An Antisense Transcript from within the *ptsG* Promoter Region in *Escherichia coli*

Carole Pennetier^a Jacques Oberto^b Jacqueline Plumbridge^a

^aUPR9073-CNRS Institut de Biologie Physico-Chimique, Paris, ^bUMR8621-CNRS Institut de Génétique et Microbiologie, Université Paris X1, Orsay, France

Key Words

Divergent promoters • sRNA • Cryptic transcripts • Mlc • Glucose • CUT

Abstract

The *ptsG* gene, encoding the major glucose uptake system in *Escherichia coli*, is expressed from 2 promoters, a minor promoter p2 and a major downstream promoter p1. Transcription from both promoters is repressed by Mlc, and expression of p1 is activated by the cAMP/catabolite activator protein complex. Expression from p1 is also regulated post-transcriptionally in response to sugar stress via an sRNA, SgrS, which results in translational inhibition and mRNA degradation. Here, we demonstrate an additional level of complexity to the transcriptional pattern surrounding *ptsG*. A third promoter, p3, located between p1 and p2, was found to express a transcript antisense to *ptsG*. This promoter was detected by in vitro transcription and by RNA polymerase footprinting techniques and in vivo by S1 analysis and fusions with a *lacZ* reporter gene. Although the intrinsic strength of the p3 promoter was comparable to that of *ptsG*, it proved difficult to identify a full-length transcript. A faint transcript of greater than 400 nt could be detected. The transcript thus has more of the characteristics of a divergently expressed cryptic unstable transcript (CUT) than a prokaryotic sRNA.

Copyright © 2010 S. Karger AG, Basel

Introduction

Work from many different laboratories over the last few years has shown that expression of *ptsG*, encoding the major transporter for glucose in *Escherichia coli*, is subject to controls at numerous levels. Moreover, the PtsG protein itself is an important regulatory factor in the expression of many other genes via its roles in catabolite repression and inducer exclusion and its control of the activity of the transcriptional repressor Mlc.

The *ptsG* gene encodes EIICB^{Glc}, the membrane-specific component of the glucose phosphotransferase system (PTS). The soluble cytoplasmic part of the glucose PTS, EIIA^{Glc}, is encoded by *crr*, which is part of the *ptsHicrr* operon, encoding the common central components of the PTS, Enzyme 1 and HPr. EIIA^{Glc} is the key player in the control of use of other carbon compounds. The phosphorylation state of EIIA^{Glc} determines whether the cell is directed towards exclusive use of glucose (via inducer exclusion preventing the use of other carbon sources) or whether it activates adenylate cyclase, thus increasing cAMP levels and favoring the use of alternative carbon sources. The transport of glucose via PtsG is the major factor determining the phosphorylation state of EIIA^{Glc} [Kimata et al., 1997], and thus PtsG is necessary for inducer exclusion (via EIIA^{Glc}) and catabolite repression by affecting the levels of cAMP/catabolite activator

Fig. 1. Organisation of the *ptsG* regulatory region. **A** The locations of the Mlc and CAP operators and p1, p2 and p3 promoters relative to the p1 transcription start site are shown. The oligonucleotides used for PCR are shown; their 5' ends are at -237 (Glc1), -58 (Glc3), +68 (Glc4) and -298 (Glc5). **B** Sequence of the *ptsG* promoter region between oligonucleotide Glc1 and the beginning of the *ptsG* ORF. The Mlc and CAP consensus sites are underlined. The -10 and -35 sequences of the p1, p2 and p3 promoters are shown in bold italics (there is no consensus -35 sequence for p1).

protein (CAP) [for reviews, see Deutscher et al., 2006; Görke and Stulke, 2008].

questering Mlc to the membrane and preventing it from repressing the expression of *ptsG* and its other targets [Lee et al., 2000; Nam et al., 2001; Tanaka et al., 2000]. Expression of the major p1 promoter requires cAMP/CAP [Kimata et al., 1998; Plumbridge, 1998b]. Thus, growth on glucose has two conflicting effects on *ptsG* expression; it enhances expression by relieving repression by Mlc but also reduces expression by lowering the concentration of cAMP/CAP.

ptsG is sensitive to oxidative, anaerobic and thermal stress.

In addition to this multitude of transcriptional regulatory mechanisms, expression of *ptsG* p1 is also controlled post-transcriptionally in response to phosphosugar stress. The accumulation of Glc6P (e.g. in a *pgi* mutant in the presence of glucose) activates a transcription factor, SgrR, which promotes the expression of a small RNA, SgrS [Vanderpool and Gottesman, 2004]. This RNA base-pairs in the presence of Hfq with the translational initiation region of *ptsG* mRNA and prevents *ptsG* translation and ultimately allows RNase E-dependent mRNA decay [reviewed in Aiba, 2007; Maki et al., 2010]. The existence of these multiple levels of regulation demonstrates how it is crucial for the bacteria to control *ptsG* expression and, in consequence, regulate the use of glucose and alternative carbon sources.

During experiments to investigate transcriptional regulation of *ptsG* in vitro, we detected the existence of a third transcript synthesised from within the *ptsG* promoter region. Mapping of the transcript showed that it arose between the p1 and p2 promoters and was expressed in the opposite direction. The gene upstream of *ptsG*, *ycfH*, is expressed in the same direction as *ptsG* and there is no obvious open reading frame (ORF) in the direction of the p3 promoter. This suggested that the product of a p3 mRNA could correspond to a new sRNA. In this work, we have characterised the p3 promoter and its transcript in vitro and in vivo. Although no known function is identified for the transcript, we discuss the similarity of this transcript to the cryptic unstable transcripts (CUTS) which are ubiquitous in yeast and most eukaryotes.

