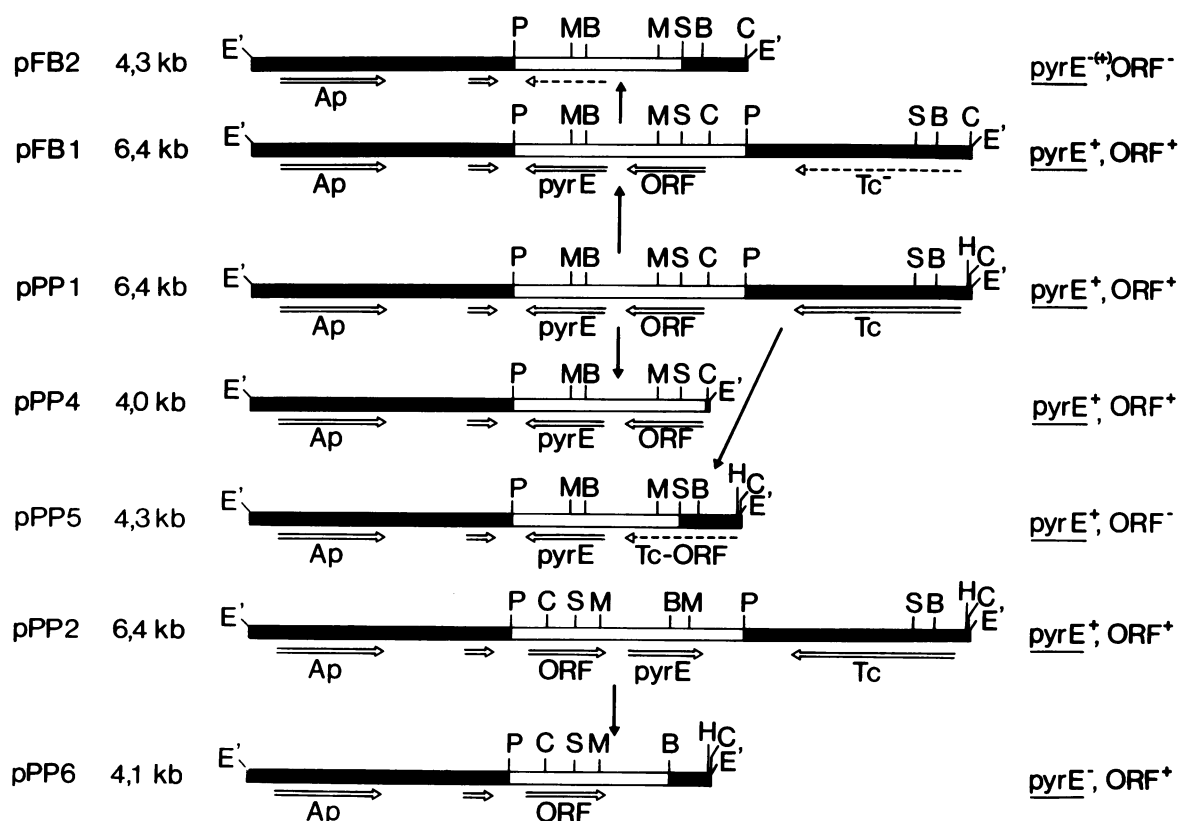


Fig. 1. Plasmids used. The plasmids were constructed as described in Materials and methods. Abbreviations: E', half site for *EcoRI*; B, site for *BamHI*; C, *ClaI*; H, *HindIII*; M, *MluI*; S, *SphI*. The arrow denoted Ap is the gene for β -lactamase, and Tc is the gene for tetracycline resistance. The small arrow without designation is the origin of replication. ORF denotes a 238 codons long open reading frame (see text).

