The use of a tRNA as a transcriptional reporter: the T7 late promoter is extremely efficient in *Escherichia coli* but its transcripts are poorly expressed

Pascal J.Lopez, Isabelle Iost and Marc Dreyfus*

Laboratoire de Génétique Moléculaire, CNRS D1302, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris Cedex 05, France

Received January 1, 1994; Revised and Accepted March 3, 1994

ABSTRACT

In gene expression studies, promoters are often fused to a protein-encoding reporter gene, the expression of which is then taken as an indirect measure of their strength. Here, we advocate the use of a tRNA reporter for the direct quantification of promoter strength. Using this method, we have studied the bacteriophage T7 gene 10 promoter in an *E.coli* strain that produces saturating amounts of T7 RNA polymerase. At 37°C in aminoacid-glycerol medium, we show that this promoter ranks amongst the strongest known, directing *ca* 1.1 transcription events per second, 2.2-fold more than the promoters for rRNA operons, or 15-fold more than the induced *lac* promoter. Surprisingly, compared to the *lac* promoter, the T7 promoter is far less efficient in driving the expression of protein-encoding genes such as *cat*, *neo* or *lacZ*. Therefore, the polypeptide yield per transcript is lower when the T7 RNA polymerase is used instead of the *E.coli* RNA polymerase. The former enzyme travels faster than the translating ribosomes, and we suggest that this desynchronization lowers the polypeptide yield per transcript.

INTRODUCTION

Bacteriophages T3, T7 and SP6 encode closely related RNA polymerases which transcribe the viral genome from highly specific promoters, during the late phase of infection (1). The prototype of the group, the T7 RNA polymerase, can be stably expressed in *E.coli*, and genes that are fused to a T7 late promoter can be expressed to high levels when introduced in such hosts (2, 3). In spite of this practical bearing, relatively little is known about the *in vivo* behaviour of these RNA polymerases. Indirect evidence suggests that the T7 RNA polymerase interacts very efficiently with its promoter in *E.coli* (4), whereas *in vitro* the interaction appears relatively weak (1, 5). The enzyme has a very high turnover for the polymerisation of ribonucleotides, both *in vitro* (6) and *in vivo* (7). As a result, in *E.coli* it runs far ahead of the ribosomes which translate its nascent transcript, a situation

which is quite unusual in this organism (8). It is not known whether this desynchronisation affects gene expression, although we recently hypothesized that it might decrease the number of polypeptides synthesized per transcript by depressing mRNA stability and/or translatability (7). The aim of the present work is to compare the frequencies of transcription from the T7 late promoter and from typical *E.coli* promoters, in an *E.coli* strain expressing the T7 RNA polymerase. Moreover, we also examine the consequences of the high elongation rate of the T7 enzyme upon the polypeptide yield per transcript.

A classical method for comparing the frequencies of transcription from different promoters consist in fusing them to a protein-encoding reporter gene, the expression of which is then taken as an indirect measure of their strength (9). However, this expression actually reflects not only the frequency of transcription, but also the number of polypeptides synthesized per transcript, which obviously cannot be assumed invariant for our present purpose. In order to quantify separately these two parameters, we propose to use two reporters in parallel, i.e. a protein-encoding gene, and a gene encoding a stable RNA, such as a tRNA. Being stable, a tRNA will accumulate much like a protein, but its accumulation does not depend upon a polypeptide yield per transcript, and therefore it will reflect the sole frequency of transcription. A further advantage of such a reporter is that it can yield absolute values for the transcription frequencies from the different promoters, rather than relative ones. This can be achieved by comparing the accumulation of the tRNA with that of rRNAs, the transcription frequency of which is well known under a variety of conditions (10).

We have therefore fused both a protein-encoding reporter gene and a tRNA reporter gene to a bacteriophage T7 late promoter, and these fusions were transferred onto the chromosome of an *E.coli* host expressing the T7 RNA polymerase. In parallel, these same constructs were put under the control of a typical *E.coli* promoter, that of the *lac* operon, in such a way that the mRNAs transcribed from either promoter are identical in sequence. Judging from the accumulation of the reporter tRNA, the T7 promoter appears to be much stronger than the *lac* promoter, and, indeed, than most *E.coli* promoters. However, judging from

*To whom correspondence should be addressed

the expression of the protein-encoding gene, it is weaker, indicating that the polypeptide yield per transcript is lower when the T7 RNA polymerase is used instead of the *E. coli* enzyme. This result indicates that mRNAs that are identical in sequence can nevertheless differ in their stability or translatability, depending upon which polymerase is used to transcribe them. More generally, it strongly advocates the usefulness of a tRNA reporter for the direct and unbiased quantification of transcription frequencies.