Results

A Third Promoter in the ptsG Promoter Region

In vitro transcription (IVT) using a fragment of DNA covering the region from -237 to +68 compared to the *ptsG* p1 promoter (Glc1–Glc4; fig. 1) in the absence of cAMP/CAP detected 2 transcripts. The longer transcript corresponded to the p2 promoter (209 nt) while the other transcript (p3) of about 130 nt was not expected (fig. 2A, lane 1). The use of a template shortened at one end demonstrated that the p3 transcript corresponded to a transcript expressed in the opposite direction to p1 and p2 (data not shown). Addition of cAMP/CAP activated expression from p1 but reduced expression from p3 (fig. 2A, lane 2). Addition of Mlc after formation of an open complex with cAMP/CAP and RNA polymerase (RNAP) had

little effect on the transcription pattern (fig. 2A, lane 3), but addition of Mlc before cAMP/CAP and RNAP reduced the intensity of all 3 transcripts (fig. 2A, lane 4). These data showed that the p1 and p2 promoters are repressed in vitro by Mlc and that the p1 promoter but not the p2 promoter requires cAMP/CAP for activity. The p3 transcript appears to be expressed constitutively in vitro but is repressed by cAMP/CAP and Mlc. The same transcription pattern was detected using supercoiled plasmid DNA as template (data not shown).

DNase I footprinting showed that cAMP/CAP bound to the previously identified CAP sites centred at position -40.5 compared to p1 (CAP site 1, CGTGA-N6-TCAAA) and also more weakly to a second site at -95.5 (CAP site 2, TGTGA-N6-TTTTG; bases corresponding to the CAP consensus are underlined) (fig. 1, 2B, lanes 2–4) [Shin et al., 2003]. RNAP alone was found to very weakly protect a region near to the p3 start (fig. 2B, lane 7). The protection around p3 by RNAP was lost in the presence of cAMP and CAP but there was an extended footprint covering p1 from about +15 to position -90. In the presence of RNAP, cAMP/CAP bound to CAP sites 1 and 2, at concentrations which alone did not bind to the DNA (fig. 2B, lanes 5 and 6, c.f. lanes 2–4). The combined presence of 100 nM cAMP/CAP and RNAP produced a hypersensitive cleavage between CAP sites 1 and 2 (indicated with a dot, fig. 2B, lane 5). Mlc alone bound to the 2 sites Mlc1 and Mlc2 previously identified (fig. 2B, lane 10). When added before open complex formation with cAMP/CAP and RNAP, Mlc reduced the extent of the RNAP footprint upstream of CAP site 1 and from +1 to +15 (fig. 2B, lane 9). CAP was still bound to CAP site 1. There was no effect of Mlc when added after open complex formation, and the hypersensitive cleavage between CAP sites 1 and 2 was still detected (fig. 2B, lane 8). The shift of RNAP from p3 to p1 in the presence of cAMP/CAP and reduction of the RNAP footprint at p1 and p3 by Mlc are consistent with the pattern of IVT.

To further confirm that RNAP binding to p3 was inhibited by cAMP/CAP and Mlc, we used KMnO₄ sensitivity to detect single-stranded DNA regions formed during open complex formation. RNAP alone produced a single KMnO₄ band corresponding to the p3 promoter (fig. 2C, lane 2). The presence of cAMP/CAP produced a series of strong bands for p1 and reduced the signal at p3. Again, Mlc added after open complex formation had little effect on the p1 or p3 KMnO₄ cleavage bands but enhanced a cleavage site within the CAP1 binding site (fig. 2C, lane 4). This cleavage site was detected even when Mlc was added before open complex formation and when

Fig. 2. Demonstration of p3 promoter activity in vitro. **A** IVT on the Glc1–4 fragment. RNAP (1 unit, Epicentre Technologies), CAP (100 nM with 400 μ M cAMP) and Mlc (100 nM) were present as shown. After open complex formation at 37°C for 10 min, a mixture of ATP, GTP, CTP (200 μ M) and [α - 32 P]-UTP (20 μ M, 2 μ Ci) was added and elongation was allowed to proceed for 5 min. The origins of the bands at about 220 nt in lane 1 and 245 nt in lanes 2–4 are unknown. **B** DNase I footprinting on the Glc1–4 PCR fragment labelled at Glc4. CAP protein was present at the

concentrations indicated (nM), and the other proteins were added at the same concentrations as in **A**. Proteins were incubated for 10 min at 37°C before attack with DNase I. **C** Detection of single-stranded regions by KMnO_4 footprinting. Complexes were formed as in **A** and **B**. The asterisks indicate that Mlc was added before open complex formation. All reactions were analysed on 6% denaturing acrylamide gels. The locations of the p1, p2 and p3 transcription start sites are indicated. The marker is pBR322 digested with *MspI*.

no open complex formation at p1 was detected by KMnO_4 (fig. 2C, lane 5), suggesting it is due to an interaction between CAP and Mlc.