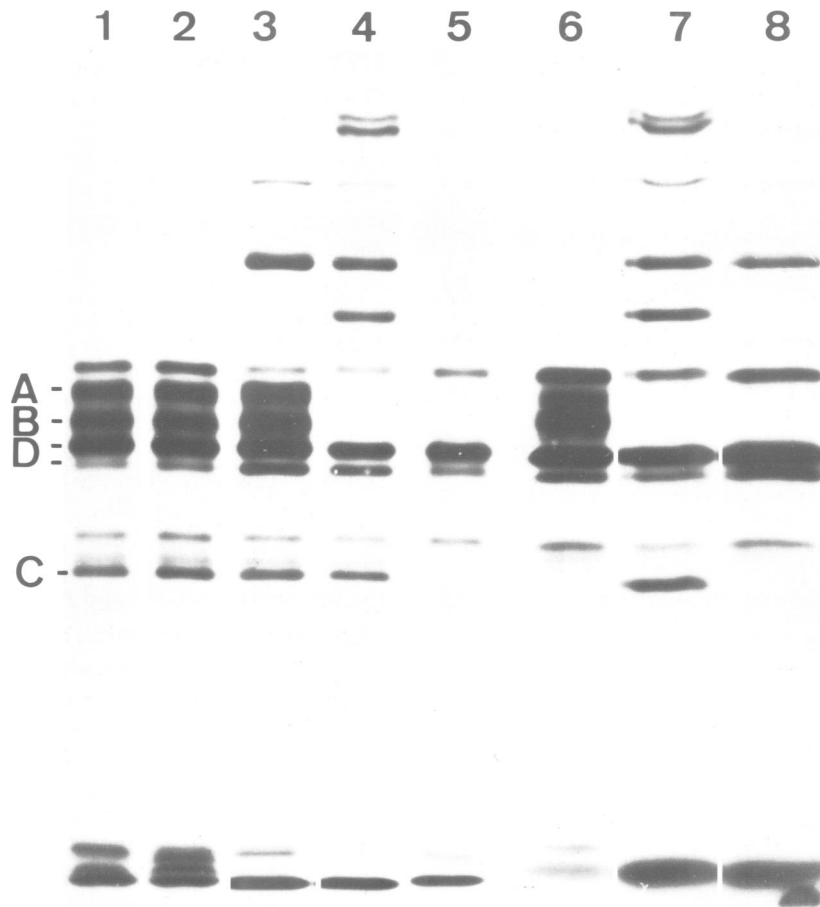


Fig. 2. Identification, using 'minicells', of a protein coding DNA segment preceding the *pyrE* gene. Proteins made by 'minicells' containing the indicated plasmids were labeled with [³⁵S]methionine and analysed by electrophoresis in SDS-polyacrylamide gels. The figure shows an autoradiographic picture: **lane 1**, pPP1; **lane 2**, pPP2; **lane 3**, pPP4; **lanes 4 and 7**, pPP5; **lanes 5 and 8**, pBR322; **lane 6**, pPP6. By comparison with Figure 1 it is seen that the protein bands A and B (apparent mol. wts. 31 500 and 30 000) both are derived from the DNA region between the *Clal* and *Bam*HI sites. The band marked C results from the subunit of orotate phosphoribosyltransferase encoded by *pyrE*, while the bands marked D originate from β -lactamase.

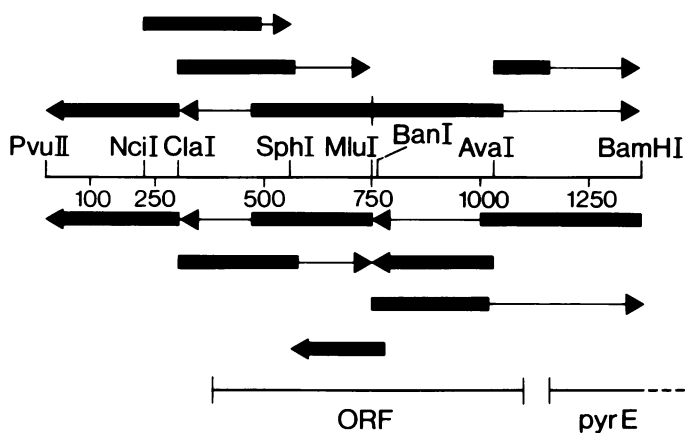


Fig. 3. Restriction endonuclease sites and fragments used for DNA sequencing. The arrows above and below the center line represent the fragments used for sequencing the upper and the lower strand. The heavy and the light lines indicate, respectively, the sequenced and the non-sequenced parts of the fragments. The proximal ends of the arrows represent the labeled ends. The coordinates refer to the numbers of base pairs.

quence, which thereby resembles amino acid sequences of authentic proteins. We cannot at present explain why the protein product of the ORF gene migrates as two species in SDS gels or why the apparent mol. wt., as determined by gel electrophoresis, exceeds that expected from the DNA sequence. However, there is ample evidence indicating that protein migration in SDS gels may not always correlate well with mol. wts. (Beyreuther *et al.*, 1980; Pollitt and Zalkin, 1983).

Operon expression of the ORF gene and pyrE

Since we were unable to find good candidates for promoter structures in the DNA sequence of 400 bp preceding the *pyrE* gene (Poulsen *et al.*, 1983) we considered the possibility that this gene might be expressed in an operon after the ORF gene. This would mean that plasmids lacking the start of ORF should not direct any synthesis of orotate phosphoribosyltransferase although they contain the coding frame of *pyrE*. Different plasmids derived from pPP1 (see Figure 1) were therefore introduced in strain S Φ 1258 (*pyrE*) and the orotate phosphoribosyltransferase was determined (Table II). Plasmid pPP1 results in a level of ~ 0.4 units/mg of this enzyme. Plasmid pPP4, lacking the *Clal*-*Clal* segment of pPP1, gives rise to a similar synthesis of the enzyme. Plasmid pPP5,

