Volume 11 Number 18 1983 Nucleic Acids Research

Nucleotide sequence of the promoter region of the *E. coli lys*C gene

Michel Cassan, Jeanine Ronceray and Jean Claude Patte

Institut de Microbiologie, Université Paris-Sud, Centre d'Orsay, 91405 Orsay Cedex, France

Received 2 August 1983; Accepted 5 September 1983

ABSTRACT

The regulatory region of the lysC gene (that encodes the lysine-sensitive aspartokinase of Escherichia coli) has been identified and purified by the use of lysC-lacZ fusions. Its regulatory sequence has been determined. No signals similar to those described in the case of an attenuation mechanism could be found in the long leader sequence existing between the starts of transcription and of translation.

INTRODUCTION

In Escherichia coli, the lysine biosynthetic pathway is composed of nine different steps. It has been shown that expression of most of the corresponding genes is repressed by lysine (1) and that these "lysine" genes are scattered along the chromosome (2). Our laboratory is currently involved in obtaining the nucleotidic sequences of the different regulatory regions located upstream the lysine genes ; this would help for determining the molecular mechanism of regulation, either by an attenuation mechanism (as in most of the amino acid biosynthetic pathways, ref. 3) or by the involvement of one (or several) regulatory molecule(s) that could share similar target sites on DNA (as in the case of argR, ref. 4). The lysC gene encodes the lysine-sensitive aspartokinase (AKIII), one of the isofunctional enzymes catalyzing the first step of the pathway. Lysine inhibits the activity of the enzyme and represses its synthesis (5). Mutants altered in this regulatory pattern have been isolated, leading either to a loss of feed-back inhibition (desensitized mutants, ref. 6) or to constitutivity of expression (7, 8). All the mutants of this last class are cis-dominant ; no trans-dominant mutation could be isolated, that would have affected a regulatory protein. The possible existence of an attenuation mechanism was thus expected.

This paper describes the nucleotidic sequence of the lysC regulatory region.

MATERIAL AND METHODS

Strains

RM4102 : F⁻ araD139 lacU169 rpsL thiA (9)

JM101 : Δ(lac-pro) supE thiA / F' traD36 proAB laci^{q}ZΔM15 (10)

PAL40 : RM4102, Φ(lysC-lacZ)1140 (11)

Enzymes assays

β-galactosidase was assayed according to Miller (12) the specific activity being calculated as nmole ONP produced/min.

β-lactamase specific activity was measured as described in Chenais et al. (13).

DNA manipulation

- Plasmids : large scale purification of plasmid DNA was made according to cleared lysate method (14) after chloramphenicol amplification (15). Rapid analysis of recombinant plasmids was performed after alkaline lysis procedure (16).
- DNA probes : fragments were 5' labelled at BamHI or BalI sites by the kinase exchange reaction as described by Beckner and Falk (17) and separated, after secondary cut, on polyacrylamid gels and electroeluted.

DNA sequencing

The chemical technique (Maxam and Gilbert, ref. 8) was performed through four reactions with the same minor modifications as Clement and Hofnung (19).

Mapping of the 5'-ends of transcripts

-S1 mapping : mapping of the transcription start point with S1 endonuclease was performed as described by Barry et al. (20) with the following modifications : the DNA probes were used in hybridization with total RNA from strain RM4102/pCM9420 during 3 hrs at 35°C in 80 % formamide. Nonhybridized DNA was digested by 5,000 u S1 nuclease (Boehringer) for 30 min at 45°C. The protected hybrids fragments were submitted to electrophoresis on 7 % or 20 % sequencing gels in parallel with the products of sequencing reactions.

- Reverse transcriptase : the procedure followed was the same that in Débarbouillé and Raibaud (21) with the following modifications : reverse transcriptase was incubated 1 h at 39°C ; 1 µg of RNase was added and further incubated 30 min at 37°C. After precipitation with ethanol the nucleic acids were analysed as described in S1 mapping experiments.