The 3 techniques, IVT, DNase I footprinting and KMnO_4 footprinting, give concordant results showing that a promoter exists within the *ptsG* regulatory region which expresses a transcript in the opposite direction to

p1 and p2. This transcript is inhibited by cAMP/CAP binding to the second CAP site, upstream of the p1 activating site, and it is also repressed by Mlc.

The p3 Transcript Is Expressed in vivo

We used S1 mapping with the Glc5*–4 PCR fragment (fig. 1, 3E) labelled at Glc5 to look for the presence of the

p3 promoter transcript in total RNA extracted from cultures growing in glucose or glycerol media. The 5' end of the p3 transcript was found to map to almost exactly the same position using mRNA synthesised in vivo as using RNA made by IVT (fig. 3A, compare lane 2 and lanes 4–6). This showed that the p3 promoter was active in vivo and was not an artefact of IVT on a fragment of DNA. Moreover, the fact that the p3 transcript was more strongly detected in the *mlc* strain grown in glycerol (fig. 3A, lane 4) and the wild-type strain grown in glucose (fig. 3A, lane 6) than in the wild-type strain grown in glycerol (fig. 3A, lane 5) is in agreement with the observation that the expression of p3 was controlled by Mlc in vitro (fig. 2A). [The extra band of 165 nt (fig. 3A, marked as X) corresponds to the centre of a palindromic sequence within the DNA which can form a cruciform structure in the double-stranded probe used for the S1 mapping. The cruciform structure in the DNA would provide a target for S1 nuclease. The same band is also seen with RNA made by IVT (fig. 3A, lane 2) and with a control containing tRNA (fig. 3A, lane 3).]

Fig. 3. Detection of the p3 transcript in total RNA. **A** S1 analysis using the Glc5 to Glc4 fragment labelled at Glc5 (**E**). RNA was made by IVT (lane 2), as in figure 2A, or is total RNA (30 µg) extracted from JM101 grown in glycerol or glucose medium or from IBPC1012 (*mlc*) grown in glycerol (Gly; lanes 4–6). Lane 1 is an A+G sequencing reaction on the probe. **B** Northern blot analysis of the p3 transcript using RNA from JM101 carrying pTSG10, pTSG20 or pBR322 and the Glc5–Glc3 fragment as probe. **C** S1 analysis of the p3 transcript using the Glc58–4 and Glc59–4 probes, where Glc58 and Glc59 were labelled with [γ^{32} P]-ATP and T4 polynucleotide kinase. p3 indicates the 336-nt transcript detected with the Glc58–4 probe and the 475-nt transcript detected with the Glc59–4 probe. **D** Primer extension analysis of *ptsG* and p3 mRNAs. Equal quantities of RNA made from JM101 carrying pTSG20 (+) or pBR322 (–) were annealed with equal amounts of the 5 oligonucleotides. Primer extension was carried out in the presence of ddATP (Glc1 and 2) and ddTTP (Glc9, 10 and 59). Glc1 (21 nt) is extended 11 nt to the first A; Glc2 (21 nt) is extended 8 nt to the first A; Glc9 (19 nt) is extended 8 nt to the first T; Glc10 (22 nt) is extended 6 nt to the first T, and Glc59 (19 nt) should be extended 11 nt to the first T (but no transcript was detected). **E** Scheme showing the location of the probes used for the S1 analysis of p3 in **A** and **C** and the primers for the abortive primer extension analysis in **D**. C = Control with 30 µg of tRNA. X indicates a cleavage within the probe corresponding to a cruciform structure forming within the double-stranded probe. The marker is pBR322 digested with *MspI*.

Fig. 4. Comparison of the effect of mutations in *cis* on p3 and *ptsG* promoter activity. Mutations affecting the Mlc1 and Mlc2 operators or the presumed -10 and -35 sequences of the p3 promoter were introduced into *ptsG-lacZ* and p3-*lacZ* fusions, and β-galactosidase activity was measured during growth on glycerol (gly) or glucose (glc) and in a strain carrying an *mlc* mutation during growth on glycerol. wt = Wild-type.

The activity of the p3 promoter was measured in vivo using a p3-*lacZ* fusion and compared to that of a *ptsG-lacZ* fusion carrying the same insert in the opposite orientation (fig. 4). The p3 promoter was found to be of comparable strength to the *ptsG* (p1 plus p2) promoter, and, moreover, expression was increased nearly twofold by growth on glucose and 5-fold by the presence of an *mlc* mutation, showing that, as in the case of the *ptsG* promoter, p3 expression was repressed by Mlc and induced by growth on glucose (fig. 4). Mutation of one or other of the *mlc* operators (mutations Mlc1 and Mlc2; fig. 4) produced derepression of both *ptsG* and p3 expression. Interestingly, the Mlc2 mutation in the upstream operator had a greater effect on p3 than on *ptsG*, whereas the reverse is true for *ptsG-lacZ*, where the Mlc1 mutation had a greater effect. Most *ptsG* expression comes from p1, which is primarily controlled by the Mlc1 operator [El Qaidi and Plumbridge, 2008].