CGTTGGGATCGACTCACTTTCGGCGCTAACATCTGACTCATAGCGGGGATACATCCCA 59
 CTTTACCGCGCGTCAGCGGAGAACGAATACGCTCGCGAAGCAGGTTCAAGGCTACGAA 119
 ACTGCTCCGACAGACGAAAGTAAGTTTCCAGATAACCGTGGTTTACCGAGCGCATTTCCA 179
 GGTTCGGCTCCCCCATTCACCCCTTGATTTACCGCCGGCGTAGGCGGTCATACTCGGGAT 239
 CATAGACGTTCCCTGTTTATAAAAAGGAAGCGTGGAAAGGATTATAGCCATCGATGCCTTGT 299
 AGGATAGGAATAACCGCGGGAAGTCGGTATAATCGCGCAACCCACATTTGTTTCAAGCCGA 359
 GATTTCAATATCGCTCCAGCAGCGCGTAGCAATAATCAGGTGGCTCCCGTTACCCGACT 419
 MetArgProAlaGlyArgSerAsnAsnGlnValArgProValThrLeuThr
 CGTAACTATACAAAACATGCAGAAGGCTCGGTGCTGGTGAATTTGGCGATACCAAAGTG 479
 ArgAsnTyrThrLysHisAlaGluGlySerValLeuValGluPheGlyAspThrLysVal
 TTGTGTACCGCCTTATTTGAAGAAGGCGTCCCGCCCTTCCTGAAAGGTCAGGGCCAGGG 539
 LeuCysThrAlaSerIleGluGlyValProArgPheLeuLysGlyGlnGlyGlnGly
 TGGATCACCGCAGAGTACGGCATGCTGCCAGTCTACCCACACCCGTAACGCTCGTGAA 599
 TrpIleThrAlaGluTyrGlyMetLeuProArgSerThrHisThrArgAsnAlaArgGlu
 GCGGCGAAAGGTAAGCAGGTTGGACGCACAATGAAATCCAGCGTCTGATCGCCCGTGT 659
 AlaAlaLysGlyLysGlnGlyArgThrMetGluIleGlnArgLeuIleAlaArgAla
 CTTCCGCGCGCAGTAGATTTGAAAGCGCTGGGTGAGTTCACCATTCAGCTGGACTCGGC 719
 LeuArgAlaAlaValAspLeuLysAlaLeuGlyGluPheThrIleThrLeuAspCysAsp
 GTGCTTCAGGCTGATGCTGACCGTACCGCTCGATTACGGGTGCCCTGCGTGGCCGTCG 779
 ValLeuGlnAlaAspAlaTrpThrArgThrAlaSerIleThrGlyAlaCysValAlaLeu
 GTAGATCGGCTACAGAAGCTGGTGGAAAACCGCAAGCTGAAAACCAATCCGATGAAAGGG 839
 ValAspAlaLeuGlnLysLeuValGluAsnGlyLysLeuLysThrAsnProMetLysGly
 ATGGTAGCCCGCAGTTTCTGTCGGAATTTGTAACCGCGAAGCGGTTTGGCATTCGAATAC 899
 MetValAlaAlaValSerValGlyIleValIleAsnGlyGluAlaValCysAspLeuGlyIle
 GTTGAAGACTCTGCCGACAGACCCGACATGAACGTAGTATGACCCAAAGACGGCGCATC 959
 ValGluAspSerAlaAlaGluThrAspMetAsnValValMetThrGluAspGlyArgIle
 ATTGAAGTGCAGGGGACGGCAGAGCGGCGTTCACCCATGAAGAGCTACTCATCTTG 1019
 IleGluValGlnGlyThrAlaGluGlyGluProPheThrHisGluGluLeuLeuIleLeu
 TTGGCTCTGGCCCGAGGGGGAATCGAATCCATTGTAGCGACGCAAGGCGCGCTGGCA 1079
 LeuAlaLeuAlaArgGlyGlyIleGluSerIleValAlaThrGlnLysAlaAlaLeuAla
 AACGATTTTTTAAAGCGACTGATGAGTCGCTTTTTTGTCTGTAGAAAAGTAAGATGA 1139
 AsnEnd at t .
 GGAGCGAAGGCATGAAACCATATCAGCGCCCAAGTTTAT . . . (pyrE)
 MetLysProLysGlnArgGlnPheIle . . .

Fig. 4. Nucleotide sequence of the DNA preceding the *pyrE* gene. The encircled nucleotide residues (positions 267 and 328) represent the starts of transcripts towards *pyrE*, as mapped by the S1-nuclease technique (see text). The lines labeled P1 and P2 indicate the -10 regions (Rosenberg and Court, 1979; Siebenlist *et al.*, 1980) of the putative promoters dictating transcriptional starts at positions 328 and 267, respectively. The translated amino acid sequence represents the longest possible ORF between the *Clal* cut (after position 286) to the start of the *pyrE* coding segment. 'att.' denotes the symmetry region of the *pyrE* attenuator (see text). The nucleotide sequence downstream to the *MluI* site (position 741) from position no. 758 has previously been published by Poulsen *et al.* (1983).

which lacks the *SphI-SphI* segment of pPP1 and does not produce the protein products of the ORF gene (bands A and B of Figure 2) when introduced in 'minicells', still encodes the synthesis of an amount of orotate phosphoribosyltransferase activity about half of that seen with pPP1 and pPP4 (Table II). However, with this plasmid (pPP5) the possibility exists that the *pyrE* gene is expressed from the *Tet^R* promoter (Figure 1). Therefore two new plasmids pFB1 and pFB2, corresponding to pPP1 and pPP5, were constructed in which the *Tet^R* promoter was destroyed (Figure 1). We opened pPP1 with *HindIII*, treated it with S1-nuclease to remove the 5' overhang, and resealed with T4-DNA ligase. This treatment resulting in plasmid pFB1 strongly decreases the resistance of transformed cells to tetracycline, but does not alter the level of orotate phosphoribosyltransferase activity (Table II) or the synthesis of the protein chains A and B by 'minicells' (not

Table II. Operon expression of the ORF and *pyrE* genes

Plasmid ^a	Orotate phosphoribosyltransferase (units/mg)
pPP1	0.40
pPP4	0.34
pPP5	0.20
pFB1	0.41
pFB2	0.02

^aThe structures of the plasmids are shown in Figure 1.

Strain SØ1258 (*pyrE*) transformed with the indicated plasmids was grown at 37°C in a glucose-casamino acid medium containing ampicillin (25 µg/ml) and uracil (20 µg/ml). At OD₄₃₆ ~0.8, the cells were harvested for determination of specific orotate phosphoribosyltransferase activity.

shown). However, when the *SphI-SphI* fragment was deleted from pFB1, resulting in the formation of pFB2 (Figure 1), the level of orotate phosphoribosyltransferase was reduced by a factor of 15–20 (Table II) showing the operon character of the expression of ORF and *pyrE*.