RESULTS

1) Purification of the lysC promoter region

The DNA of the lysC promoter region was purified from a λ transducing phage carrying a lysC-lacZ fusion (pal40) previously isolated in our labora-

tory (11) according to the in vivo method of Casadaban (22) ; in such a strain expression of the lacZ gene that has lost its own promoter is dependent upon the functioning of the lysC promoter. A crude restriction map of this phage (see Fig. 1) allowed us to postulate the position of two known EcoRI sites, one on the left of the SmaI site in the exo gene region of λ (23) and the other at the distal end of lacZ (23). Due to the mechanism of lacZ expression by the gene fusion the internal EcoRI fragment must carry the lysC promoter region, the direction of transcription being from left to right. This fragment was cloned in the EcoRI site of pBR322 (plasmid pCM4, Fig. 1), identifying Lac^+ clones among Amp^R transformants of strain JM101 (that harbours a F'lacZΔM15). As the uncomplete lacZ cloned region does not allow the expression of an active β-galactosidase but contains all the Z(α) sequence, Lac^+ clones would result from complementation of this Z(α) fragment and the Z(ω) polypeptide

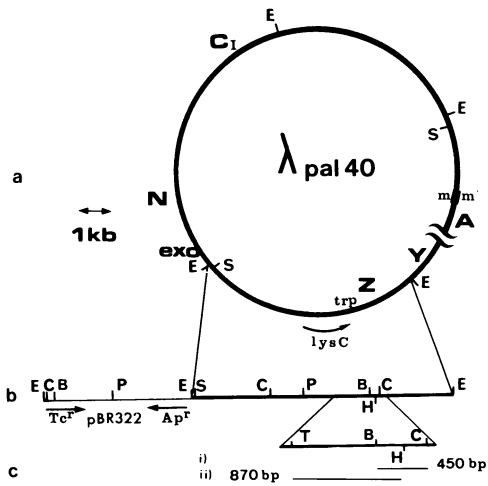

Figure 1 : Phage and plasmids used for sequencing the lysC promoter region.
a) EcoRI and SmaI physical map of λpal40 (λtransducing phage carrying a lysC-lacZ fusion isolated from strain PAL40 as described in Cassan et al., ref.34). Some λ(exo,N,CI) and bacterial (trpA, B ; lacZ, Y) genes are indicated. The arrow gives the presumed position of the lysC regulatory region (see text).
b) plasmid pCM4 : the 7.7 kb EcoRI fragment of λpal40 was cloned in the EcoRI site of pBR322 (heavy line : bacterial and λDNA)
c) enlargement of part of plasmid pCM4. i) : the 450 bp BamH1-Cla1 fragment from pCM4 was inserted between the BamH1 and Cla1 sites of pBR322 to give pCM402 ; ii) : the 870 bp Taq1-Hpa1 fragment of pCM4 was cloned from pCM4 using the Cla1 and PvuII sites of pBR322 to give pCM420 (see Fig.2).
The abbreviations for restriction sites are as follows : B, BamH1 ; C, Cla1 ; E, EcoR1 ; H, Hpa1 ; P, PvuII ; S, Sma1 ; T, Taq1.

TABLE I

Strain	β-galactosidase specific activity after growth in		lacZ gene copy number per chromosome
	minimal medium	40 mM L-lysine	
PAL 40	200	50	1
JM101/pCM4	32 (a)	10 (a)	60 (b)
RM4102/pCM9420	22100	10200	210 (b)

Regulation of β-galactosidase synthesis in different strains . β-galactosidase specific activities were determined on cell-free crude extracts. Values are the mean of several independent experiments. For growth conditions, see ref. 13.
(a) : in this case, β-galactosidase activity results from complementation between Z (α) and Z (ω) peptides ; see text ;
(b) : approximate values of the plasmid copy number determined as in ref. 13.

synthetized by the ZΔM15 gene (24). Measurements of β-galactosidase activity were performed on cell-free crude extracts of strain JM101/pCM4 grown with IPTG (in order to induce the synthesis of Z (ω) that must be in excess) and in the presence or in the absence of lysine. Values, given in Table I, indicate that a repressive effect of lysine is indeed obtained, confirming the presence of the lysC promoter fused to the Z(α) coding region.