Characterisation of the p3 Promoter

The 5' end of the p3 transcript as measured by IVT (fig. 2) and by S1 (fig. 3A) corresponded to a series of Ts at position -106 to -108 compared to p1 (fig. 1). Inspection of the DNA sequence suggested that the best candidates for the -35 and -10 promoter sequences were TTGCAC - 16 bp - TGTCAC - 4 bp - TTT. Mutation of the potential -35 sequence from TTGCAC to ACCCAC reduced expression from p3-*lacZ* 3-fold, while mutation of the putative -10 sequence from TGTCAC to CGCCAC reduced expression 6-fold during growth on glycerol

(fig. 4). This is consistent with their assignment as the -35 and -10 promoter elements. Introduction of an *mlc* mutation into the strain with the -35 mutation increased p3 expression to almost the level of the wild-type p3 fusion in the *mlc* strain. This suggests that the lack of a good -35 sequence decreases the ability of RNAP to bind to the p3 promoter and compete with Mlc repression, but in the absence of Mlc, RNAP can still bind to the mutated p3 promoter. Neither the -10 nor -35 mutation had any significant effect on *ptsG-lacZ* expression (fig. 4). Changing the rather poor consensus p3 -10 sequence to an optimal -10 sequence, TATAAT, produced a construct that was not stable in the p3-*lacZ* fusion, presumably due to excessive β-galactosidase activity, since the same sequence change was stable in the *ptsG-lacZ* fusion.

Length of the p3 Transcript

The *lacZ* fusion experiments suggested that the p3 promoter was rather strong and comparable to the *ptsG* promoter during growth on both glycerol and glucose. We used Northern blotting to try to detect the full-length p3 transcript. However, using the same RNAs as used for the S1 experiment (fig. 3A), no discrete transcripts were detected. Using RNA prepared from strains carrying multicopy plasmids including several kilobases of the *ptsG* chromosomal region (pTSG10 and pTSG20 [Erni and Zanolari, 1986]), we detected a faint signal of about 400 bp (fig. 3B). Inspection of the DNA sequence showed that the putative 3' end fell within the *ycfH* ORF (on the antisense strand) and that there was no obvious factor-

independent terminator. To try to confirm that the p3 promoter produced a discrete transcript of 400 nt, we repeated the S1 mapping using 2 longer probes which ended before and after the putative 3' end of a 400-bp p3 transcript (fig. 3E). One probe, Glc58*-4, was designed to detect a transcript of about 366 nt from p3, while the other, Glc59*-Glc4, would detect a 475-nt transcript but only if the p3 transcript were longer than 400 nt. Signals of the expected sizes were detected with both probes (fig. 3C). They were essentially only detected in the RNAs made from strains carrying the pTSG10 and pTSG20 plasmids, although a very faint signal could be seen from the chromosomal copy of the *mlc* strain (fig. 3C, lanes 2 and 8).

The difficulty to detect a discrete transcript from the p3 promoter implied that although the initial transcription was rather strong, the transcript itself was not stable. This view was supported by an abortive primer extension experiment. We compared the levels of p3 transcript at promoter proximal and distal positions with those of the *ptsG* mRNA. Using oligonucleotide Glc2, which hybridises to the beginning of *ptsG*, Glc10, which hybridises within *ptsG*, and Glc9, which overlaps the *ptsG* terminator (fig. 3E), we detected strong signals from the mRNA prepared from the pTSG20 strain but not from a strain carrying pBR322 (fig. 3D, lanes 1, 5 and 7). Using the Glc1 oligonucleotide, which hybridises near the 5' end of the p3 transcript, we detected a weak signal (fig. 3D, lane 3) but we detected nothing using the Glc59 oligonucleotide, which hybridises at 475 bp from the p3 promoter. This indicated that although the fusions with *lacZ* detected comparable β -galactosidase activity for *ptsG* and p3 fusions, implying similar promoter strength, the native p3 transcript is rapidly degraded (in the absence of the *lacZ* mRNA, which presumably stabilises the p3-*lacZ* fusion transcript). On the other hand, the *ptsG* transcript (corresponding to its uninduced level during growth on glycerol) is detected at both its 5' and 3' ends.

Discussion

We have demonstrated in vitro and in vivo the existence of a promoter within the *ycfH-ptsG* intergenic region of *E. coli* synthesising a transcript antisense to *ptsG*. The *lacZ* fusion data show that the p3 promoter is of similar strength to the promoters for *ptsG*, which encodes the major transporter for glucose. The basal levels of expression of p3-*lacZ* and *ptsG-lacZ* are almost identical, and the derepressed expression of p3 in an *mlc* strain is more than 50% that of *ptsG* (fig. 4). This suggested that the p3

promoter is rather strong and might be expected to have some function. However, attempts to detect the native p3 transcript by Northern blotting, S1 analysis or abortive primer extension analysis revealed only a low amount of a long transcript.

A revelation of the last 10 years has been the abundance and multiple functions of small, non-coding RNAs in bacteria. Very many have now been identified in *E. coli* using various techniques, both theoretical and experimental [reviewed in Vogel and Sharma, 2005]. The detection of an unexpected antisense promoter in the *ptsG* promoter region immediately suggested that it corresponded to a new sRNA. However, our analysis of the p3 promoter and its product does not suggest that it has many of the characteristics of typical sRNAs.