Identification of ORF-*pyrE* promoters

To identify the promoter(s) of the *ORF-pyrE* operon we prepared RNA from a *pyrH* strain (SO1240) and from strain SO1258 grown in the presence or absence of uracil. Both strains harbored plasmid pPP1, and Table I gives the specific activity of orotate phosphoribosyltransferase in the three cultures. Aliquots of the RNA preparations were first hybridized to the *MluI-BamHI* DNA fragment that covers the start of the *pyrE* gene (Figure 1). The fragment was labeled at the 5' overhang of the *BamHI*-end with [³²P]phosphate. The hybridization mixtures were then treated with S1-nuclease and the pattern of protected DNA fragments was analyzed by electrophoresis in a denaturing polyacrylamide gel (Figure 5A). No start of a transcription towards the *pyrE* gene can be detected in this region of the DNA. However, the probe fragment is protected along its entire length to an extent that parallels the level of orotate phosphoribosyltransferase in the cultures (compare Figure 5A with the data of Table I). Transcription must therefore be initiated upstream to the *MluI* site and a pyrimidine controlled amount of RNA chains reach the *BamHI* site in the beginning of *pyrE*. To map the promoter, aliquots of the RNA preparations were hybridized to the *PvuII-MluI* fragment that covers the start of the ORF gene and which was labeled at the 5' end of the *MluI* overhang. After digestion with S1-nuclease to remove single-stranded nucleic acids the mixtures were analysed by electrophoresis (Figure 5B). Transcription towards *pyrE* is initiated at two positions in this region of the DNA. The more intense transcript starts with a G residue at position 338 of Figure 4, apparently directed by a promoter (P1) at the -10 region (TATAATG, position 327–333 in Figure 4) identical to the accepted sequence (Rosenberg and Court, 1979; Siebenlist *et al.*, 1980) of an *E. coli* promoter, but with no homology to the consensus sequence at its -35 region. The weaker transcript is initiated at position 266–268 of Figure 4 dictated apparently by a promoter (P2) that seems to lack the canonical T residue in its -10 region (TATAAAA, position 256–262 of Figure 4) and which also lacks homology to the consensus sequence of its -35 region. Judged from the intensities of the protected bands in Figure 5C the frequency of initiations at P1 and P2 is not controlled by the supply of pyrimidine to the cells.

