Volume 16 Number 9 1988 Nucleic Acids Research

3′ End of the *malEFG* operon in *E.coli*: localization of the transcription termination site

Eric Francoz and Elie Dassa

Unité de Programmation Moléculaire et de Toxicologie Génétique, CNRS UA271, INSERM U163,
Insitut Pasteur, 28 rue du Dr Roux, 75015 Paris, France

Received January 7, 1988; Revised and Accepted March 9, 1988 Accession no.X06663

ABSTRACT

The nucleotide sequence of a 981 bp's *Hinc*II-*Pvu*II DNA fragment containing the 3' end of the *malEFG* operon in *E. coli* was determined. This sequence displayed a putative Rho-independent transcription termination site localized 87 bp's after the stop codon of *malG*. When cloned into plasmid pKG1800, the *Hinc*II-*Pvu*II fragment containing this structure acted as a strong transcription termination signal. By S1 mapping, we demonstrated that the 3' end of the *malEFG* transcript coincided with the putative transcription termination site. One short open reading frames *orf1* (123 bp) and and the beginning of another one *orf2* were localized after *malG*. The transcription termination site is localized within *orf1*. Consequently *malG* is the last gene of the *malEFG* operon. *orf2* corresponds exactly to the 5' part of the *xylE* gene reported independantly (Davis & Henderson, 1987) as the gene coding for the XylE protein, the xylose-proton symport of *Escherichia coli.*

INTRODUCTION

In *Escherichia coli*, high affinity transport of maltose and maltodextrins is mediated by a shock sensitive, periplasmic binding protein-dependent transport system (reviewed in 1). Five proteins, localized in the three layers of the bacterial envelope, are involved in the transport and are encoded by genes clustered in the *malB* region of the chromosome. These genes are organized in two divergent operons : the *malKlamBmalM* operon and the *malEFG* operon (2 ; 3 ; figure 1). *lamB* encodes the maltoporin, which is necessary for maltodextrins uptake and serves as a receptor for bacteriophage Lambda (4). *malE* codes for the periplasmic maltose binding protein (5). *malF, malG* and *malK* encode three membrane proteins that are supposed to form a complex in the inner membrane (6 ; 7). The protein encoded by the *malM* gene has been recently shown to be localized in the periplasmic space but its role is unknown (8).

The DNA sequence of all genes known to be involved in maltose transport has been established (8-13). The complete DNA sequence of the *malKLamBMalM* operon is known and the transcription termination site is localized (8).

The question of the end of the *malEFG* operon was still left open. From genetic complementation and mapping of mutations affecting growth on maltose, *malG* was considered as the last gene of the *malEFG* operon (3). However, from the genetic analysis of mutants impaired in maltodextrins but not in maltose transport, Wandersman et al. (14) suggested the existence of an additional gene localized after *malG* and involved in

© IRL Press Limited, Oxford, England.

maltodextrins transport. Moreover, down to 60 base-pairs (bp) after the stop codon of *malG*, no potential transcription termination signal had been found, while the beginning of a new open reading frame had been detected (13), that could, for example, encode a protein involved in maltodextrins but not in maltose entry.

To decide about the existence of a such supplementary gene, we sequenced the 980 bp's 3'-flanking region of *malG* and we determined the 3' end of the *malEFG* transcript. We found a typical potential Rho-independent transcription termination signal 87 bp's after the *malG* stop codon. This signal acts as a transcriptional terminator when cloned in plasmid pKG1800. By S1 nuclease mapping of the 3' end of a *malG* transcript, we concluded that the *malEFG* operon ends 87 base pairs after the translational stop of the *malG* gene.

MATERIALS AND METHODS

1) Bacterial and phage strains

The bacterial strains used in this work are listed in Table I.

Strain RR1ΔM15 was used as recipient for pUR250 derivative plasmids. Strain TG1 was used for the propagation of phage M13mp8 derivatives. Strains ED51 and ED52 are derivatives of strain CSR603 (15), containing the F'*lacI*^{Q} episome and transformed respectively with plasmids pTAC12 (16) and pTMG1 (see below).

The M13mp8 (17) replicative form, cut with *Sma*I and dephosphorylated, was purchased from Amersham. Competent cells were prepared according to the method of Hanahan (18).

TABLE 1 : *Escherichia coli* strains used for this work.
All strains are derivatives of *Escherichia coli* K-12.