Of the numerous theoretical and experimental approaches employed to detect sRNA [Argaman et al., 2001; Kawano et al., 2005; Rivas et al., 2001; Tjaden et al., 2002; Vogel et al., 2003; Wassarman et al., 2001; Yachie et al., 2006; Zhang et al., 2003], none, as far as we are aware, has predicted or detected an sRNA corresponding to the p3 promoter. Two reasons can be suggested for this. Firstly, the theoretical approach looks for conserved signals within the sequence, e.g. promoters, terminators or secondary structure. The proposed promoter for p3 does not conform well to the consensus promoter sequence and there is no obvious factor-independent terminator sequence to define a 3' end. Secondly, the experimental approaches require that the sRNA be of significant abundance to be found by shotgun cloning, transcriptome analysis or Northern blotting. However, although the p3 promoter appears to be rather strong, we were unable to detect any distinct transcript from wild-type cells and only a weak transcript of about 400 nt even in cells carrying multicopy plasmids. Moreover, the 400-nt transcript detected by Northern blotting does not correspond to the full-length transcript, because S1 analysis detected even longer transcripts. Most sRNAs are in the range of 50–250 nt, although a few known transcripts are longer.

Considered together, the Northern blotting, S1, abortive primer extension and *lacZ* fusion data imply that RNA transcripts corresponding to near the 5' end of the p3 transcript are relatively abundant but become progressively weaker as the transcript becomes longer. This could be either because RNAP is less progressive and subject to pause sites and drop-off or because the transcript is rapidly subject to nucleolytic degradation.

One characteristic that p3 does share with many sRNAs is that it is expressed from an intergenic region in the opposite direction to the protein-coding genes [Vogel

Fig. 5. Multiple alignment of the region upstream of *ptsG* from 9 genera of Enterobacteriaceae (obtained with Clustal W). Operator sequences for Mlc and CAP in *E. coli* are shaded, and bases conserved in other bacteria are indicated by asterisks. The -10 and -35 promoter sequences for p1, p2 and p3 are underlined and in italics (note that the p3 promoter is on the other strand). The AUG for 8 of the bacteria is indicated, except for *Citrobacter koseri*, where the annotated translational start site is located 87 bp down-

stream. The dots above the *E. coli* sequence immediately upstream of the *ptsG* start codon indicate the bases involved in base pairing with SgrS sRNA. E.c. = *E. coli* K-12 substr. MG1655; S.b. = *Shigella boydii* CDC 3083 94; S.t. = *Salmonella typhimurium* LT2; C.k. = *Citrobacter koseri* ATCC BAA-895; K.p. = *Klebsiella pneumoniae* MGH 78578; E.s. = *Enterobacter sakazakii* ATCC BAA-894; S.p. = *Serratia proteamaculans* 568; Y.p. = *Yersinia pestis* CO92; E.t. = *Erwinia tasmenensis*.

et al., 2003]. The expression of sRNAs is often precisely controlled (e.g. SgrS [Vanderpool and Gottesman, 2004] and RhyB [Massé and Gottesman, 2002]), but because their targets are usually elsewhere on the chromosome they are not regulated in parallel with the surrounding genes but by specialised transcription factors. However, the p3 transcript is subject to the same control signals as *ptsG*; in particular, p3 is repressed by Mlc and activated by growth on glucose.

In vitro, the p3 promoter is active in the absence of cAMP/CAP and it appears to be repressed by cAMP/CAP binding to the site called CAP2, which displaces RNAP to the *ptsG* p1 promoter. However, we failed to detect any repression of p3 by cAMP/CAP in vivo by examining expression in *cya* strains in the presence and absence of exogenously added cAMP (data not shown). Moreover, as

CAP1 has a higher affinity for cAMP/CAP than CAP2, p1 should be activated before RNAP is chased from p3. We think that the apparent repression by cAMP/CAP in vitro could be due to competition for a limited amount of RNAP in the IVT assay.

The fact that p3 is subject to the same glucose-mediated controls as *ptsG* suggests that p3 could have a role in glucose utilisation. The p3 transcript would hybridise with transcripts coming from p2 and might regulate the level of the longer *ptsG* transcript, e.g. by providing a target for nucleases. However, as p2 accounts for only 10% of the total *ptsG* mRNA [El Qaidi and Plumbridge, 2008; Plumbridge, 1998b], its impact on overall *ptsG* expression would seem to be limited. Longer p3 transcripts are also antisense to the 3' end of *ycfH*, which terminates just 80 bp downstream of the p3 start (fig. 1A), so that p3 tran-

Table 1. Oligonucleotides used

Name	Use	Sequence (5'–3')
Glc1	IVT template, primer extension	CCGTCTGTTTCACATCGACGC
Glc2	primer extension	GCAGCATCAGCGATTACCGA
Glc3	Northern blot probe	CAGATCATTGGAGAAGTTGCAC
Glc4	IVT template, S1 probe	TACGTTCTCACGCGTGGCAA
Glc5	S1 probe, Northern blot probe	GGCTGTGTTGAAAGGTGTTGC
Glc9	primer extension	ATCTGGCTGCCTTAGTCTC
Glc10	primer extension	CAGAGTGCCAGATAGCAATTGC
Glc44	mutate Mlc1	GGCACTGAATTAT <u>G</u> CTACTCTGTGTAATAAATAAAGGGC
Glc45	mutate Mlc2	GAATGAGTTTTTCCAAAGCTCGTAATTAATGGCTA
Glc46	mutate –10 p3	CCACGTTTGAAATATTGTGG <u>C</u> GATGTTTTGTCAAAATGTG
Glc48	mutate –35 p3	TTGTCAAAATGTGGTTCTTCTCCAATGATCTGAA
Glc58	S1 probe	ACATCTCCTTTTCCGGCATTG
Glc59	S1 probe, primer extension	ATACGCTGGCAATTCTGCG