MATERIALS AND METHODS

Plasmid construction

Plasmid pMAMA5 is a pUC18 derivative carrying a truncated version of the *lac* operon (from the 8th *lacZ* codon to the middle of *lacY*), in-phase fused downstream of a 51 bp fragment carrying the Ribosome Binding Site (RBS) of the *lamB* gene (11). This construct is bracketed by the T7 gene 10 promoter-leader sequence and by a T7 late transcription terminator ($T\phi$; (3)). In the *Bst*BI site located within the *lacY* sequence of pMAMA5, we inserted a synthetic *Bst*BI fragment encompassing a modified tRNA^{Arg5} gene preceded by a unique *Bgl*II site (Fig. 2). Compared to the genuine tRNA^{Arg5} gene, this gene carries a A to U substitution at position +73. The *Cla*I site lying 93 nt upstream of the *Bst*BI site was then filled in, shifting the *lacY* translation into the -1 reading frame and causing translation termination 88 nt upstream of the tRNA gene. Next, the synthetic sequence 5' TAGCCCGCCTAATGAGCGGGCTTTT 3' encompassing the *trpA* terminator (12) was inserted in the *Xba*I site immediately downstream of the T7 terminator. The extremities of this fragment were so designed that only one *Xba*I site is regenerated downstream of the terminator. Finally, we replaced the *Kpn*I–*Bam*HI fragment encompassing the T7 gene 10 promoter-leader sequence, by a 52 bp synthetic *Kpn*I–*Bam*HI fragment carrying a shortened version of this same promoter (from positions -21 to +2 with respect to the gene 10 natural transcription start) followed by the *lac* leader sequence (from positions +1 to +22 with respect to the major natural *lac* transcription start) (7). This leads to plasmid pTLacZ-Arg5. Alternatively, this same *Kpn*I–*Bam*HI fragment was replaced by the *Pvu*II–*Bam*HI fragment of pEMBLΔ46 (13) carrying the *lac* promoter-leader sequence from nt -122 to +22. Prior to insertion, this fragment was made *Kpn*I–*Bam*HI compatible with a suitable *Kpn*I–*Pvu*II adaptor. Moreover, it was oligo-mutagenized to insert two Cs between nt -3 and -2, which shifts the transcription start by two nt upstream (see below). The final plasmid is pLLacZ-Arg5. Hereafter, pTLacZ-Arg5 and pLLacZ-Arg5 are collectively referred to as the '*lacZ*' plasmids.

We then constructed derivatives of these plasmids in which the *lacZ* gene is replaced by the *neo* gene encoding neomycin phosphotransferase II (NPT II) or the *cat* gene encoding chloramphenicol acetyltransferase (CAT). To this end, most of the *lacY* sequence downstream of the tRNA^{Arg5} gene was first removed (deletion of the fragment *Bst*BI–*Sna*BI corresponding to Genebank Ecolac 4826 to 5199; Fig. 1). The *Bam*HI–*Bst*BI fragment encompassing the RBS and all the *lacZ* and *lacY* sequence was then replaced by the 845 bp *Bgl*II–*Bst*BI fragment from pNEO (Pharmacia). This fragment carries the *neo* coding sequence and 23 nt and 35 nt of downstream and upstream flanking sequences, respectively, including the *neo* RBS. This yields the *neo* plasmids pLNeo-Arg5 and pTNeo-Arg5. Alternatively, the *Bam*HI–*Bgl*II fragment from the *lacZ* plasmids

was replaced by the 977 bp *Sau*3A fragment from pACYC184 (Biolabs), encompassing the *cat* coding sequence and 262 and 59 nt of downstream and upstream flanking sequence, respectively, including the *cat* RBS. This yields the *cat* plasmids pLCat-Arg5 and pTCat-Arg5. The relevant regions of the *lacZ*, *neo* and *cat* plasmids, together with the positions of the restriction sites used in their construction, are depicted on the lower part of Fig. 1.

Strain construction

The *Kpn*I–*Xba*I fragments carrying the *lacZ*, *neo* and *cat* constructs from upstream of the promoters to downstream of the terminators (Fig. 1), were then transferred onto the chromosome of the BL21(DE3) derivatives MO00 or MO20. BL21(DE3) is an *E. coli* B strain harbouring the T7 RNA polymerase gene under the control of the P_{lacUV5} promoter (4). MO00 is a *lacZ*::Tn10 ($Lac^- Tet^r$) derivative of BL21(DE3), whereas MO20 ($Lac^- Tet^s$) was obtained from MO00 by Tn10 excision as in (14). The technique used for the chromosome transfer is described elsewhere (11, 15). Briefly, the shuttle plasmid pOM43 harbours a large DNA fragment originating from the *malA* chromosomal region with a polylinker inserted in the middle, and the constructs were first cloned into this polylinker. Plasmid pOM43 allows an easy selection of those host cells in which the chromosomal *malA* region has been replaced by the plasmid-borne one, thus permitting the transfer of any insert cloned in its polylinker onto the host chromosome. The recombinant *malA* region was first transferred onto the chromosome of strain JM101, and then P1-transduced into MO00 or MO20, using suitable markers (11). This procedure ensures that the final strains are free of unwanted changes outside the *malA* region. For several constructs the whole transfer procedure was repeated twice: the transcription and expression of the constructs in the final strains were identical in each case.