Strains	genotypes	origin
RR1ΔM15	*leu, pro, thi, rpsL, hsd* (r-,m-) *lacZΔM15* F'(*lacI*^{Q} *lacZΔM15*)	U. RUTHER
EFF46	same as RR1ΔM15 but containing pEFF46	This work
TG1	*pro, thi, hsd* (r-) *lacZΔM15*	S. WAIN-HOBSON F'(*lacI*^{Q} *lacZ ΔM15*)
N100	*pro, recA, galK, rpsL*	K. MAC KENNEY
EF 3	same as N100 but containing pEF3	This work
ED 51	*thr1, leuB, proA2(Δ), argE3, thi1, ara14, lacY, galK2, xyl5, man1, rpsL 31, tsx33, supE44, phr1, recA1, uvrA 56,* F'(*lacI*^{Q}*Z::Tn5*) (pTMG1)	This work
ED52	same as ED52 but containing pTAC 12	This work
ED11	*argH, his, recA, rpsL, malG1, F' lacI*^{Q}	This work

2). Plasmids

- Plasmid pMB3 (described in 13) carried an insertion of a 6 kilobases (kb) *Bgl* II fragment from the *malEFG* operon of phage λap*malB13* (19) in the *BamHI* site of pBR322 (Niels Fiil, unpublished). It contained DNA sequences extending more than 2 kb after the translational stop signal for *malG*. This plasmid complemented mutations in *malF* and *malG* and was used as source of DNA in subsequent subcloning experiments.
- Plasmid pED4 was described in (13).
- Plasmid pUR250 (20) was described in (13) and was used for subcloning.
- Plasmid pKG1800 (21) was used to test the presence of transcription termination sites.
- Plasmid pTAC12 (16) was used to overproduce MalG and its mRNA under the control of the *tac* promoter. In order to efficiently repress the *tac12* promoter, *E. coli* strains overproducing the *lac* repressor (*lac* I^Q strains) were used.

3) DNA cloning and sequencing techniques

General procedures for the isolation of DNA fragments, cloning, rapid purification of plasmids and their characterisation by restriction enzyme analysis have been described in (22).

The DNA sequence of the 3' flanking region of *malG* was determined using a "shot-gun" strategy (23 ; 24).

Plasmid pEF2 were sonicated in a Vibracell sonicator (Sonics and Materials) and DNA was fractionated on a 5% polyacrylamide gel. Fragments between 400 bp and 700 bp were purified, end-repaired by a treatment with *T4* DNA-polymerase, followed by a filling-in reaction with *E. coli* DNA-polymerase (Klenow fragment) and ligated to dephosphorylated, *Sma* I cleaved, M13mp8 replicative form. The products of the reaction were used to transfect *E. coli* strain TG1. Recombinant clones carrying fragments from the 3'-end of the *malEFG* operon were screened by plate hybridization using a 979 bp *Hinc*II-*Pvu*II probe (fragment A on the figure 1) labelled by nick-translation (22).

Single-stranded templates were prepared from plaques exhibiting positive hybridization signals and were sequenced by the dideoxy chain termination procedure (25) using $\alpha^{35}S$ dATP (Amersham 400 Ci/mmol) and buffer gradient gels (26). Sequences were compiled and analyzed using the programs of Staden adapted by B. Caudron for the Centre de calcul de l'Institut Pasteur (27).

4) Assay of ß-galactosidase

The activity of ß-galactosidase was determined according to Miller (28) with the following modifications.

- Bacteria were lysed by sonication to preserve the activity of an hypothetic chimeric protein.
- Bacterial debris were removed by centrifugation before reading the absorbance of the reaction mixture.
- Proteins were determined according to Bradford (29).

Strains were grown in minimal M63 medium containing glucose and casaminoacids and were induced by Isopropyl β-D-thiogalactopyranoside (IPTG) (10^{-3}M) and cyclic AMP (10^{-3}M) during two generations before the lysis to ensure a complete induction.

5) Maxicells

Maxicells were prepared according to (15). Labelled proteins were analysed by autoradiography after electrophoresis on SDS-polyacrylamide gels (30).

6) RNA preparation and nuclease S1 mapping

Total RNA's were prepared by the hot phenol method (31) from strain ED51 either uninduced or induced with IPTG (10^{-3}M). S1-nuclease mapping of the 3' end of *malEFG* mRNA was realized as described (31). Two kinds of probes were prepared : 1- a *Fok*I fragment (figure 6A) from plasmid pTMG1 3'-labelled only on the coding strand with the Klenow fragment of the *E. coli* DNA polymerase and ^{32}P dTTP (32) ; 2- a single stranded probe from phage HP2, a derivative of phage M13mp8 containing in its *Sma*I site a 979 bp *Hinc*II-*Pvu*II fragment from plasmid pED4 prepared as described (33)

7) Materials

All enzymes were purchased from Boehringer (Mannheim) except for the *T4* DNA Ligase which was purchased from Biolabs. They were used according the specifications of the supplyer. Labelled deoxyribonucleotides (^{35}S and ^{32}P) were purchased from Amersham. Dideoxyribonucleotides and cold deoxyribonucleotides were from Boehringer. Nitrocellulose filters were purchased from Schleicher and Schuell.

RESULTS

1. Nucleotide sequence

A 2.8 kb *Hinc* II fragment, named C (figure 1) from pMB3, starting 30 nucleotides before the *malG* stop codon and ending near the hybrid *BamH*I-*Bgl*II site of this plasmid was subcloned into the *Hinc*II site of pUR250. By restriction enzyme analysis of recombinant plasmids, we characterized plasmid pEF1, in which the *malB* fragment is cloned on its natural orientation downstream to the *lac* promoter and plasmid pEF2, where the same insert was present on the opposite orientation. As pEF1 gave poorer DNA yields, we used pEF2 as source of DNA to subclone random DNA fragments from the distal part of the *malEFG* operon into M13mp8, as described in Materials and Methods. Using this shotgun method, we determined the nucleotide sequence of 981 bp's 3' to the *Hinc*II site in *malG* (figure 2).