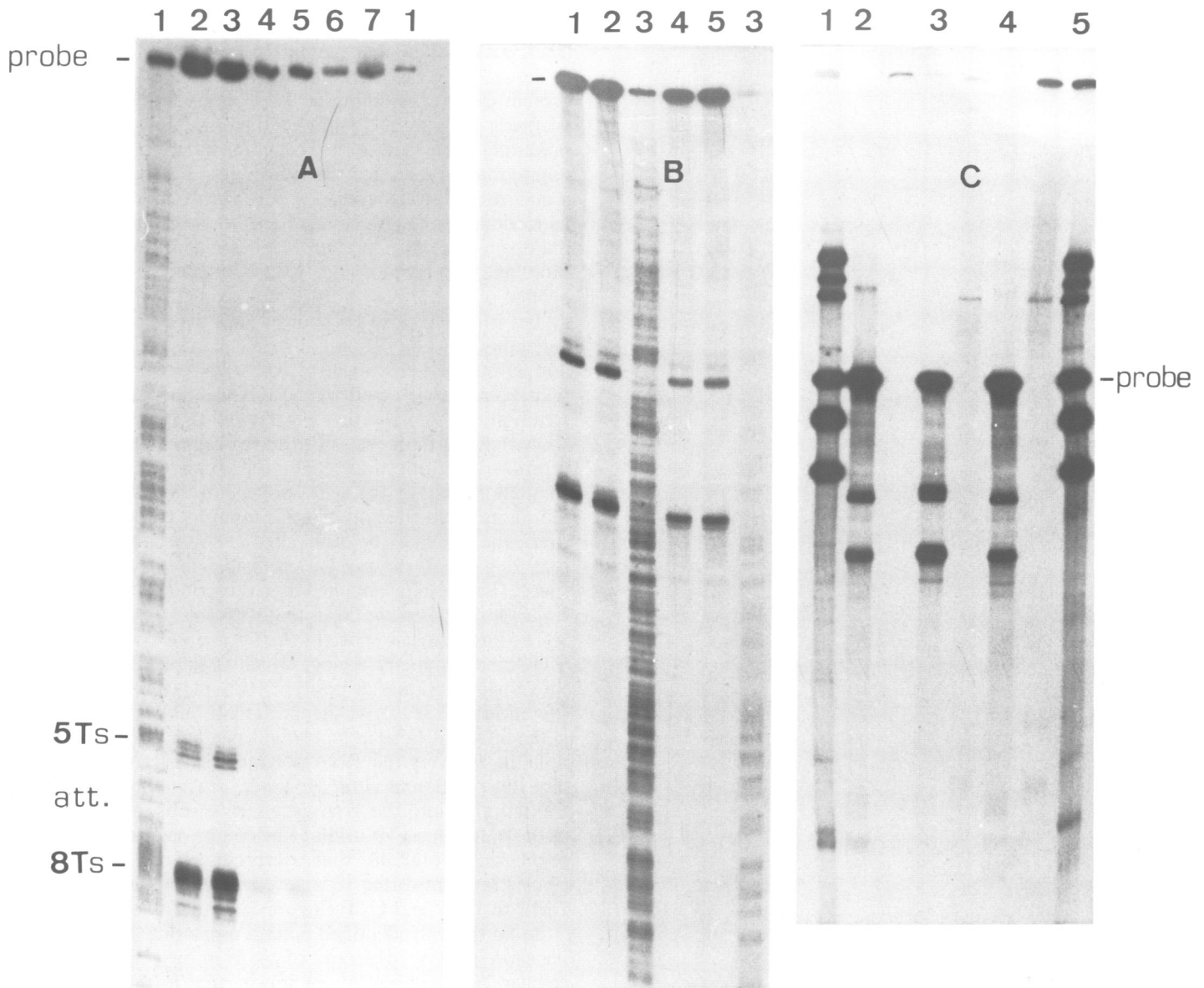


Fig. 5. Mapping of promoter activity in the *ORF-pyrE* operon by the S1-nuclease technique. (A) S1-nuclease digested hybrids between RNA isolated from the cultures in Table I with different levels of *pyrE* expression and the *MluI-BamHI** [5'-³²P]fragment covering the start of *pyrE*: the lanes marked 1 contain partially depurinated probe DNA; lanes 2 and 3, the probe DNA was hybridized to RNA from S ϕ 1240 (+pPP1) before treatment with S1 nuclease; lanes 4 and 5, the probe DNA was hybridized to RNA from S ϕ 1258 (+pPP1) grown without uracil; lanes 6 and 7, the DNA probe was hybridized to RNA from S ϕ 1258 (+pPP1) grown in the presence of uracil (20 μ g/ml). The samples in lanes 2, 4 and 6 were digested with S1-nuclease for 4 min at 30°C prior to electrophoresis in 6% polyacrylamide gels, while samples 3, 5 and 7 were digested for 8 min with S1-nuclease. We ascribe the bands at the positions marked '5Ts' and '8Ts' to double-stranded cleaving by S1 at the many AT base pairs flanking the *pyrE* attenuator (see Figure 4). (B) S1-digested hybrids between RNA from S ϕ 1240 (+pPP1) and the *PvuII-MluI** [5'-³²P]fragment covering the start of ORF. Lanes 1 and 4, the hybrids were digested with S1 for 4 min; lanes 2 and 5, the samples were treated with S1 for 12 min; lane 3 contain partially depurinated probe DNA. (C) The three RNA preparations used in A were hybridized to the 5'-³²P-end labeled *PvuII-MluI* fragment used as probe in B and digested with S1-nuclease for 4 min before electrophoresis parallel to ³²P-labeled DNA size markers (lanes 1 and 5). Lane 2, RNA from S ϕ 1240 (+pPP1); lane 3, RNA from S ϕ 1258 (+pPP1) grown without uracil; lane 4, RNA from S ϕ 1258 (+pPP1) grown in the presence of uracil (20 μ g/ml).

Termination of transcription in vitro in the space between the two cistrons of the ORF-pyrE operon

When the 2.0-kb *PvuII-PvuII* DNA fragment containing the ORF and *pyrE* genes was mixed with purified RNA polymerase and the nucleoside triphosphate substrates (ATP, GTP, UTP, and [α -³²P]CTP) one transcript of a discrete size and considerable intensity was seen after analysis by electrophoresis in urea-containing polyacrylamide gels (Figure 6, lanes 1, 3 and 4). The size of this predominant RNA species is

~780 nucleotides when compared with the migration of the denatured DNA fragments used as markers. When the *PvuII-BamHI* fragment was used as the template a similar pattern of transcription was observed (Figure 6, lane 5). On the other hand, the use of the *PvuII-MluI* fragment as template for RNA polymerase gives rise to the formation of a transcript 400–410 residues long (Figure 6, lane 6) corresponding nicely to the distance between the start at position 338 (Figure 4) dictated by promoter P1 and the *MluI* cut (Figure 1) in the middle of the ORF gene. Apparently, however, the *PvuII-*