Base replacements in mutagenic oligos are underlined.

scripts could affect the expression of the *ycfH* gene. For example, the GadY sRNA overlaps the 3' end of the *gadX* gene and was found to stabilise the *gadX* transcript [Opdyke et al., 2004]. We wondered if p3 had some function in the sugar stress regulation involving SgrS and SgrR. Phosphosugar stress produced by α -methylglucoside (10 mM) resulted in induction of the p3-*lacZ* fusion similarly to that during growth on glucose, while a *ptsG-lacZ* translational fusion was not induced due to translational inhibition and mRNA degradation provoked by binding of the SgrS sRNA to the translation initiation region of *ptsG* (data not shown).

An indication of the importance of features within genomes can be deduced from their conservation in closely or more distantly related species. Orthologs of PtsG are widespread in both Gram-positive and Gram-negative bacteria. The organisation *ycfH-ptsG* is found in all Enterobacteriales and Vibrionales sequenced to date but absent in the other Gammaproteobacteria. We tried aligning the *ycfH-ptsG* intergenic regions and found that the sequences corresponding to Mlc1 and Mlc2 operators, p1, p2 and CAP site 1, are well conserved in 9 out of the 13 sequenced genera composing the Enterobacteriaceae (fig. 5). However, the region between Mlc2 and CAP1, although very rich in AT in most species, is less well conserved and it is difficult to identify the CAP2 site and the p3 promoter in anything other than *Escherichia* and *Shigella*. This suggests that if the p3 promoter function has been conserved across other species, the sequence has also evolved significantly. A recent survey of SgrS and

SgrR homologs has shown that regulation by the SgrS sRNA is likely to be widely distributed in enteric bacteria, but the exact sequences involved in the base pairing interaction have evolved and their precise functions have been modified [Horler and Vanderpool, 2009; Wadler and Vanderpool, 2009]. The same could be true for the p3 transcript.

A recent discovery in the RNA world is the pervasive nature of transcription in yeast and other eukaryotes. Cryptic unstable transcripts have been found to be widespread but have been best studied in yeast. They are often derived from divergent promoters at nucleosome-free regions (i.e. transcriptionally active sites) [Neil et al., 2009; Xu et al., 2009]. Although functions have been identified for only a few [Berretta et al., 2008], their most pronounced characteristic is that they are subject to rapid degradation unless stabilised by mutations in the exosome or capping machinery. Moreover, the vast majority are expressed antisense to the downstream gene and generate divergent transcripts [reviewed in Berretta and Morillon, 2009]. Although cryptic transcripts have not been extensively studied in bacteria, there is evidence from microarray studies and cDNA deep sequencing that there is widespread transcription corresponding to both sense and antisense strands [Selinger et al., 2000; Sharma et al., 2010]. Moreover, a recent survey by genomic Selex and deep sequencing found that the majority of Hfq-binding RNAs in *E. coli* corresponded to antisense transcripts [Lorenz et al., 2010]. The p3 transcript expressed divergently from the active *ptsG* promoter seems to have more

the character of a cryptic unstable transcript (CUT) than a prokaryotic sRNA. In either case it is presently in search of a function.

Experimental Procedures

Bacteriological Methods

E. coli JM-G301 carrying a *ptsG-lacZ* transcriptional fusion on a λ lysogen has been described previously [Pennetier et al., 2008]. The same insert, as an *EcoRI-BamHI* fragment, was inserted in the opposite direction in pRS1274 (similar to pRS557 [Simons et al., 1987]) to make a p3-*lacZ* fusion, which was transferred to λ RS45 and used to lysogenise *E. coli* JM101 to give JM-G401. Mutations within the *ptsG* regulatory region were made by oligonucleotide-directed mutagenesis as described elsewhere [El Qaidi and Plumbridge, 2008] and inserted into the fusions in both orientations. Oligonucleotides used in this work are described in table 1. Bacteria were grown in MOPS medium containing 0.5% casamino acids and 0.2% glucose or 0.4% glycerol. β -Galactosidase activities were measured throughout the growth curve, and results are presented for bacteria with A_{650} between 0.5 and 0.8. The plasmids pTSG10 and pTSG20 are derivatives of pBR322 carrying approximately 4- and 10-kb inserts surrounding the *ptsG* gene [Erni and Zanolari, 1986]. JM101 and IBPC1012 (JM101*mlc::tc*) [Plumbridge, 1998b] were used for RNA preparations.

In vitro Methods

DNase I footprinting was carried out as described previously [El Qaidi and Plumbridge, 2008] using the Glc1-Glc4 PCR fragment labelled at Glc4 (fig. 1). The same DNA fragment was used as a template for IVT [Plumbridge, 1998a] and to detect open complex formation by KMnO₄ sensitivity as described elsewhere [Plumbridge and Pellegrini, 2004]. Open complexes between DNA and RNAP, with other proteins as indicated in the figure legends, were formed by incubation for 10 min at 37°C in a buffer containing 20 mM Tris, pH 8.0, 10 mM MgCl₂, 100 mM KCl, 5 mM DTT and 200 μ g/ml BSA. Transcription was initiated by the addition of nucleotides (200 μ M ATP, CTP and GTP and 20 μ M [α -³²P]-UTP) for 5 min and terminated by the addition of formamide loading dyes. The 5' ends of p3 transcripts were detected