We found total agreement with the sequence previously determined (13) for positions 1 to 87. At position 36, two nucleotides after the stop codon of *malG*, we found an ATG codon preceeded by a sequence GAAAGGT that is complementary to the 3' end of the 16S RNA of *E. coli* (34). This ATG was followed by an open reading frame (*orf1,* 123 bp long) having a coding capacity for a polypeptide with 41 amino acids.

We found within *orf 1*, at position 103, a GC-rich palindromic structure followed by a stretch of T's that have all the characters of a typical Rho-independent transcription termination

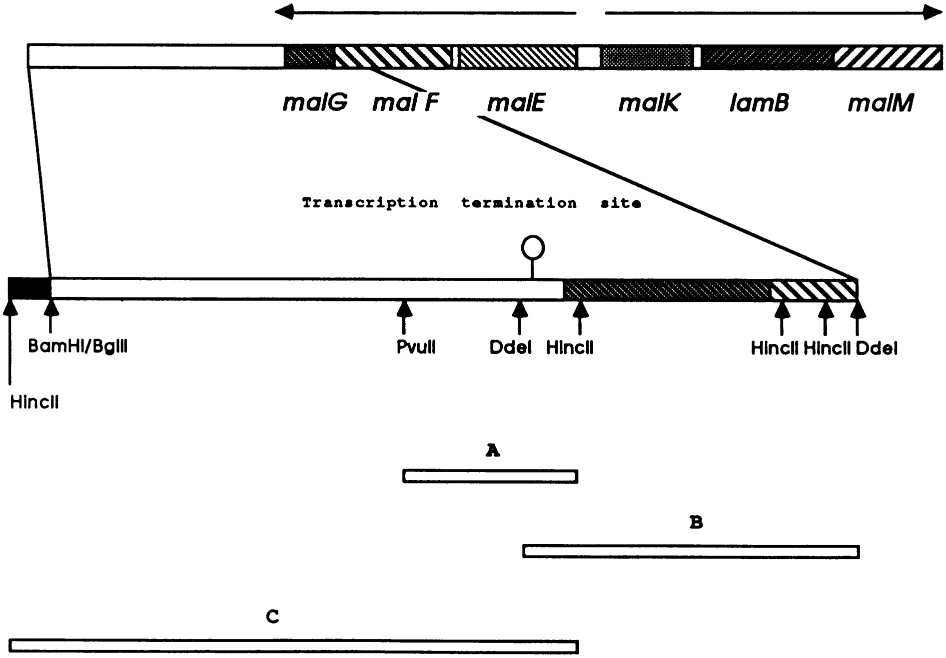

Figure 1 : Structure of the *malB* region and fragments cloned in differents plasmids.
Upper part : Structure of the *malB* region.
Lower part : Blow up of the end of the *malEFG* operon

The DNA fragments clones in the plasmids used in this work are represented.

- pTMG1 was constructed by cloning the 1350 bp fragment B under the *tac* promoter in the *Pvu*II site of pTAC12.
- pEF1 and pEF2 were constructed by cloning in both orientations the 2.8 kb fragment C in the *Hinc*II site of pUR250.
- pEF3 was constructed by cloning, in its natural orientation, the 979 bp fragment A in the *Sma*I site of pKG1800

site (figure 3). In accordance with Brendel and Trifonov (35) , we found two nucleotides after the run of thymine residues a TCTG motif and five nucleotide before a CGG motif.This structure has a high stabilization energy (36) ΔG = -12.6 kilocalories (-52.6 kilojoules) and could constitute a transcription termination signal for the *malEFG* operon.

At position 402, the beginning of a second open reading frame (*orf* 2) preceeded by a potential ribosome binding site GCAGG was detected. The region comprised between the stop codon for *orf1* and the start codon for *orf2* is 241 bp's long and is AT-rich (69.1%). Translation stop codons were distributed rather evenly within the three reading frames. A search by eye for regions of homology with the consensus sequence of *Escherichia coli* promoters indicated three possible occurences of a promoter.