The structure of the promoter/RBS and of the tRNA^{Arg5}/terminator regions was checked in the final strains after PCR amplification of the relevant genomic segments. The transcription startpoint, determined by primer extension, was 5' pppGG-AATT... for the both the T7 and the modified *lac* promoter, instead of 5' pppAATT... for the genuine *lac* promoter. The chromosome structure of the final strains is schematized on Fig. 1, and their current nomenclature for future reference is:

ENS133=BL21(DE3) (*lacZ*::Tn10 *malPp*Δ534:: P_{lac} *lacZ*-Arg5)
 ENS134=BL21(DE3) (*lacZ*::Tn10 *malPp*Δ534:: P_{T7} *lacZ*-Arg5)
 ENS201=BL21(DE3) (Δ*lacZ* *malPp*Δ534:: P_{lac} *cat*-Arg5)
 ENS202=BL21(DE3) (Δ*lacZ* *malPp*Δ534:: P_{T7} *cat*-Arg5)
 ENS203=BL21(DE3) (Δ*lacZ* *malPp*Δ534:: P_{lac} *neo*-Arg5)
 ENS204=BL21(DE3) (Δ*lacZ* *malPp*Δ534:: P_{T7} *neo*-Arg5)

Cell growth and protein assays

Cells were grown at 37°C in MOPS medium containing glycerol (0.2% w/v), inorganic phosphate, vitamins and all amino acids (16). Unless otherwise stated, the medium also included 1 mM IPTG for the inactivation of the *lac* repressor. The generation time at 37°C was ca 40 min whether IPTG was present or not.

For β-galactosidase, NPT II and CAT assays, cells were grown in the presence of IPTG to an OD₆₀₀ of 0.5, chilled, centrifuged, resuspended in cold phosphate-buffered saline and lysed by pulsed sonication. After removal of debris, lysates were assayed for total proteins using the Biorad assay reagent. CAT

and β -galactosidase activities were measured as in (17) and (18), respectively. The NPT II, CAT or β -galactosidase proteins were assayed by Western blotting. Lysates were loaded on 7.5% (for β -galactosidase) or 12.5% (for NPT II and CAT) polyacrylamide-SDS gels (19) together with serial dilutions of the corresponding pure protein. Gels were electroblotted onto nitrocellulose (BA85: Schleicher and Schuell) as in (20), and the membranes were incubated with diluted (1/7500) commercial rabbit antibody directed against the assayed protein. Antibodies and pure proteins were from 5'-3', Boulder, Colorado, except for β -galactosidase (Boehringer, Mannheim). The second antibody was a goat anti-rabbit IgG coupled to horseradish peroxidase (1:7500, Promega). For the CAT or NPT II assays, the antibody solutions were supplemented with 25 μ g/ml of crude proteins extracted from an *E. coli* strain not carrying the *neo* or *cat* genes, to reduce non-

specific binding. The different proteins were detected with the ECL kit of Amersham, and quantified by film densitometry.

RNA isolation and analysis

Samples of the same cultures used for protein assays were rapidly mixed with an equal volume of boiling lysis buffer (5% SDS, 1.5 M sodium acetate, 12 mM EDTA, pH 7.0), and total RNA was extracted as in (21).

RNA samples (5 or 20 μ g) were electrophoresed on 2% agarose-formaldehyde gels, blotted and UV cross-linked onto a nylon membrane (Hybond-N⁺, Amersham) as in (18). For the visualisation of 5S rRNA and tRNA^{Arg5}, the blots were probed and washed as in (22). The two following oligonucleotides were used: 5' ACTACCATCGGCGCTACGGC 3' (complementary to the 5S rRNA from nt 61 to 80); and 5' TGGAGTCCCCTGC-AG 3' ('U73' probe, complementary to the tRNA^{Arg5} from nt 62 to 76, including the U73 mutation; Fig. 2). Oligonucleotides were 5' labelled with polynucleotide kinase (Pharmacia) and ATP- γ -³²P (Amersham) and were used in large excess in all cases. For the U73 probe, stringent washing conditions (56°C, 20 min) were used to eliminate hybridization with the genuine tRNA^{Arg5}. The hybridization signals were quantified with a PhosphorImager System (Molecular Dynamics).

To study the *neo* and *cat* mRNA decay, cells were grown as above, rapidly filtered through 0.45 μ HA circular filters (Millipore) and resuspended in an equal volume of prewarmed IPTG-free medium. At timed intervals, aliquots of the culture were harvested and lysed as above. The *neo* and *cat* mRNAs were visualized as a function of time on Northern blots. In the experiment shown on Fig. 4, the 845 bp *Bgl*III/*Bst*BI fragment from the pNEO plasmid was used as a probe after being ³²P labelled with the BRL multiprime labelling kit. The membrane was processed as in (18).

In vitro transcription

Pure tRNA^{Arg5} carrying the U73 point mutation was synthesized *in vitro* as in (23) using two synthetic oligonucleotides. The first one carried the sequence of the T7 promoter from nt -17 to +1 (upper strand), whereas the second encompassed the sequence complementary to the tRNA^{Arg5} (including the U73 mutation), fused to the lower strand of the T7 promoter (from nt +1 to -17). Conditions for *in vitro* transcription and tRNA purification were as in (24). The concentration of the pure tRNA was determined spectrophotometrically.