by S1 analysis using the Glc5-Glc4 (366 bp), Glc58-Glc4 (540 bp) and Glc59-Glc4 (649 bp) PCR fragments (fig. 3E) with the 5' end of the Glc5, 58 or 59 oligonucleotides labelled with [γ -³²P]-ATP and T4 polynucleotide kinase. The labelled DNA fragments were hybridised with 30 μ g of total RNA overnight at 51°C (Glc5-4) or 54°C (Glc58-4 and 59-4), and then the temperature was slowly reduced over 2 h to 49°C the next morning, before treatment with S1 nuclease for 45 min at 30°C [El Qaidi and Plumbridge, 2008]. Total RNA was prepared by the hot phenol method. Reactions from footprinting, IVT and S1 analysis were analysed on 6% denaturing polyacrylamide gels.

For Northern blotting, RNA was prepared from JM101 carrying the pTSG10 and pTSG20 plasmids [Erni and Zanolari, 1986] or pBR322 and treated with RNase-free DNase I (Promega AQ1) to remove plasmid DNA. Then, 20 μ g of RNA was separated on a 6% polyacrylamide/8 M urea gel and transferred to Hybond N+ membranes. The probe was the Glc5-3 PCR fragment (fig. 1) labelled by random priming, and the hybridisation temperature was between 53°C and 47°C during a period of 24 h in 5 $\times$ SSPE, 5 $\times$ Denhardt's solution, 0.5% SDS and 50% formamide with 0.2 mg/ml yeast tRNA.

Abortive primer extension reactions with 3 dNTPs and 1 ddNTP were performed on the same mRNA preparations using oligonucleotides Glc1, Glc2, Glc9, Glc10 and Glc59 (table 1) labelled with [γ -³²P]-ATP and T4 polynucleotide kinase. Equal amounts of the 5 oligonucleotides (3 pmol) were annealed to 20 μ g of RNA prepared from JM101 carrying pTSG20 or pBR322 and then treated with AMV reverse transcriptase (Finnzymes) for 30 min at 42°C and analysed on a 15% denaturing polyacrylamide gel. For Glc1 and Glc2, the extension reaction contained 0.5 mM dGTP, dCTP, dTTP and 0.08 mM ddATP. For Glc9, Glc10 and Glc59, the dideoxynucleotide was ddTTP. All radioactive gels were analysed by phosphorimager.

Acknowledgements

We are extremely grateful to Annie Kolb for the gift of purified CAP protein and helpful discussion and Marie Cerciati, Alex Maes and Eliane Hajnsdorf for help and advice on Northern blotting. This work was supported by grants from the CNRS and Université Paris 7 to UPR9073.

References

- Aiba H: Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr Opin Microbiol* 2007;10:134–139.
- Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner E, Margalit H, Altuvia S: Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol* 2001;11:1–20.
- Becker A-K, Zeppenfeld T, Staab A, Seitz S, Boos W, Morita T, Aiba H, Mahr K, Titgemeyer F, Jahreis K: Yeel, a novel protein involved in modulation of the activity of the glucose-phosphotransferase system in *Escherichia coli* K-12. *J Bacteriol* 2006;188:5439–5449.
- Berretta J, Morillon A: Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep* 2009;10:973–982.
- Berretta J, Pinskaya M, Morillon A: A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev* 2008;22:615–626.
- Deutscher J, Francke C, Postma P: How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* 2006;70:939–1031.
- El Qaidi S, Plumbridge J: Switching control of expression of *ptsG* from the Mlc regulon to the NagC regulon. *J Bacteriol* 2008;190:4677–4686 (erratum published in *J Bacteriol* 2008;190:5733).