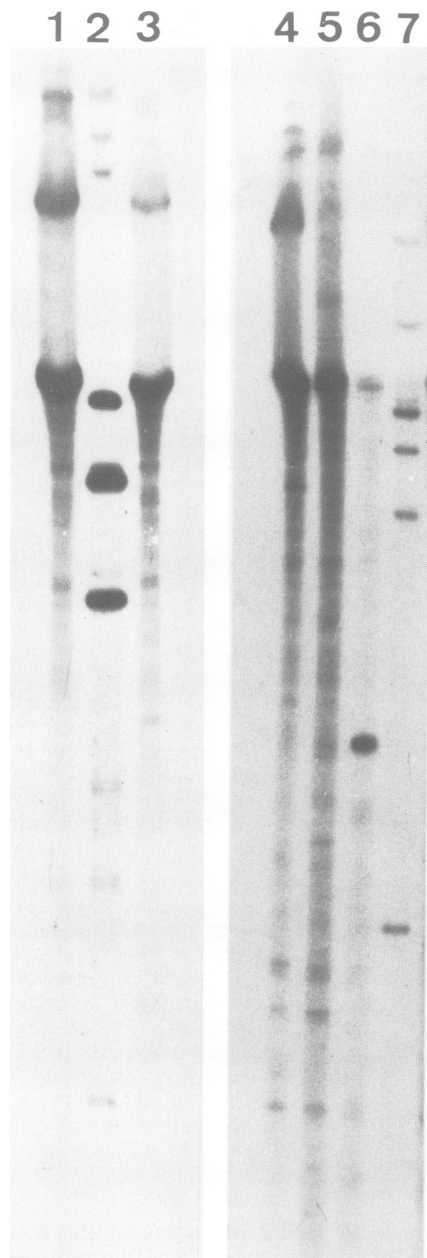


Fig. 6. *In vitro* transcription of isolated DNA fragments harboring parts of the *ORF-pyrE* operon. The reactions were composed as described in Materials and methods. **Lanes 1, 3 and 4**, transcripts from the 2.0-kb *PvuII-PvuII* fragment; **lane 5**, transcripts from the 1.35-kb *PvuII-BamHI* DNA fragment; **lane 6**, transcripts directed by the 0.75-kb *PvuII-MluI* DNA fragment. For the sample in **lane 1** the reaction was carried out for 6 min, while the reaction was for 3 min for the other samples. **Lane 2**, radioactive markers, sizes close to 510, 625, and 735 residues; **lane 7** markers ~335, 550, 650 and 735 residues.

MluI template also results in formation of small amounts of a transcript that corresponds to the full length of the DNA fragment (~760 bp). The reason for this behaviour is not known.

These results show that transcription is initiated *in vitro* from promoter P1 in front of the ORF gene, at position 338 in Figure 4, and is terminated some 780 residues downstream to this site, i.e., in the intercistronic space of the operon.

Discussion

These results show that the *pyrE* gene of *E. coli* is expressed as the second gene of an operon after the coding frame for a protein chain, 238 amino acid residues long. Thus, protein synthesis in 'minicells' showed that the DNA immediately preceding *pyrE* directs the formation of a polypeptide that migrates as a 30-kD protein in SDS gels. Furthermore, DNA sequence analysis revealed the existence of an ORF, 238 triplet codons long. The function of the encoded protein is unknown since no mutation has been identified at this position of the chromosome. Measurements of the level of orotate phosphoribosyltransferase directed by different plasmids derived from pPP1 showed that the DNA around the beginning of the ORF gene is needed for *pyrE* expression, and S1-nuclease treatment of RNA-DNA hybrids identified two promoters, P1 and P2, at this location, directing transcription of the operon. Furthermore, these experiments show that the frequency of mRNA chain initiations at these promoters in front of the ORF gene is independent of the pyrimidine supply of the cells, while, however, only a pyrimidine controlled fraction of the transcripts are continued into *pyrE*; the rest must be terminated at some point before this second gene of the operon, probably in the intercistronic space. The results obtained by transcribing isolated DNA fragments *in vitro* are consistent with this.

The intercistronic space of the *ORF-pyrE* operon contains a GC-rich symmetry region, which is flanked by blocks of thymidylate residues at position 1086–1118 in Figure 4. As the nucleotide sequence suggests (Platt, 1981), this structure apparently functions as a *rho*-independent transcription terminator and very few RNA polymerase molecules transcribe past this position *in vitro*. However, *in vivo* the frequency of transcription into the *pyrE* gene seems to be inversely correlated to the supply of uridine nucleotides as evidenced by the *pyrH* strain, SØ1240, that overproduces orotate phosphoribosyltransferase due to a low endogenous conversion of UMP to UDP and UTP (Figure 5A and Table I).

So far the only identified mutations causing high constitutive expression of *pyrE* (and of *pyrB*) in the presence of high cellular UTP concentrations are located in the RNA polymerase genes *rpoBC* (Jensen *et al.*, 1982), and the purified RNA polymerase from such a mutant has been found to harbor a K_m defect for UTP in the transcription of T7 DNA and synthetic DNA templates (K.F.Jensen, unpublished observations). This indicates that the regulatory signal in the control of *pyrE* (and *pyrB*) expression involves the degree of saturation of the elongating RNA polymerase with its UTP substrate (Jensen *et al.*, 1982). Together with the present results, this indicates that the regulation of *pyrE* expression occurs by an RNA polymerase-modulated attenuation in the intercistronic space of the *ORF-pyrE* operon in such a way that the RNA chain termination is rare, when RNA polymerase transcribes slowly due to shortage of UTP, while the termination frequency is high, when RNA polymerase is not starved of UTP during elongation.

The regulatory region of the *pyrB* gene also contains a *rho*-independent transcription terminator located between the promoter and the beginning of the structural gene, within an ORF for a hypothetical leader peptide of 44 codons (Roof *et al.*, 1982; Navre and Schachman, 1983; Turnbough *et al.*, 1983). Since the degree of transcription termination at this attenuator *in vitro* is independent of the UTP concentration

(Turnbough *et al.*, 1983) it has been suggested that the degree of attenuation is controlled by UTP-induced variations in the tightness of coupling between the elongating RNA polymerase and the leading ribosome translating the message for the leader peptide (Turnbough, 1983; Turnbough *et al.*, 1983).

In the *ORF-pyrE* operon the coding frame of the ORF gene ends just prior to the block of five thymidylate residues in front of the GC-rich stem of the *pyrE* attenuator (Figure 4), and it is tempting to speculate that the end of the ORF gene may act as a probe for the tightness of coupling between transcription and translation in the modulation of *pyrE* attenuation, similar to that proposed for the coding frame for the leader peptide of the *pyrB* gene. We propose that termination *in vivo* is suppressed by translation at the end of the ORF gene if the coupling of this translation to the elongating RNA polymerase is very tight, while the termination frequency is increased if the coupling is less tight. Thus, when RNA polymerase is starved of UTP and transcribes more slowly than when the UTP concentration is high, this slow transcription rate may increase the tightness of coupling and thereby reduce the frequency of termination at the *pyrE* attenuator. Indeed we have found that translation at the carboxy-terminal end of the ORF gene is an absolute prerequisite *in vivo* for expression of *pyrE* in a *cis* position even when the cellular UTP pool is low (Bonekamp, Clemmesen, Karlström and Jensen, in preparation).