Calculations of transcription frequencies

Let p be the proportion of total RNA mass which constitutes the reporter tRNA (76 nt in length). About 85% of the total RNA mass constitutes the 23S, 16S and 5S rRNA (total length 4560 nt) (10). On a molar basis, the ratio of reporter tRNA over the three rRNA is then $(4560p)/(76 \times 0.85) = 71p$. Since all these RNAs are stable, this figure also reflects their relative frequency of synthesis. The constructs used here are located 9 min away from the replication origin, and for the seven rRNA operons this distance averages 11 min, a very similar value. Hence, we can assume that there are 7 rRNA operons per copy of our constructs. Then, per operon, the relative frequency of synthesis of the reporter tRNA vs. that of the three rRNA is $7 \times 71p$, or $497p$. Under the growth conditions used ($t_{1/2}$ ca 40 min), each rRNA operon is transcribed every 2 sec (10), and therefore the reporter is transcribed every $(2/497p)$ or $(4 \times 10^{-3}/p)$ sec.

Figure 1. (upper) Schematic representation of the chromosome from the strains used in this work. Along with the phage lambda DE3 which harbours the T7 RNA polymerase gene (black box) these strains harbour any one of the constructs depicted on the lower part of the figure. These constructs are inserted upstream of the *malP* gene. The endogenous *lacZ* gene has been inactivated (see text), as indicated by the symbol ' Δ *lacZ*'. The symbols + and - stand for positive and negative control by the T7 RNA polymerase and *lac* repressor, respectively. (lower) Structure of the constructions used in this work. P ('Promoter') stands for either the promoter of the bacteriophage T7 late gene 10 (P_{T7}) or the *lac* operon promoter (P_{lac}). Op ('operator') is the *lac* repressor binding site; Ter ('terminators') stands for the two tandemly arranged transcription terminators from the T7 late region and from the *E. coli trpA* gene (see text). The thick, hatched boxes correspond to the coding sequence of the *lacZ-lacY* (upper), *cat* (middle), or *neo* (lower) gene, whereas the thick open boxes stand for the tRNA^{Arg5} gene. Other non coding sequences are shown as thin boxes. Relevant restriction sites are abbreviated as follows; Ba, *Bam*HI; Bg, *Bgl*III; Bs, *Bst*BI; C, *Cla*I; K, *Kpn*I; Sa, *Sau*3A; Sn, *Sna*BI; X, *Xba*I. Those sites that have been destroyed in the course of the constructions are parenthesized.

RESULTS

Design of the experimental system

Our aim is to compare a T7 late promoter and a typical *E. coli* promoter in their capacity to drive the expression of a reporter tRNA gene on one hand, and that of a protein-encoding reporter gene on the other. For this purpose, the two genes were associated in tandem, with the longer protein-encoding gene lying upstream. The bicistronic operon was then fused to either promoter in such a way that the transcribed sequence is identical in each case.

We have used here the promoter from the bacteriophage T7 gene 10 (P_{T7}), and the *E. coli lac* operon promoter (P_{lac}). These promoters are immediately followed by the binding site for the *lac* repressor ('Op' on Fig. 1). This makes P_{T7} , like P_{lac} , inducible by the *lac* operon inducer IPTG (see e.g. (7)). To insure that the transcripts from either P_{T7} or P_{lac} have identical 5' ends, we slightly modified the natural sequence of P_{lac} (see Materials and Methods). A control experiment (O. Makarova, unpublished observation) showed that this modification did not alter the P_{lac} -driven synthesis of β -galactosidase, nor that of the tRNA^{Arg5} (see below).

As protein-encoding genes, we have used three genes commonly used as reporters, i.e. the *lacZ* gene encoding β -galactosidase, the *neo* gene encoding neomycin phosphotransferase (NPT II) and the *cat* gene encoding chloramphenicol acetyltransferase (CAT). All three are preceded by the repressor binding site and followed by the tRNA gene. Due to construction constraints, the *lacZ* construct differs from others in two respects: (i) the *lacZ* gene is not translated from its own RBS but from that of the *lamB* gene, which is nearly equivalent in efficiency (18); and (ii) it is followed by a truncated version of the *lacY* gene, making this construct tri- rather than bicistronic (Fig. 1).

Figure 2. Sequence of the reporter tRNA^{Arg5} gene and of the immediately flanking region. The sequence is written in the RNA form and is folded into the cloverleaf structure. The 5' and 3' ends of the mature tRNA are arrowed. The nt 73 which has been mutated from A into U, is in boldface. The first five nt upstream of the tRNA originate from the natural tRNA^{Arg5} precursor (25). The sequence shown correspond to the *lacZ* construct (Fig. 1). In the *neo* and *cat* constructs, the downstream *Bst*BI site has been blunt-end ligated to a *Sna*BI site located further downstream (Fig. 1), and the fragments harbouring the *neo* and *cat* genes have been cloned upstream from the tRNA, in the *Bgl*II (*cat*) or the *Bst*BI (*neo*) sites, respectively. Therefore, the *neo* and *cat* constructs diverge from the sequence shown starting from the 7th nucleotide following the tRNA 3' end, and from the 11th (*cat*) or the 17th (*neo*) nt preceding the tRNA 5' end. The region of the acceptor arm which is complementary to the U73 probe used for hybridization, as well as relevant restriction sites, are overlined.