The fragment of *orf* 2 was 578 bp long and encoded the 192 first amino acids of a

```
                                                          60
AACGGCCTGACGGCAGGTGGTGTGAAAGGTTAAAGATGTTGTTCTGCCAATGTTATGCCG
AsnGlyLeuThrAlaGlyGlyValLysGly***  MetLeuPheCysGlnCysTyrAlaA
HincII                                    orf1
                                                         120
CTGCACCCTCAACTTACGTTATCCCAACTTGTGACTGTTATTCGGCGCTCCACGGAGCGC
laAlaProSerThrTyrValIleProThrCysAspCysTyrSerAlaLeuHisGlyAlaL
                                                         180
TTTTTTTCTTTCGTCTGCAATCTGAATCGTTCGCCGGTTAATATTTCCATCATAGAGCTT
euPhePhePheArgLeuGlnSerGluSerPheAlaGly***
▲    ▲
                                                         240
ATTATTTTTACGTTATTTGTTTTCCCACTTACGATAATTCTCTTTCGTGCTCTGAGTCAC
                                                         300
GGCAATAGTATTGTTTTTATCAATTTTGGATAATTATCACAATTAAGATCACAGAAAAGA
                                                         360
CATTACGTAAACGCATTGTAAAAAATGATAATTGCCTTAACTGCCTGACAATTCCAACAT
                                                         420
CAATGCACTGATAAAAGATCAGAATGGTCTAAGGCAGGTCTGAATGAATACCCAGTATAA
                            DdeI           MetAsnThrGlnTyrAsn
                                            xylE
                                                         480
TTCCAGTTATATATTTTCGATTACCTTAGTCGCTACATTAGGTGGTTTATTATTTGGCTA
 SerSerTyrIlePheSerIleThrLeuValAlaThrLeuGlyGlyLeuLeuPheGlyTyr
                                                         540
CGACACCGCCGTTATTTCCGGTACTGTTGAGTCACTCAATACCGTCTTTGTTGCTCCACA
 AspThrAlaValIleSerGlyThrValGluSerLeuAsnThrValPheValAlaProGln
                                                         600
AAACTTAAGTGAATCCGCTGCCAACTCCCTGTTAGGGTTTTGCGTGGCCAGCGTTCTGAT
 AsnLeuSerGluSerAlaAlaAsnSerLeuLeuGlyPheCysValAlaSerValLeuIle
                                              BalI       660
TGGTTGCATCATCGGCGGTGCCCTCGGTGGTTATTGCAGTAACCGCTTCGGTCGTCGTGA
 GlyCysIleIleGlyGlyAlaLeuGlyGlyTyrCysSerAsnArgPheGlyArgArgAsp
                                                         720
TTCACTTAAGATTGCTGCTGTCCTGTTTTTTATTTCTGGTGTAGGTTCTGCCTGGCCAGA
 SerLeuLysIleAlaAlaValLeuPhePheIleSerGlyValGlySerAlaTrpProGlu
                                                     BalI
                                                         780
ACTTGGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTATCTGGCAGGTTATGT
 LeuGlyPheThrSerIleAsnProAspAsnThrValProValTyrLeuAlaGlyTyrVal
                                                         840
CCCGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCT
 ProGluPheValIleTyrArgIleIleGlyGlyIleGlyValGlyLeuAlaSerMetLeu
                                                         900
CTCGCCAATGTATATTGCGGAACTGGCTCCAGCTCATATTCGCGGGAAACTGGTCTCTTT
 SerProMetTyrIleAlaGluLeuAlaProAlaHisIleArgGlyLysLeuValSerPhe
                                                         960
TAACCAGTTTGCGATTATTTTCGGGCAACTTTTAGTTTACTGCGTAAACTATTTTATTGC
 AsnGlnPheAlaIleIlePheGlyGlnLeuLeuValTyrCysValAsnTyrPheIleAla

CCGTTCCGGTGATGCCAGCTG
              PvuII
 ArgSerGlyAspAlaSer
```

Figure 2 : Sequence and structure of the 3' flanking region of *malG*.
The non-coding strand is represented and the protein sequence is shown under the DNA sequence.The numbering starts in the *Hinc*II cutting site located at the 3' end of *malG*. Relevant restriction sites are also shown and underlined. Stop codons are symbolized by three stars. The transcription termination site falls between the two arrows (▲). Putative ribosome binding sites are boxed.

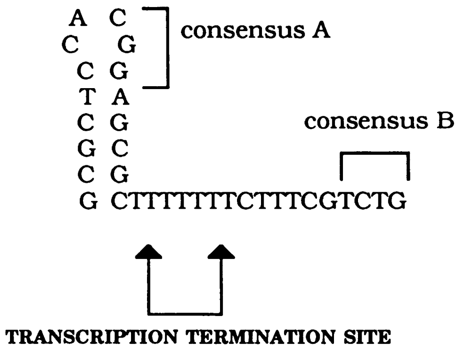

Figure 3 : Secondary structure of the transcription termination site.
The consensus sequences A and B described by Brendel and Trifonov (34) are shown. The 3' end of the *malEFG* transcript is located within a zone delimited by black arrows. This secondary structure is highly stabilized, its free energy is -52,6 kilojoules/mole.

polypeptide. In this fragment, the frequency of optimal codons (37) had a low value (*f*=0.48). Hence *orf 2* could be a gene expressed at low levels. It contained a large excess of hydrophobic amino acids and could determine part of an hydrophobic, probably membrane-bound protein. *orf2* corresponded exactly to the *xylE* gene, coding for xylose-proton symport in *Escherichia coli.*, whose nucleotide sequence has been recently reported by Davis and Henderson (39).