A more precise restriction map of plasmid pCM4 was performed (Fig. 1) that led to the following conclusions : i) on the left of the inserted fragment, the existence of the characteristic EcoRI-SmaI fragment (23) confirmed the

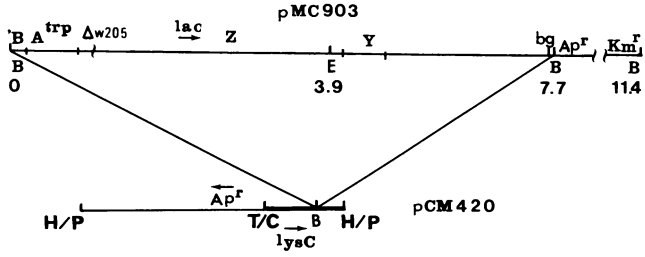

Figure 2 : Construction of plasmid pCM9420.
Insertion of the BamH1 fragment of pMC903 in the BamH1 site of pCM420 (see fig.1). The arrows indicate the polarity of transcription (presumed for lysC, see text). Numbers give the distance (kb) from the left BamH1 site. For abbreviations, see Fig.1 ; bg is used for BglII.

location of the λ exo gene ; on the contrary the ClaI and PvuII sites do not exist in λ DNA in this region and must belong either to bacterial or to Mu DNA, as resulting from illegitimate excision of the λ phage from the parental lysC-lacZ fusion PAL40 (as detailed by Casadaban, 22). The existence of Mu DNA in this ClaI-PvuII fragment was confirmed by blot hybridization (data not shown); ii) in the right part, in front of the lacZ gene, the ClaI and HpaI sites were identified as those of the trpB DNA existing in the parental λP1(209) phage used for the construction of the fusion (Casadaban, 22). This is not the case for the BamHI site, that must belong to lysC (or Mu) DNA. The event of deletion of part of the Mu genome leading to the gene fusion was thus located in this region. So the bacterial insertion was limited between the ClaI site (on the left) and the HpaI site (Fig. 1). We decided to study a 870 bp TaqI-HpaI fragment upstream the fusion joint, that was cloned on pBR322 (plasmid pCM420, Fig. 1). In order to ensure the entire lysC promoter was indeed contained in this fragment, we inserted downstream the BamHI site of the bacterial insert of pCM420 a BamHI fragment taken from plasmid pMC903 (25), that carries an entire lacZ coding sequence without the lac promoter (Fig. 2). Among Amp^R Lac^- transformants of strain RM4102, Lac^+ clones were identified. Restriction analysis of the plasmid present in these strains indicated the lacZ gene was inserted in the correct orientation (deduced from the direction of lysC transcription as explained for pCM4). A promoter had to be present in front of the lacZ gene in this plasmid, called pCM9420. Regulation of expression of the lacZ gene was studied in strain RM4102 harbouring pCM9420, grown in the presence or in the absence of lysine. Results are given in Table I. A decrease of β-galactosidase specific activity is observed when growth was effected in the presence of lysine, indicating the presence of the lysC regulatory signals in front of lacZ. As this coding sequence is present in multicopies β-galactosidase expression is largely enhanced (when compared to the parental fusion strain), being compatible with a gene dosage effect. However one has to note that the extent of regulation (minimal level / repressed level) is lower (around 2) in the case of the plasmid than of the fusion (around 4). Two explanations are possible to explain this difference : i) some readthrough transcription, not due to the lysC promoter ; a background level of expression is indeed obtained with the parental plasmid pMC903 ; ii) or a titration of a regulatory molecule in limiting amount, by increasing its target number.

2) Nucleotidic sequence

The determination of the nucleotidic sequence was thus undertaken on the TaqI-ClaI fragment of 1,050 bp. The DNA fragments used were isolated from

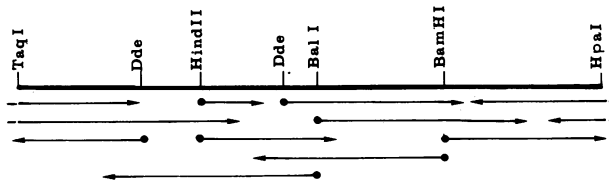

Figure 3 : Sequencing strategy of the lysC promoter region.
Relevant restriction sites are given. The arrows indicate the direction of sequencing and the length of sequence determined. 5'-labelled ends are indicated by dots.