The tRNA used here is tRNA^{Arg5} (25), which recognizes the rare arginine codon AGG. This tRNA ranks amongst the least abundant in *E. coli* (26), and therefore its genuine expression interferes minimally with that of the reporter. Moreover, we have introduced an A to U substitution at the 'discriminator' nt 73 (Fig. 2). This nt is not directly involved in secondary or tertiary structure, but the substitution decreases the capacity of the tRNA to load arginine (27), thereby minimizing possible effects of the tRNA overproduction upon cell growth. In addition, the substitution permits the discrimination between the reporter and the genuine tRNA^{Arg5} by hybridization (see Materials and Methods). Noteworthy, our three constructs diverge in sequence starting from 11 and 7 nt upstream and downstream from the mature tRNA^{Arg5} sequence, respectively (cf. Fig. 2 legend).

The constructs end with the major late T7 terminator followed by the strong rho-independent terminator from the *E. coli trpA* gene. The former is moderately efficient upon T7 RNA polymerase (28), whereas the second is very efficient upon the *E. coli* RNA polymerase (12) and moderately so upon the T7 enzyme (28).

All constructs were finally introduced as single copy into the *malA* chromosomal region of a strain expressing the T7 RNA polymerase (11) (Fig. 1). In all cases, the transcription of our constructs can be IPTG-induced without affecting cell growth, and unless otherwise stated, cells were continuously grown with saturating concentrations (1 mM) of this inducer. Hereafter, the qualifiers ' P_{lac} ' or ' P_{T7} ' are used to collectively designate the strains carrying the P_{lac} or the P_{T7} promoters, as well as the corresponding transcripts, mRNA patterns, etc. Similarly, a qualifier such as ' P_{T7} -*neo*' shortly designates a strain carrying the *neo* gene under the control of P_{T7} .

Figure 3. (A) Autoradiogram of a Northern blot showing the probing of total RNA from different strains with the U73 probe (Fig. 2). The protein-encoding gene and the promoter present in each IPTG-induced strain are indicated above the corresponding lane. The control lane corresponds to the IPTG-induced parental strain BL21(DE3). 5 μ g of total RNA were loaded on each lane. The four rightmost lanes correspond to the indicated amounts (ng) of pure reporter tRNA synthesized *in vitro*. Aside from the mature tRNA (76 nt), unmaturing precursors are also visible. Their positions are arrowed and their approximate size (nt) are indicated. (B) Same as (A), except that the membrane was stripped and reprobed with an oligonucleotide complementary to the 5S rRNA.

The expression of the reporter tRNA

We used Northern blots to assay the reporter tRNA^{Arg5} expression in exponentially growing *P_{lac}* and *P_{T7}* cells. For calibration purposes, known amounts of pure, *in vitro* synthesized tRNA^{Arg5} carrying the U73 mutation were loaded on the same gel. As a probe, we used a 15-mer oligonucleotide complementary to the reporter tRNA around the U73 mutation ('U73' probe; Fig. 2). Under stringent washing conditions, the probe did not hybridize with the RNA from the parental strain BL21(DE3) (Fig. 3A, control lane), or from *P_{lac}* or *P_{T7}* strains grown without IPTG (not shown). In contrast, it detected a species corresponding in size to a mature tRNA in IPTG-induced *P_{lac}* or *P_{T7}* strains, indicating that the reporter tRNA is transcribed and matured. The intensity of this band was far stronger in *P_{T7}* than in *P_{lac}* strains (Fig. 3A). In addition, unmaturing tRNA precursors were readily detected in RNA extracted from the *P_{T7}-cat* and *P_{T7}-neo* strains. In either case, the highest molecular weight species migrated as expected for the full-length operon transcript. The additional precursor species, which are about 0.25 kb in length (for *neo*) or 0.4 kb in length (for *cat*), must consist in relatively stable decaying intermediates encompassing both the tRNA and the 3' region of the *neo* or *cat* constructs. Upon longer exposures, the same precursors were also seen in the *P_{lac}-neo* and *P_{lac}-cat* strains; in addition, a band corresponding to the full length *P_{T7}-lacZ* transcript became visible.

The hybridization signals were then quantified. In all lanes, the hybridization signal from the mature tRNA exceeded 92% of the total. This total signal was then compared to the signals from samples of pure tRNA^{Arg5} in order to estimate the amount of reporter tRNA present. The membrane was then stripped and reprobed with an oligonucleotide complementary to the 5S rRNA (Fig. 3B). The 5S hybridization signals were used to estimate the actual total RNA loading in each lane, assuming that the averaged signal corresponds to the nominal loading. Finally, by combining the RNA loading and the amount of reporter tRNA in each lane, we evaluated the proportion of total RNA mass which consists of reporter tRNA. In *P_{T7}* strains, irrespectively of which protein-encoding gene is present, this proportion is $(0.44 \pm 0.04)\%$. In the *P_{lac}-neo*, *P_{lac}-lacZ*, and *P_{lac}-cat* strains, this figure is 15-, 20- and 32-fold lower, respectively (Fig. 3 and Table I). Therefore, in contrast to *P_{T7}* strains, the expression of the tRNA reporter varies amongst *P_{lac}* strains, being distinctly

lower in the *P_{lac}-cat* strain than in the other two. The rationale for this difference is discussed below.

Using the same probe, we also checked that the mature reporter tRNA^{Arg5} is stable. The *P_{T7}-cat* strain was IPTG-induced at zero time, and aliquots of the culture were subsequently assayed for reporter tRNA concentration. Compared to total RNA, this concentration rose steadily with time and did not reach a steady-state even after one generation (not shown), indicating that the chemical tRNA lifetime exceeds the doubling time (29).