2. XylE is not co-expressed with MalG

To monitor its eventual expression, *xylE* was fused to the α-*lacZ* coding sequence present on pED4. On this plasmid, a translational fusion between XylE and the α-fragment of β-galactosidase could be constructed by deleting DNA fragments starting from the 5' proximal *Bal*I site in *xylE* and ending at *Eco*RI site in the *lacZ* DNA sequence (figure 4 top). From the DNA sequence of pED4, one expects to construct an in frame fusion by digesting the 5'-protruding ends of the *Eco*RI site with the Mung Bean nuclease and ligating the generated flush end to the 3' end of the *Bal*I site. The mixture was used to transform strain RR1ΔM15. All ampicillin-resistant transformants presented a white Lac$^-$ phenotype on 5-Bromo-4-Chloro-3-Indolyl β-D-Galactoside (Xgal), ampicillin, IPTG plates.

We characterized plasmid pEFF46 from one white clone. A *Hinc*II-*Pvu*II fragment was subcloned in M13 mp8. The sequence at the fusion joint indicated that *xylE* was fused in frame with *lacZ* (figure 4, bottom). One codon of *lacZ* was lost in the course of the fusion experiment. This anomaly could be attributed to a limited double strand digestion by the Mung Bean nuclease. The activity of the fusion protein was assayed and found to be undistinguishable from that present in the *lacZ* deletion strain RR1ΔM15, even in the presence of the inducers IPTG and cAMP (table II). Addition of maltose (0.2 %) did not change the

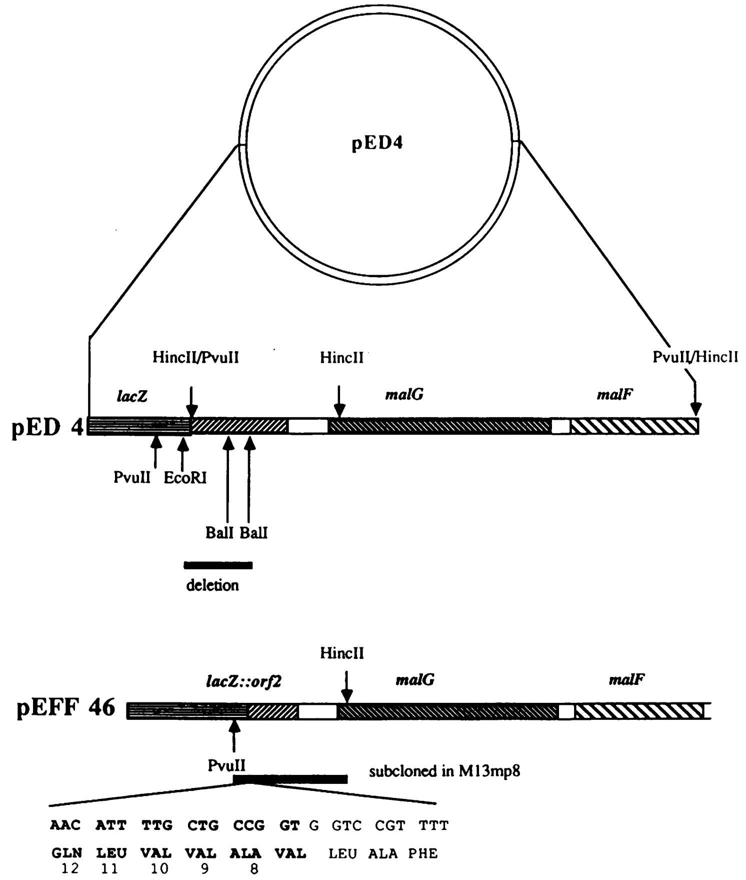

Figure 4 : Construction of a translational *xylE:lacZ* fusion in pEFF46.
The top of the figure represents the structure of the insert present in pED4, a derivative of pUR250. An in-frame fusion is realized by digesting the plasmid DNA with *Eco*RI and *Bal*I, removing the 5' protruding end of the *EcoR*I site with the Mung Bean Nuclease and ligating. To check the nucleotide sequence of the fusion joint, a *Hinc*II-*Pvu*II fragment of pEFF46 starting from the *Hinc*II site in *malG* and ending in the *Pvu*II site in *lacZ* was subcloned in M13mp8. The nucleotides from the *lacZ* sequence are in bold characters. The amino acid sequence is shown under the DNA sequence and the numbers refer to the position of the residues in the mature β-galactosidase.

basal level of ß-galactosidase. Expression of the *xylE*::*lacZ* fusion could however be induced by xylose (data not shown). These result suggest that *xylE* does not belong to the *malEFG* operon since it is not co-expressed with *malG*.

We then undertook to determine directly the transcription termination site.

3. The potential transcription termination site detected 3' to *malG* is functional.

A 979 bp *Hinc*II-*Pvu*II fragment, named fragment A (figure 1), carrying the potential

TABLE 2 : Specific activity of β-galactosidase and of the hybrid *xylE*::*lacZ* protein fusion. Activities are expressed in μmole min^{-1} mg^{-1} o-nitrophenol formed.

strains	no induction	IPTG cAMP
RR1ΔM15	**0.045**	**0.060**
RR1ΔM15 pUR250	**7.8**	**114**
RR1ΔM15 pEFF46	**0.21**	**0.057**

transcription terminator was cloned in the unique *SmaI* site present in the *galE* gene of the plasmid pKG1800, between the *gal* promoter and the *galK* gene. We characterized, by restriction enzyme analysis, the recombinant plasmid pEF3 which contained this fragment in its natural orientation under the control of the *gal* promoter .