- Erni B, Zanolari B: Glucose-permease of the bacterial phosphotransferase system. Gene cloning, overproduction and amino acid sequence of enzyme II^{Glc}. *J Biol Chem* 1986; 261:16396–16403.
- Görke B, Stülke J: Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* 2008;6:613–624.
- Horler R, Vanderpool C: Homologs of the small RNA SgrS are broadly distributed in enteric bacteria but have changed in size and sequence. *Nucleic Acids Res* 2009;37:5465–5476.
- Jeong J-Y, Kim Y-J, Cho N, Shin D, Nam T-W, Ryu S, Seok Y-J: Expression of ptsG encoding the major glucose transporter is regulated by ArcA in *Escherichia coli*. *J Biol Chem* 2004; 279:38513–38518.
- Kawano M, Reynolds AA, Miranda-Rios J, Storz G: Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res* 2005;33:1040–1050.
- Kimata K, Inada T, Tagami H, Aiba H: A global repressor (Mlc) is involved in glucose induction of the ptsG gene encoding major glucose transporter in *Escherichia coli*. *Mol Microbiol* 1998;29:1509–1519.
- Kimata K, Takahashi H, Inada T, Postma P, Aiba H: cAMP receptor protein-cAMP plays a crucial role in glucose-lactose diauxie by activating the major glucose transporter gene in *Escherichia coli*. *Proc Natl Acad Sci USA* 1997;94:12914–12919.
- Lee S-J, Boos W, Bouché J-P, Plumbridge J: Signal transduction between a membrane bound transporter and a soluble transcription factor, Mlc, of *Escherichia coli*. *EMBO J* 2000;19: 5353–5361.
- Lorenz C, Gesell T, Zimmermann B, Schoeberl U, Bilusic I, Rajkowsch L, Waldsich C, von Haeseler A, Schroeder R: Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominately in antisense transcripts. *Nucleic Acids Res* 2010, in press (doi: 10.1093/nar/gkq032).
- Maki K, Morita T, Otaka H, Aiba H: A minimal base-pairing region of a bacterial small RNA SgrS required for translational repression of ptsG mRNA. *Mol Microbiol* 2010;76:782–792.
- Massé E, Gottesman S: A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA* 2002;99:4620–4625.
- Matos C, Di Cola A, Robinson C: TatD is a central component of a Tat translocon-initiated quality control system for exported FeS proteins in *Escherichia coli*. *EMBO Rep* 2009;10: 474–479.
- Nam T-W, Cho S-H, Shin D, Kim J-H, Jeong J-Y, Lee J-H, Roe J-H, Peterkofsky A, Kang S-O, Ryu S, Seok Y-J: The *Escherichia coli* glucose transporter enzyme IICB^{Glc} recruits the global repressor Mlc. *EMBO J* 2001;20:491–498.
- Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz L, Jacquier A: Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 2009; 457:1038–1042.
- Opdyke J, Kang J-G, Storz G: GadY, a sRNA regulator of acid response genes in *Escherichia coli*. *J Bacteriol* 2004;186:6698–6705.
- Pennetier C, Domínguez-Ramírez L, Plumbridge J: Different regions of Mlc and NagC, homologous transcriptional repressors controlling expression of the glucose and N-acetylglucosamine phosphotransferase systems in *Escherichia coli*, are required for inducer signal recognition. *Mol Microbiol* 2008;67:364–377.
- Plumbridge J: Control of the expression of the manXYZ operon in *Escherichia coli*: Mlc is a negative regulator of the mannose PTS. *Mol Microbiol* 1998a;27:369–381.
- Plumbridge J: Expression of ptsG, the gene for the major glucose PTS transporter in *Escherichia coli*, is repressed by Mlc and induced by growth on glucose. *Mol Microbiol* 1998b;29:1053–1063.
- Plumbridge J, Pellegrini O: Expression of the chitobiose operon of *Escherichia coli* is regulated by three transcription factors: NagC, ChbR and CAP. *Mol Microbiol* 2004;52:437–449.
- Rivas E, Klein R, Jones T, Eddy S: Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* 2001;11:1369–1373.
- Rungrasamee W, Liu X, Pomposiello P: Activation of glucose transport under oxidative stress in *Escherichia coli*. *Arch Microbiol* 2008;190:41–49.
- Selinger D, Cheung K, Mei R, Johansson E, Richmond C, Blattner F, Lockhart D, Church G: RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nat Biotechnol* 2000;18:1262–1268.
- Sharma C, Hoffman S, Darfeuille F, Reignier J, Findeis S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler P, Vogel J: The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 2010;464:250–255.
- Shin D, Cho N, Heu S, Ryu S: Selective regulation of ptsG expression by Fis. *J Biol Chem* 2003; 278:14776–14781.
- Shin D, Lim S, Seok Y-J, Ryu S: Heat shock RNA polymerase (σ^{32}) is involved in the transcription of mlc and crucial for induction of the Mlc Regulon by glucose in *Escherichia coli*. *J Biol Chem* 2001;276:25871–25875.
- Simons RW, Houman F, Kleckner N: Improved single and multicopy lac based cloning vectors for protein and operon fusions. *Gene* 1987;53:85–96.
- Tanaka Y, Kimata K, Aiba H: A novel regulatory role of glucose transporter of *Escherichia coli*: membrane sequestration of a global repressor Mlc. *EMBO J* 2000;19:5344–5352.
- Tjaden B, Saxena R, Stolyar S, Haynor D, Kolker E, Rosenow C: Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res* 2002;30:3732–3738.
- Vanderpool C, Gottesman S: Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol* 2004; 54:1076–1089.
- Vogel J, Bartels V, Tang T, Churakov G, Slagter-Jager J, Hüttenhofer A, Wagner E: RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* 2003;31:6435–6443.
- Vogel J, Sharma C: How to find small non-coding RNAs in bacteria. *Biol Chem* 2005;386: 1219–1238.
- Wadler C, Vanderpool C: Characterization of homologs of the small RNA SgrS reveals diversity of function. *Nucleic Acids Res* 2009; 37:5477–5485.
- Wassarman K, Repoila F, Rosenow C, Storz G, Gottesman S: Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev* 2001;15:1637–1651.
- Wexler M, Sargent F, Jack R, Stanley N, Bogsch E, Robinson C, Berks B, Palmer T: TatD is a cytoplasmic protein with DNase activity. *J Biol Chem* 2000;275:16717–16722.
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Cambong J, Guffanti E, Stutz F, Huber W, Steinmetz L: Bidirectional promoters generate pervasive transcription in yeast. *Nature* 2009;457:1033–1037.
- Yachie N, Numata K, Saito R, Kanai A, Tomita M: Prediction of non-coding and antisense RNA genes in *Escherichia coli* with gapped Markov Model. *Gene* 2006;372:171–181.
- Zhang A, Wassarman K, Rosenow C, Tjaden B, Storz G, Gottesman S: Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbiol* 2003;50:1111–1124.