The kinetics of tRNA excision

In addition to the U73 probe, we used fragments encompassing the *lacZ*, *cat*, and *neo* coding sequences to probe the cognate transcripts. As shown in Fig. 4, when the *P_{T7}-neo* transcript was probed with the whole *neo* gene, the major species detected was the full-length molecule retaining the tRNA (Fig. 4). Similar observations were made with the *P_{lac}-neo* as well as with *P_{lac}*- and *P_{T7}-cat* strains. In the *P_{T7}-lacZ* strain, only the full-length and a trimmed transcript encompassing the proper *lacZ* mRNA were seen (O. Makarova *et al.*, in preparation). Thus, no processed derivative of the full-length species that is severed from the tRNA ever accumulated, implying either that the excision of the tRNA from the full-length transcript is slower than some alternative nucleolytic attack, or that it yields short-lived products. The former view is consistent with the occurrence of decaying intermediates retaining the tRNA (Fig. 3). In any event, the lifetime of the full-length species sets a minimum time for the tRNA excision from this species.

We then measured the lifetime of the full-length *P_{T7}-neo* and *P_{T7}-cat* transcripts, taking advantage of the fact that *P_{T7}* can be rapidly switched off by removing IPTG from the medium. The subsequent decay of the *P_{T7}* transcripts was monitored on Northern blots. The full-length *P_{T7}-neo* and *P_{T7}-cat* transcripts decayed exponentially with half-lives of (0.6 ± 0.1) and (1.0 ± 0.05) min, respectively (Fig. 4 and results not shown). Therefore, the excision of the tRNA from these species is not a fast process compared to the decay of typical *E. coli* mRNAs.

Table I. Expression of the tRNA^{Arg5} gene and of the protein-encoding genes in the different strains^a

Promoter	Expression of the tRNA ^b			Expression of protein-encoding gene ^c		
	<i>lacZ</i>	<i>cat</i>	<i>neo</i>	<i>lacZ</i> ^d	<i>cat</i> ^e	<i>neo</i>
<i>P_{T7}</i>	44.0	48.0	42.0	5	16	0.75
<i>P_{lac}</i>	2.2	1.5	2.8	10	2.7	0.25
Ratio <i>P_{T7}/P_{lac}</i>	20	32	15	0.5	6.0	3.0

^a The values indicated are averaged from two to five determinations, which are generally within 10% of the mean. ^b Proportion of the total RNA ($w/w \times 10^4$) which consist in reporter tRNA^{Arg5}. The tRNA^{Arg5} was quantified vs. known quantities of *in vitro*-synthesized tRNA. ^c Proportion in total protein as estimated from Western blots ($w/w \times 103$). ^d The averaged β -galactosidase activities from the same cultures were 2100 for *P_{T7}* and 4000 for *P_{lac}* (in nmol ONPG hydrolyzed per min and per mg of total protein). ^e The averaged CAT activities from the same cultures were 100 for *P_{T7}* and 13 for *P_{lac}* (in arbitrary units).

Figure 4. (A) Northern blot showing the decay kinetics of the full-length *P_{T7}-neo* transcript (1150 nt). After IPTG removal, samples were harvested at the times indicated above each lane. Total RNA loading was 20 μ g per lane. The probe used here encompasses the *neo* gene (see Materials and Methods). (B) Rehybridization of the same membrane with the 5S probe (see Fig. 3B).

Figure 5. Western analysis of β -galactosidase (upper), CAT (middle) and NPT II (lower panel) in crude extracts of the cognate P_{lac} and P_{T7} strains. Proteins were electroblotted onto nitrocellulose, reacted with specific antibodies, and visualized using enhanced chemiluminescence (see text). The position of each protein is arrowed on the left side of the corresponding panel. The starred arrow in the middle panel points out an unidentified *E. coli* antigen which binds the anti-CAT antibody. The exact strain used (P_{T7} or P_{lac}), and the total load in crude proteins (in μ g) are indicated above each lane. In lanes 'C' (control), we loaded extracts of each P_{T7} strains grown without IPTG. The rightmost part of each panel shows serial dilutions of the pure cognate protein, with the amounts loaded (ng) indicated above each lane.

Protein synthesis from P_{T7} and P_{lac} transcripts

The expressions of *lacZ*, *neo* and *cat* gene products in the respective P_{lac} and P_{T7} strains were next assayed on Western blots, using ECL detection (see Materials and Methods). In order to allow an absolute quantification from the ECL signals, known amounts of the pure polypeptides were loaded in parallel on the same gels (Fig. 5). To validate this quantification, we also measured the CAT and β -galactosidase activities in extracts of the corresponding P_{lac} and P_{T7} strains. These activities paralleled the abundance of the corresponding polypeptides on Western blots (Table I).