Plasmid pEF3 was unable to complement the *galK* mutation of strain N100. This indicated that the *galK* gene present on the plasmid pEF3 was not expressed and, consequently, that fragment A contained a strong transcription termination signal.

To confirm this result, we undertook a precise localization of the 3' end of the *malEFG* transcript. For this purpose, we constructed the recombinant plasmid pTMG1 in order to amplify mRNA's from the distal end of the operon. A 1350 bp *Dde*I fragment containing *malG*, named B on figure 1 was end-repaired using *E.coli* DNA polymerase Klenow fragment in the presence of the four dNTP's and ligated to pTAC12 cleaved by *Pvu*II and dephosphorylated. The ligation products were used to transform strain ED11. Mal^+, ampicillin-resistants derivatives were screened on Mac Conkey agar plates supplemented with maltose and ampicillin. Plasmid DNA was isolated from Mal^+ colonies and analyzed with restriction enzymes. All Mal^+ clones yielded a recombinant plasmid in which the *Dde*I fragment was cloned in the right orientation under the the control of the *tac* promoter.

malG was correctly expressed on pTMG1 (figure 5, lane 3). Maxicells prepared from strain ED52 (containing pTMG1) were shown to direct the synthesis of a 24 kd polypeptide, with the same migration as MalG on SDS-polyacrylamide gels, as determined previously in an

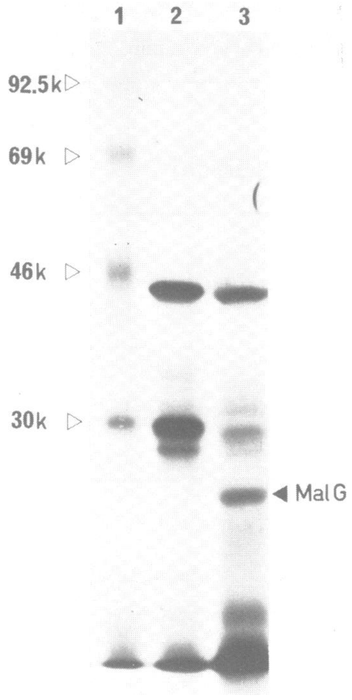

Figure 5 : Synthesis of polypeptides directed by pTMG1.
Polypeptides synthetized in maxicells analyzed by autoradiography after electrophoresis on SDS-polyacrylamide gels (29).
- Lane 1 : Size markers; phosphorylase b (M.W. 92,500), bovine serum albumine (M.W. 69,000), ovalbumine (M.W. 46,000), carbonic anhydrase (M.W. 30,000).
- Lane 2 : Maxicell total extract of strain ED 52 containing the plasmid pTAC12. The polypeptide with M.W. 30 000 is probably the β-lactamase and that with M.W. 45 000 is encoded by the F'*lac* factor.
- Lane 3 : Maxicell total extract of strain ED51 containing the recombinant plasmid pTMG1. The apparent M.W. of MalG is 24 000.

in vitro coupled transcription-translation system (Dassa and Hofnung, 1985). This polypeptide was absent in extracts made in parallel from strain ED51 (containing pTAC12).

Total RNA's were prepared from strain ED51, carrying plasmid pTMG1 as described in Materials and Methods. After hybridization of either induced or uninduced mRNAs with the *Fok*I probe and treatment with S1 nuclease, a family of fragments ranging between 221 and 226 bp's was protected (figure 6B). This size corresponds exactly to the distance between the end of the probe and the putative transcription termination site. The 3' end of the transcript is localized within the stretch of thymidine residues following the palindrome (figure 3). A band that could correspond to the native double-stranded probe is present on the top of the gel.