Recently, a mutant was isolated which exhibited a pyrimidine requirement due to a mutation (*pyrS*) outside any of the known *pyr* genes. The pyrimidine requirement was found to be caused by a very low level of expression of *pyrE* and *pyrF* resulting from the mutation in *pyrS* (Nowlan and Kantrowitz, 1983). In the light of our results the (unknown) product of the *pyrS* gene may either be activating transcription from the *ORF-pyrE* promoters or participating in the process of modulation of attenuation in the intercistronic space of the operon.

Materials and methods

Bacterial strains

Strain S ϕ 1258 (*pyrE*) lacking a functional *pyrE* gene was made from strain MC4100 (*araD139*, Δ (*lac*)U169, *rpsL*, *thi*) by insertion of Mu *cts62* into *pyrE* and subsequently selecting for ability to form colonies at 42°C, as previously described (Jensen *et al.*, 1984). The strain is not entirely deleted for its *pyrE* gene. Rather, it harbors a mutation in the lysogenic Mu phage, as it releases plaque-forming phage particles at a low frequency. Strain SO1240 (*metB*, *purE*, *pyrH*⁶) was a gift from Per Nygaard. The 'minicell'-producing strain BD1854 (*minA*, *minB*) has previously been used (Jensen *et al.*, 1984) and was a gift from Borge Diderichsen.

Cell growth

The cells were grown at 37°C in the (A+B) medium of Clark and Maaløe (1967) supplemented with casamino acids (0.2%), glucose (0.2%), and thiamine (1 μ g/ml). Sometimes uracil (20 μ g/ml) or hypoxanthine (30 μ g/ml) were present. Growth of the cells was monitored at 436 nm in an Eppendorf photometer.

Orotate phosphoribosyltransferase

The activity of this enzyme was measured by a spectrophotometric monitoring of the absorbance at 295 nm resulting from the conversion of orotate to OMP, as described by Poulsen *et al.* (1983). One unit is defined as the activity that converts 1 μ mol of orotate to OMP per min at the prescribed conditions.

Protein synthesis in 'minicells'

The proteins made by 'minicells' containing different plasmids were labeled with [³⁵S]methionine and the extracts were analysed by electrophoresis in 12.5% polyacrylamide gels containing SDS and run in parallel with non-radioactive markers, as described previously (Jensen *et al.*, 1984).

DNA sequence determinations

The DNA sequencing was carried out according to the procedure of Maxam and Gilbert (1977, 1980). The A/G, G, C/T, and C specific cleavages were modified as described by Poulsen *et al.* (1983).

S1-nuclease mapping of transcriptional initiation

The isolation of RNA for hybridization was done as described by Valentin-Hansen *et al.* (1982). For hybridization, 150 μ g RNA and ~0.1 μ g of the 5'-³²P-labeled DNA probe fragment was mixed, denatured, and allowed to reanneal by slow cooling to 40°C, overnight. The mixture was then treated with S1-nuclease (Boehringer, 2000 units), and the pattern of protected DNA fragments was analysed by electrophoresis in 8 M urea containing polyacrylamide (6%) gels (Berk and Sharp, 1978; Manley *et al.*, 1979; Valentin-Hansen *et al.*, 1982).

Transcription of DNA fragments in vitro

After digestion of plasmid pPP2 with restriction endonucleases the desired DNA fragments were separated by electrophoresis in polyacrylamide gels. The DNA bands were detected by u.v. absorbance and the proper gel piece was cut out. Following electroelution from the gel, the DNA was concentrated using ethanol precipitation onto small columns of cellulose (Jensen *et al.*, 1976). To perform a transcription of the DNA, 0.5–1 μ g of the fragment was mixed with 5 μ g of purified RNA polymerase (Borges and Jendrzak, 1975) from *E. coli* K-12, 0.4 mM each of ATP, GTP, UTP and of 0.1 mM [α -³²P]CTP (~500 Ci/mol) in a buffer consisting of 44 mM Tris-HCl pH 8.0, 20 mM NaCl, 28 mM 2-mercaptoethanol, 14 mM MgCl₂ at 37°C in a total volume of 50 μ l. The reaction was started by adding RNA polymerase to the pre-warmed mixture of the other components. At 3 min and 6 min, 25 μ l samples were pipetted into 2.5 μ l sodium acetate plus 50 μ l 95% ethanol and left to precipitate at -20°C overnight. Following centrifugation the pellet was dissolved in 25 μ l of (7.5% glycerol, 0.025% SDS, 0.025% Bromophenol blue, 25 mM EDTA pH 7, 0.2 mg/ml of sonicated T7 DNA) (Chamberlin *et al.*, 1979) and mixed with 25 μ l 8 M urea-0.1% xylene cyanol. Then 1–2 μ l samples were analysed by electrophoresis in 6% polyacrylamide gels (60 cm x 20 cm x 0.03 cm) containing 8 M urea in 0.8 times the buffer of Peacock and Dingman (1968). The gel was run at 50 W constant power for various times and subjected to autoradiography.