As expected, the P_{lac} and P_{T7} constructs yielded no protein in the absence of IPTG ('C' lanes in Fig. 5 and results not shown). Even in the presence of saturating IPTG concentrations, these proteins usually constituted less than 1% of the total proteins (w/w) (Table I). Thus, the protein expression from our single copy constructs causes little drain on the cell resources, consistent with its negligible effect upon the growth rate. However, a most striking result arose when these expressions were compared in P_{lac} and P_{T7} strains. As shown in Table I, compared to the cognate P_{lac} strains, P_{T7} strains consistently accumulate less protein than tRNA. The P_{T7} strain harbouring the *lacZ* gene is most impressive in this respect, since the level of β -galactosidase expression is twice lower than in the corresponding P_{lac} strain, whereas the level of tRNA expression is 20-fold higher. Insofar as the latter reflects faithfully the level of transcription (see Discussion), these results show that the genes under P_{T7} control

Figure 6. The NPT II protein (upper), and the reporter $tRNA^{Arg5}$ (lower) were analyzed in extracts of the $P_{T7}neo$ strain as in Fig. 5 and 3, respectively, except that the cells were grown with the IPTG concentrations shown, instead of the standard one (1 mM).

yield fewer polypeptides per transcript than their P_{lac} counterparts. Therefore, the expressions of the *lacZ*, *cat* and *neo* genes from P_{lac} and P_{T7} do not faithfully report their relative frequencies of transcription from these two promoters.

The low polypeptide yield from P_{T7} transcripts presumably cannot reflect an overload of the cell resources for protein synthesis, since the cell growth is unaffected by the induction of P_{T7} . To confirm this point, we examined the polypeptide yield per P_{T7} transcript in the presence of subsaturating IPTG concentrations. As judged by the accumulation of the reporter tRNA in the $P_{T7}neo$ strain, the level of transcription from P_{T7} decreased markedly when the concentration of IPTG dropped below 0.1 mM. However, the NPT II expression faded in proportion, implying that the polypeptide yield per transcript is alike whatever the transcription level (Fig. 6).

DISCUSSION

The use of a tRNA as a transcriptional reporter

Ideally, the expression of a transcriptional reporter should reflect only the transcription frequency of the gene to which it is fused. It should not depend upon the growth conditions or upon the exact sequence of the operon — insofar as they do not affect the transcription frequency — nor upon the stability of the transcript. The following arguments initially advocated the use of tRNAs as transcriptional reporters. The tRNAs can adopt their compact, nuclease-proof folding even within their precursors (30), and presumably this folding takes place in a very short time compared to the breakdown of the unfolded sequence (31). The sequences flanking the tRNAs are not conserved in natural tRNA precursors, and hence they should be largely irrelevant to the tRNA maturation, even though they can affect its rate (32). Hence, a tRNA embedded within a large transcript is likely to be ultimately matured and recovered from it, with the only possible caveat that, for certain transcripts, alternate structures involving the flanking sequences might be kinetically favoured over the regular tRNA folding, resulting in tRNA degradation (30). Finally, once matured, tRNA molecules are stable, and therefore their accumulation reflects only their frequency of synthesis, together with the cell doubling time.

On these grounds, we tested the usefulness of a tRNA to measure transcription frequencies. The gene encoding $tRNA^{Arg5}$, a very minor arginine isoacceptor in *E. coli*, was fused downstream of either the P_{T7} or the P_{lac} promoter. In

between, we inserted either of three protein-encoding genes (*lacZ*, *neo* or *cat*), and the constructs were so designed that the sequence flanking the tRNA was different in each case (Fig. 1 and 2). The following argument suggests that in these constructs the expression of the tRNA reflects the frequency of its transcription. This expression was alike in all P_{T7} strains. Since, in the absence of specific terminators the T7 RNA polymerase is highly processive *in vivo* whatever the level of translation (4, 11, 33), we expect the reporter tRNA to be equally transcribed in these strains. In the three P_{lac} strains, the tRNA expression spans a twofold range, ranking in the increasing order *cat* < *lacZ* < *neo*. The *cat* gene is unique in being separated from the tRNA^{Arg5} gene by a long (269 nt) untranslated spacer (Fig. 1), and the *E. coli* RNA polymerase is prone to premature termination when transcribing such sequences (polarity; (34)). A weak natural polarity also exists within translated sequences, and in the very long *lacZ* gene it is large enough to become detectable (35). In contrast, polarity should be minimal in the *neo* construct, which is both short and virtually devoided of noncoding sequence. Therefore, amongst both the P_{T7} and the P_{lac} strains, the variations of the tRNA^{Arg5} expression parallel the expected variations of the primary transcription.

A straightforward explanation for this parallelism is that each primary transcript ultimately yields a mature tRNA. This assumption is supported by the fact that it yields absolute transcription frequencies that agree with independent estimates. In the P_{lac} -*neo* strain, the reporter tRNA accounts for about 2.8 10–2% of the total RNA mass. This corresponds to one reporter tRNA synthesized every 14 sec on each copy of the P_{lac} -*neo* construct (see Materials and Methods for the detail of the calculation). This figure agrees reasonably with a former estimate of the transcription frequency from P_{lac} under similar growth conditions (a frequency of *ca* 1 transcription event per 13 sec can be calculated from (36)).

The above data indicate that, in our constructs, the reporter tRNA is quantitatively recovered from the primary transcript, whatever the tRNA flanking sequences. It is noteworthy, however, that the excision is not particularly rapid. Despite the fact that they harbour the excisable tRNA, the full-length P_{T7} -*cat* and P_{T7} -*neo* transcripts are not unusually short-lived compared to other *E. coli* mRNAs (see in particular (37)). It is possible that, when embedded within the large primary transcript, the reporter tRNA is inaccessible to the maturation machinery (32, 38) and can only be matured after the transcript has largely decayed via independent routes (30).