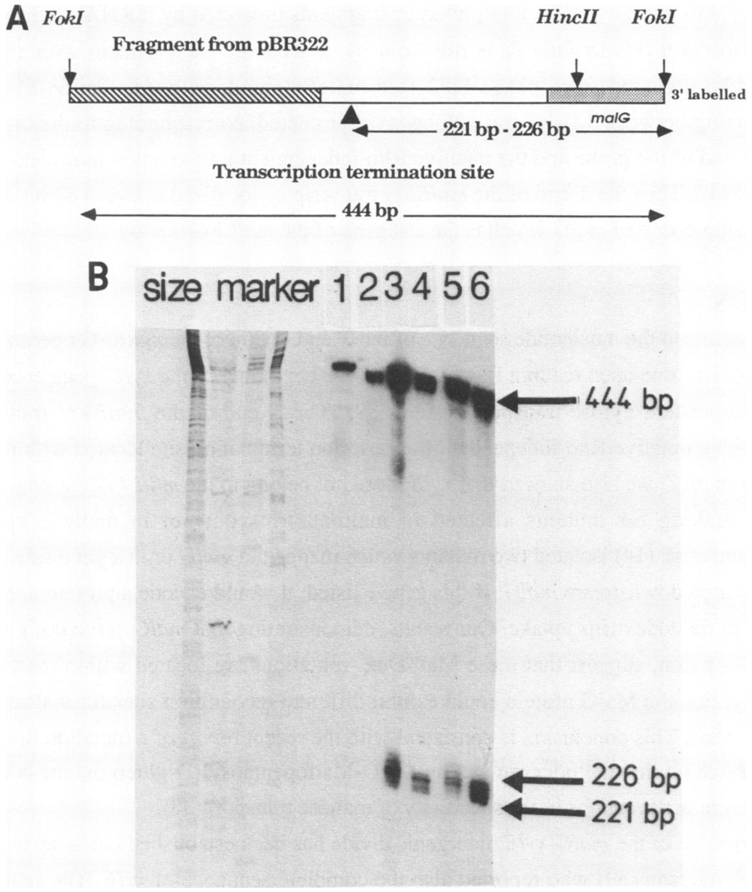

Figure 6 : S1 nuclease mapping of the 3' end of the *malEFG* operon.
-part A : Structure of the *Fok*I probe used for the S1 mapping. A 444 bp's *Fok*I fragment from pTMG1 was 3'end-labelled. By using α ^{32}P dTTP, only the *Fok*I site in *malG* was labelled. The deduced location of the 3' end of the *malEFG* transcript is indicated ( )
-part B : Autoradiography of a gel showing S1 nuclease mapping of the 3' end of transcripts terminating within the *malEFG* operon. The *Fok*I probe (20 000 cpm in each experiment) was denatured and hybridized to RNA's of strains ED51 and ED52. Hybrids were digested with nuclease S1 and run on a sequencing gel. Reaction products of a dideoxy DNA sequence were loaded as size markers.
Lane 1 : Double stranded probe without RNA and S1 nuclease treatment.
Lane 2 : Probe without RNA but treated with 300 units of S1 nuclease.
Lane 3 : Probe with ED 52 RNA prepared after IPTG induction treated with 300 units of S1 nuclease.
Lane 4 : Same experiment as in Lane 3 but treated with 1000 units of S1 nuclease.
Lane 5 : Probe with ED 52 RNA prepared without induction treated with 300 units of S1 nuclease.
Lane 6 : Same experiment as in Lane 5 but treated with 1000 units of S1 nuclease.

Alternatively, this band could be interpreted as fragments protected by mRNA's longer than the probe. However this hypothesis is ruled out by a second S1 experiment with a single stranded probe prepared from phage HP2 (see materials and methods, section 6). Only fragments sizing between 121 bp and 126 bp were protected, corresponding to the distance between the end of the probe and the putative Rho-independent transcription termination site (data not shown). Thus the 3' end of the *malEFG* transcript is localized between 87 bp and 92 bp after the stop codon of *malG* which is the last gene of the *malEFG* operon.

DISCUSSION

We determined the nucleotide sequence of the 3' end region of the *malEFG* operon. This sequence contains one open reading frame *orf1* and the beginning of the *xylE* gene encoding the proton-dependent xylose transport system (39). The 3' end of the *malEFG* transcript coincides with a putative Rho-independent transcription termination site located within *orf1*. By fusion with *lacZ*, we also showed that *xylE* does not belong to the *malEFG* operon.

While seaking for mutants affected in maltodextrins but not in maltose uptake, Wandersmann et al. (14) isolated two mutants which mapped in *malG* or in a yet unidentified new gene located downstream *malG*. If this gene existed, it would encode a protein involved specifically in maltodextrins uptake. Our results, demonstrating that *malG* is the last gene of the *malEFG* operon, suggest that these $Mal^{+}Dex^{-}$ mutations are located within *malG* and consequently, that the MalG protein could exhibit different recognition sites for maltose and for maltodextrins. This conclusion is consistent with the recent report of a mutation in *malG* allowing growth on the maltodextrin analog 4-β-D-Maltopentaosyl-D-gluco pyranoside that presented severe perturbations in the specificity of maltose transport (38).

The sequence of the *malG-xylE* intergenic divide has been established independently by Davis and Henderson (39) who reported also the complete sequence of *xylE*. The sequence they determined for the 981 bp's fragment is identical to that presented in this paper excepting two differences in the stem and loop structure defining the transcription termination signal of the *malEFG* operon. At position 116, we found a GCGC motif rather than GCCG as in their sequence. If the latter was true, the stem of the terminator would be destabilized by the presence of a non pairing GC sequence (107-108). This difference may be due to base compression on the gels while sequencing DNA fragments with secondary structures (the DNA sequence for positions 1 to 175 was covered only once in the work from Davis and Henderson) or to allelic differences between the strains used. Our gels show clearly that the actual sequence is GCGC and this is substantiated by our results showing that the Rho-independent transcription termination signal acts as a strong terminator.

The complete sequence of the *malB* regulon involved in maltose transport is now established. A function for all the genes except *malM* has been proposed. The complete molecular description of the regulon opens the way to an exhaustive study of this system.