plasmids pCM420 and pCM402 (Fig. 1). The strategy is summarized in Fig. 3. The DNA sequence was determined on both strands, except for the region coding for part of the trpB gene, that has been already published (26). At position 585 (Fig. 4) an ATGTCT sequence was found that agrees well with the N terminal Serine of AKIII determined previously (27) ; this was confirmed by unpublished

```
        10          20          30          40          50          60
TCGAAGTGTT TCTGTAGTGC CTGCCAGGCA GCGGTCTGCG TTGGATTGAT GTTTTTCATT
        70          80          90         100         110         120
AGCAATACTC TTCTGATTTT GAGAATTGTG ACTTTGGAAG ATTGTAGCGC CAGTCACAGA
       130         140         150         160         170         180
AAAATGTGAT GGTTTTAGTG CCGTTAGCGT AATGTTGAGT GTAAACCCTT AGCGCAGTGA
       190         200         210         220         230         240
AGCATTTATT AGCTGAACTA CTGACCGCCA GGAGTGGATG AAAAATCCGC ATGACCCCAT
       250         260         270         280         290         300
CGTTGACAAC CGCCCCGCTC ACCCTTTATT TATAAATGTA CTACCTGCGC TAGCGCAGGC
       310         320         330         340         350         360
CAGAAGAGGC GCGTTGCCCA AGTAACGGTG TTGGAGGAGC CAGTCCTGTG ATAACACCTG
       370         380         390         400         410         420
AGGGGGTGCA TCGCCGAGGT GATTGAACGG CTGGCCACGT TCATCATCGG CTACAGGGGC
       430         440         450         460         470         480
TGAATCCCCT GGGTTGTCAC CAGAAGCGTT CGCAGTCGGG CGTTTCGCAA GTGGTGGAGC
       490         500         510         520         530         540
ACTTCTGGGT GAAAATAGTA GCGAAGTATC GCTCTGCGCC CACCCGTCTT CCGCTCTTCC
       550         560         570         580         590         600
CTTGTGCCAA GGCTGAAAAT GGATCCCCTG ACACGAGGTA GATTATGTCT GAAATTGTTG
                                                  MetSer GluIleVal
       610         620         630         640
TCTCCAAATT TGGCGGTACC AGCGCTAGCG ATTTTGACGC CGAAG
ValSerLysPhe
```

Figure 4 : Nucleotide sequence of the lysC promoter region.
The nucleotide sequence of the non-coding strand is given. Numerotation starts at the first letter of the TaqI site. The amino acids indicated are deduced from the codon rule. -35, -10 and Shine-Dalgarno sequences are underlined.The wavy arrow indicates the beginning of mRNA (see Fig.6). Palindromes are indicated by arrows and dots.

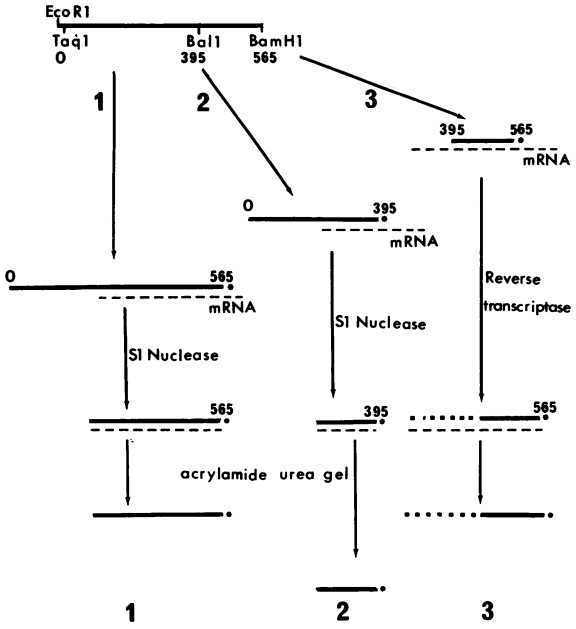

Figure 5 : Strategy used for S1 nuclear digestion and reverse transcriptase extension mapping experiments.
The DNA probes used are : 1 : the 565 bp EcoRI-BamH1 fragment ; 2 : the 395 bp EcoRI-Ball fragment ; 3 : the 170 bp Ball-BamH1 fragment. Numbers give the distance (bp) from the EcoRI site. * indicate the 5'labelled sites. Interrupted lines represent an (hypothetical) molecule of lysC mRNA. Dotted lines indicate DNA synthesized by reverse transcriptase.
Total RNA was extracted as indicated in Material and Methods from strain RM4102/pMC9420 grown in minimal medium plus Casamino acid lysine-free (Difco).

results of M. Bruschi (pers. comm.) showing the N terminal sequence of AKIII to be Ser Glu Ile Val Val Ser Lys Phe. Although the reading frame remains open for 34 nucleotides after Phe, up to the junction with trpB, it is not known if these nucleotides belong to the lysC gene or to Mu DNA.