Construction of plasmids

The construction of plasmid pPP1 has been described previously (Poulsen *et al.*, 1983). Plasmid pPP2, which contains the 2.0-kb fragment cloned in opposite orientation relative to the vector, was obtained in the same selection. Plasmid pPP4 was made by cutting pPP1 with the restriction endonuclease *Clai*, transforming with the linear fragments, and picking ampicillin-resistant, tetracycline-sensitive transformants. Plasmid pPP5 was made in a similar way after cutting with *SphI*. Plasmid pPP6 was made from pPP2 as a *Bam*HI-generated deletion derivative. Plasmid pFB1 was constructed from pPP1 by opening with *Hind*III, digesting with S1-nuclease to remove the 5' overhanging unpaired nucleotides, and resealing with T4 DNA ligase. This treatment destroys the *Tet*^R promoter (Rodrigues *et al.*, 1979) and, after transformation, ampicillin-resistant, tetracycline-sensitive colonies were picked. pFB2 was now made by deleting the *SphI-SphI* fragment from pFB1.

Acknowledgements

We wish to thank L.G.Lundberg for contributing to some sequence determinations, O.Karlström for making computer analyses of DNA sequences and for helpful discussions, K.G.Welinder for calculating hydropatic profiles, R. Fast for teaching us how to purify RNA polymerase, P.Valentin-Hansen for introducing us to the S1-nuclease technique, and A.Munch-Petersen for reading the manuscript. The work was supported by grants from the Danish Natural Science Council and from the Carlsberg Foundation by a Scholar stipendium to F.B.

References

- Bachman, B.J. and Brooks Low, K. (1980) *Microbiol. Rev.*, **44**, 1-56.
- Berk, A.J. and Sharp, P.A. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 1274-1278.
- Beyreuther, K., Bieseler, B., Ehring, R., Griesser, H.W., Miesehendahl, M., Müller-Hill, B. and Triesch, I. (1980) *Biochem. Soc. Trans.*, **8**, 675-676.
- Borges, R.R. and Jendrzak, J.J. (1975) *Biochemistry (Wash.)*, **14**, 4634-4638.
- Chamberlin, M.J., Nierman, W.C., Wiggs, J. and Neff, N. (1979) *J. Biol. Chem.*, **254**, 10061-10069.
- Clark, D.J. and Maaløe, O. (1967) *J. Mol. Biol.*, **23**, 99-112.
- Grosjean, H. and Fiers, W. (1982) *Gene*, **18**, 199-209.
- Hoover, T.A., Roof, W.D., Foltermann, K.F., O'Donovan, G.A., Bencini, D.A. and Wild, J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2462-2466.
- Jensen, K.F. (1979) *J. Bacteriol.*, **138**, 731-738.

- Jensen,K.F., Nes,I.F. and Wells,R.D. (1976) *Nucleic Acids Res.*, **3**, 3143-3155.
- Jensen,K.F., Neuhard,J. and Schack,L. (1982) *EMBO J.*, **1**, 69-74.
- Jensen,K.F., Larsen,J.N., Schack,L. and Sivertsen,A. (1984) *Eur. J. Biochem.*, **140**, 343-352.
- Kelln,R.A., Kinahan,J.J., Foltermann,K.F. and O'Donovan,G.A. (1975) *J. Bacteriol.*, **124**, 764-774.
- Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105-132.
- Manley,J.L., Sharp,P.A. and Gefter,M.L. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 160-164.
- Maxam,A. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.
- Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-580.
- Navre,M. and Schachman,H.K. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1207-1211.
- Nowlan,S.F. and Kantrowitz,E.R. (1983) *Mol. Gen. Genet.*, **192**, 264-271.
- Peacock,A.C. and Dingman,C.W. (1968) *Biochemistry (Wash.)*, **7**, 668-674.
- Pierard,A.N., Glansdorff,N., Gigot,D., Crabeel,M., Halleux,P. and Thiry,L. (1976) *J. Bacteriol.*, **127**, 291-301.
- Platt,T. (1981) *Cell*, **24**, 10-23.
- Pollitt,S. and Zalkin,H. (1983) *J. Bacteriol.*, **153**, 27-32.
- Poulsen,P., Jensen,K.F., Valentin-Hansen,P., Carlsson,P. and Lundberg,L.G. (1983) *Eur. J. Biochem.*, **135**, 223-229.
- Rodrigues,R.L., West,R.W., Heyneker,H.L., Bolivar,F. and Boyer,H.W. (1979) *Nucleic Acids Res.*, **6**, 3267-3287.
- Roof,W.D., Foltermann,K.F. and Wild,J. (1982) *Mol. Gen. Genet.*, **187**, 391-400.
- Rosenberg,M. and Court,D. (1979) *Annu. Rev. Genet.*, **13**, 319-353.
- Schwartz,M. and Neuhard,J. (1975) *J. Bacteriol.*, **121**, 814-822.
- Siebenlist,U., Simpson,R.B. and Gilbert,W. (1980) *Cell*, **20**, 269-281.
- Turnbough,C.L., Jr. (1983) *J. Bacteriol.*, **153**, 998-1007.
- Turnbough,C.L., Jr., Hicks,H.L. and Donahue,J.P. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 368-372.
- Valentin-Hansen,P., Aiba,H. and Schümperli,D. (1982) *EMBO J.*, **1**, 317-322.

Received on 11 April 1984; revised on 11 May 1984