AKNOWLEDGEMENTS

We are grateful to Maurice Hofnung for his constant interest, for helpful discussions and for reviewing the manuscript. We thank W. Saurin for help in computer analysis.

This work was supported by grants from the Fondation pour la Recherche Médicale, the Ligue Nationale contre le Cancer and the Association pour le Développement de la Recherche sur le Cancer.

REFERENCES

1 - Hengge R. & Boos W. (1983) Biochem. Biophy. Acta, **732**, 443-478.
2 - Hofnung M. (1974) Genetics, **76**, 169-184.
3 - Silhavy T.J., Brickman E., Bassford P.J. Jr., Casadaban M.J., Shuman H.A., Schwartz V., Guarente L., Schwartz M. & Beckwith J. (1979) Molec. Gen. Genet., **174**, 249-259.
4 - Szmelcman S. & Hofnung M. (1975) J. Bacteriol.,**124**, 112-118.
5 - Kellerman O. & Szmelcman S. (1974) Eur. J. Biochem. **47**, 139-149.
6 - Shuman H.A. (1982) Ann. Microbiol. (Institut Pasteur), **133A**, 153-159.
7 - Shuman H.A., Silhavy T.J. & Beckwith J. (1980) J. Biol. Chem. **256**, 168-174.
8 - Gilson E., Rousset J.P., Charbit A., Perrin D. & Hofnung M. (1986) J. Mol. Biol., **191**, 303-311.
9 - Clément J.M. & Hofnung M. (1981) Cell, **27**, 507-514.
10 - Gilson E., Nikaido H. & Hofnung M. (1982). Nucl. Acid. Res., **10**, 7449-7458.
11 - Duplay P. Bedouelle H., Fowler A., Zabin I., Saurin W. & Hofnung M. (1984) J. Biol. Chem., **259**, 10606-10613.
12 - Froshauer S. & Beckwith J. (1984) J. Biol. Chem., **259**, 10896-10903.
13 - Dassa E. & Hofnung M. (1985). EMBO J., **4**, 2287-2293.
14 - Wandersmann C., Schwartz M. & Ferenci T. (1979) J. Bacteriol., **140**, 1-13.
15 - Sancar A., Mack A.M. & Rupp W.D. (1979) J. Bacteriol., **137**, 692-693.
16 - Amann E., Brosius J. & Ptashne M. (1983) Gene, **25**, 167-178.
17 - Messing J. & Vieira J. (1982) Gene, **19**, 269-276.
18 - Hanahan D. (1983) J. Mol. Biol., **166**, 557-580
19 - Marchal C., Greenblatt J. & Hofnung M. (1979) J. Bacteriol. **136**, 1109-1119.
20 - Rüther U. (1982) Nucl. Acid. Res., **10**, 5765-5772.
21 - Mac Kenney K., Shimatake H., Court D., Schmeissner U., Brady C. & Rosenberg M. (1981) in Gene Amplification and Analysis, Chirikjian J.C. & Papas T.S. eds., Vol. II, pp. 383-414, Elsevier North Holland.
22 - Maniatis T., Frisch E.F. & Sambrook J. (1982) Molecular cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor N.Y.
23 - Messing J., Crea R. & Seeburg P.H. (1981) Nucl. Acid. Res., **9**, 309-321.
24 - Deininger P.L. (1983) Anal. Biochem., **129**, 216-223.
25 - Sanger F., Nicken S. & Coulson A.R. (1977) Proc. Natl. Acad. Sci.USA, **74**, 5463-5467.
26 - Biggin M., Gibson T. & Hong G. (1983) Proc. Natl. Acad. Sci.USA, **80**, 3963-3965.
27 - Staden R. (1982) Nucl. Acid. Res., **10**, 2951-2961.
28 - Miller J. (1972) Cold Spring Harbor Laboratory Press, Cold Spring Harbor N.Y.
29 - Bradford M. (1976) Anal. Biochem., **72**, 248-254
30 - Laemmli U.K., (1970) Nature, **227**, 680-685.
31 - Aiba H., Sankar A. & de Crombrugghe B. (1981) J. Biol. Chem., **256**, 11905-11910.
32 - Burke J.F. (1984) Gene, **30**, 63-63.
33 - Weaver R.F. & Weissman C. (1979) Nucl. Acid. Res., **7**,1175-1193.
34 - Shine J. & Dalgarno L. (1974) Proc. Natl. Acad. Sci.USA, **75**, 3737-3742.
35 - Brendel V. & Trifonov E.N. (1984) Nucl. Acid. Res., **12**, 4411-4427.
36 - Tinoco I. Jr., Borer P.N., Dengler B., Levine M.D., Uhlenbeck O.C., Crothers D.M. & Gralla J. (1973). Nature New Biology, **246**, 40-41.
37 - Ikemura T. (1981) J. Mol. Biol., **151**, 389-409.
38 - Ferenci T., Muir M., Lee K. S. & Maris D. (1986) Biochim. Biophys. Acta, **840**, 44-50
39 - Davis E. O. & Henderson P. J. F. (1987) J. Biol. Chem., **262**, 13928-13932