This ATG is preceded by a convenient Shine-Dalgarno (28) sequence GAGGT, confirming this position as the translational start.

The point of initiation of transcription was determined using S1 nuclease mapping and reverse transcriptase primer extension experiments. Total mRNA extracted from RM4102/pCM9420 strain was hybridized with the following DNA fragments (Fig. 5) : i) for S1 mapping, a BamHI-EcoRI fragment and a BalI-EcoRI fragment, labelled by ^{32}P at the 5' ends of the BamHI and BalI sites respectively ; ii) for the reverse transcriptase experiments, a BalI-BamHI fragment, ^{32}P labelled at the 5' end of the BamHI site.

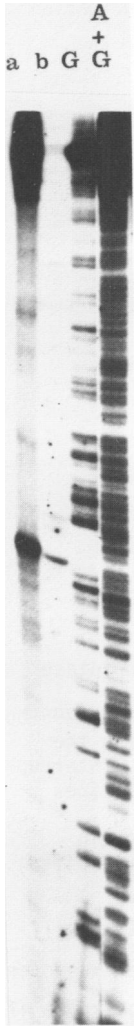

Figure 6 : Position of the 5' end of the lysC transcripts. Results of experiments 1 (S1 nuclease digestion) and 3 (reverse transcriptase extension) of Figure 5 are shown on lanes a and b respectively. Results of experiment 2, run on another gel, are not given. The labelled fragments were loaded in parallel with the products of the G and A+G reactions.

The 5' initiation point of the mRNA was thus localized at nucleotide 278 (Fig. 6). It is preceded by a convenient and well-distanced -35 sequence GTTGACA starting at position 242, with several acceptable Pribnow boxes (29) around position 270. One has to note that the interconsensus region is very unusually GC rich (13/16). That might account for a weak expression of this promoter, as supposed from the relatively low β-galactosidase activity expressed from the fusion (Table I).

Two other experiments (data not shown) has confirmed the position of initiation of transcription : two plasmids were constructed, containing either a

Figure 7 : Comparison between lysC and asd regulatory regions. Numerotations begin at the transcriptional starts (lysC : this paper ; asd : ref. 33 and C. Haziza, personnal comm.). Dots indicate the nucleotides found at similar positions in both sequences.

deletion upstream the BalI site, or a modification of the HindII site (position 245) by fusion to the PvuII site of pBR322 (the remaining lysC sequence being fused to lacZ). In both cases an important decrease in β-galactosidase activity was observed. This confirms the role of the -35 sequence located in the HindII site.

These results reveal the existence of a quite long leader region, the translational start being 308 bp far from the 5' end of the mRNA. Examination of this unusually long leader sequence (longer than the ones till now described, as found in araC and aroH, ref. 30 and 31) does not display any of the characteristics of an attenuation-like structure (3) : no typical transcriptional stop could be found even if some possible secondary structures appear (see Fig. 4) ; in any of the various short open reading frames found exist several adjacent Lys codons. The role of this leader is yet unknown.

Comparison with other lysine regulatory sequences

Computer analysis between lysC and asd sequences have revealed a noteworthy homology that lies approximately at the same distance from the transcription initiation site, upstream the -35 region (Fig. 7). A possible role of this sequence as a target for a putative regulatory molecule common to these two genes has to be supported by in vitro mutagenesis that could affect the regulatory pattern.

Till now, except for lysA where a positive regulation has been demonstrated (32), no molecular mechanism can be evoked by examination of the nucleotidic sequence of any of the lysine genes we have studied (asd, ref. 33 ; lysC, this paper ; dapB and dapD, unpublished results of this laboratory). We are currently cloning and sequencing the cis-dominant mutations previously isolated in the lysC gene (7, 8) in order to get indications on the mechanism by which lysine repression is established.

ACKNOWLEDGMENTS

We wish to thank J.M. Clement, M. Hofnung, C. Jacq and P. Stragier for

their advices and their interest during this work, and G. Bogdanoff for typing the manuscript. This work was supported by grants from the Centre National de la Recherche Scientifique (L.A. 136), the Delegation à la Recherche Scientifique et Technique and the Fondation pour la Recherche Médicale Française.

REFERENCES

1. Patte, JC (1983) in "Amino-Acid Biosynthesis and Genetic Regulation" Somerville & Herrmann Eds, in press.
2. Theze, J., Margarita, D., Cohen, G.N., Borne, F., Patte, JC (1974) J. Bacteriol. 117, 133-143.
3. Yanofsky, C. (1981) Nature 289, 751-758.
4. Cunin, R., Eckaardt, T., Piette, J., Boyen, A., Pierard, A. & Glandsdorff, N. Nucleic Acids Res., submitted.
5. Stadtman, E., Cohen, G.N., Le Bras, G. & Robichon-Szulmajster, H. (1961) J. Biol. Chem., 236, 2033-2038.
6. Boy, E., Patte, JC (1972) J. Bacteriol., 112, 84-92.
7. Cassan, M., Boy, E., Borne, F., Patte, JC (1975) J. Bacteriol, 123, 391-399.
8. Boy, E., Borne, F., Patte, JC (1979) Biochimie, 61, 1151-1160.
9. Patte, JC, Morand, P., Boy, E., Richaud, C., Borne, F. (1980) Mol. Gen. Genet., 179, 319-325.
10. Messing, J. (1979) Recombinant DNA Technical Bulletin 2, 43-48.
11. Richaud, F., Phuc, NH, Cassan, M., Patte, JC (1980) J. Bacteriol. 143, 1, 513-515.
12. Miller, J. (1972) in Experiments in Molecular Genetics Cold Spring Harbor Laboratory, N.Y.
13. Chenais, J., Richaud, C., Ronceray, J., Cherest, H., Surdin-Kerjan, Y., Patte, JC (1981) Mol. Gen. Genet., 182, 456-461.
14. Humphreys, G.O., Wilshaw, G.A., Anderson, E.S. (1975) Biochem. Biophys. Acta, 383, 457-463.
15. Clewell, D.B. (1972) J. Bacteriol. 110, 667-676.
16. Birnboim,H.C., Doly, J. (1979) Nucleic Acids Res., 7, 1515-1523.
17. Beckner, K.C., Falk, W.R. (1977) J. Biol. Chem., 252, 3176-3184.
18. Maxam,A.M., Gilbert, W. (1980) in Methods in Enzymology. Grossman C. & Moldave K. Eds. vol. 65, Academic Press, N.Y., 499-560.
19. Clement, JM, Hofnung, M. (1982) Cell, 27, 507-514.
20. Barry, G., Squires, C.L. (1980) Proc. Natl. Acad. Sci. U.S.A., 77, 3331-3335.
21. Débarbouillé, M., Raibaud, O. (1983) J. Bacteriol., 153, 1221-1227.
22. Casadaban, M. (1976) J. Mol. Biol., 104, 541-555.
23. Maniatis, T., Fritsch, E.F., Sambrook, J. in Molecular cloning a laboratory manual, Cold Spring Harbor, Lab., Cold Spring Harbor N.Y.
24. Ullman, A., Perrin, D., (1970) in Lactose Operon, Beckwith J.R. & Zipser, D. Eds. Cold Spring Harbor Lab. N.Y.
25. Casadaban, M., Chou, J., Cohen, S.N. (1980) J. Bacteriol. 143, 971-980.
26. Crawford, I.P., Nichols, B.P., Yanofsky, C. (1980) J. Mol. Biol., 142, 489-502.
27. Richaud, C., Mazat, JC, Gros, O., Patte, JC (1973) Europ. J. Biochem., 40, 619-629.
28. Shine, J., Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A., 71, 1342-1346.
29. Rosenberg, M., Court, D. (1979) Ann. Rev. Genet., 13, 319-353.
30. Wallace, R.G., Lee, N., Fawler, A.V. (1980) Gene, 12, 179-190.
31. Zurawski, G., Gunsalys, R.P., Brown, K.D., Yanofsky, C. (1981) J. Mol. Biol., 145, 47-73.
32. Stragier, P., Patte, JC (1983) J. Mol. Biol., 167, in press.
33. Haziza, C., Stragier, P., Patte, JC (1982) EMBO J. 1, 3, 379-384.
34. Cassan, M., Richaud, F., Patte, JC (1982) FEMS Microbiol. Letters, 13, 177-180.