In summary, the fusion of a tRNA tag to a gene of interest can provide a direct measurement of its transcription frequency, presumably irrespectively of the stability or translatability of the transcript. In fact, this may be the only practical way of quantifying the synthesis of very unstable RNAs.

The strength of the P_{T7} promoter

In P_{T7} strains, the reporter tRNA accounts for about 0.44% of the total RNA mass. In control experiments, we introduced into the same strains the multicopy plasmid pAR1219, which carries the T7 RNA polymerase gene (39). This yields a marked overexpression of the T7 RNA polymerase (as judged by denaturing gel electrophoresis), but not of the reporter tRNA (O. Makarova *et al.*, in preparation). Hence, in P_{T7} strains, P_{T7} is saturated with its cognate polymerase and the tRNA is maximally transcribed.

Following the calculation given in Materials and Methods, we find that in P_{T7} strains the reporter is transcribed 1.1-fold per sec, or 2.2-fold as frequently as each rRNA operon. Hence, P_{T7} ranks amongst the strongest promoters known in *E. coli*. This high *in vivo* efficiency is consistent with its biological role. During the few minutes that separate the onset of late transcription from that of replication, it must drive enough gene 10 mRNA synthesis to provide a progeny of several hundred virions with its complement of capsid protein (40). In contrast, *in vitro*, P_{T7} appears much weaker, since a lag of *ca* 10–15 sec is observed before transcription initiates (5). The reason for this difference is unknown. However, Schneider *et al.* (41), noted that the sequence conserved amongst the T7 late promoters is more information-rich than needed for the specific binding of the T7 RNA polymerase. They hypothesized that these promoters contain a binding site for an additional unknown factor. Conceivably, such binding might increase the efficiency of P_{T7} *in vivo*.

Inefficient protein expression from P_{T7} transcripts

The synthesis of the *neo*, *lacZ* and *cat* genes products are 3-, 0.5- and 6-fold higher in the P_{T7} strains than in their cognate P_{lac} strains. In contrast, as judged from the reporter tRNA accumulation, the corresponding transcription frequencies are 15-, 20- and 32-fold higher (Table I). This latter value may actually overestimate the ratio in *cat* transcription between the two promoters, because some P_{lac} -*cat* transcripts may not extent into the tRNA (polarity; see above). Nevertheless, it is fairly clear that each P_{T7} transcript yields fewer polypeptides than its P_{lac} counterpart, particularly for the P_{T7} -*lacZ* transcripts. Yet, P_{lac} and P_{T7} transcripts are expressed under the same physiological conditions and they should be identical in sequence (except possibly at their 3' end, due to the different susceptibility of the *E. coli* and T7 polymerases to the terminators used). What, then, causes the low polypeptide yield from P_{T7} transcripts?

This low yield is not due to an overloading of the translational machinery by the abundant P_{T7} transcripts (cf. Fig. 6). Neither is it related to the particular design of our constructs: we show elsewhere, in a study of the stability and translatability of the P_{lac} -*lacZ* and P_{T7} -*lacZ* transcripts, that the β -galactosidase expression is alike whether the tRNA is present or not (I.I. and M.D., in preparation). It is also unlikely that the T7 RNA polymerase is extremely error-prone under our conditions, resulting in misincorporations, frameshifting of premature termination during translation, and ultimately in a low yield of stable and functional polypeptides. Indeed, Western blots yielded no evidence of a specific accumulation of truncated peptides in P_{T7} strains (Fig. 5). A more likely explanation, then, is that the poor protein yield from P_{T7} transcripts originates from the high elongation rate of the cognate polymerase. *In vivo*, this enzyme travels about 8-fold faster than the *E. coli* RNA polymerase, or than the translating ribosomes (7). Therefore, behind the polymerase, the nascent P_{T7} transcript is ribosome-free. This might reduce its chemical and hence its functional lifetime by facilitating the attack of endonucleases (42). Alternatively, the naked, rapidly growing transcript might fold into complex structures, eventually involving the RBS. Should this happen, either the frequency of translation of the mRNA or its functional lifetime might be reduced, depending upon whether the RBS sequestration is reversible or not (43). These different effects are nonexclusive and either of them may predominate in the *lacZ*,

cat and *neo* transcripts. It is noteworthy that when P_{T7} is used *in vivo* to transcribe rRNA operons, the resulting rRNA is poorly incorporated into ribosomes (44). This low functionality has also been ascribed to the high elongation rate of the T7 RNA polymerase.

The present study suggests that, in the expression systems that use the T7 RNA polymerase (2, 3), the transcripts encoding the gene of interest may be utilized inefficiently. Presumably, these systems owe their success to the very high frequency of transcription from P_{T7}.

ACKNOWLEDGEMENTS

We are much indebted to O. Makarova and L. Coscoy for their participation in several experiments, and to O. Raibaud, J. Guillerez, R. Buckingham and E. Makarov for reviewing the manuscript. We are also grateful to R. Buckingham and his colleagues for the unlimited access to the PhosphorImager. This work has been partly funded by grants from ANRS and from the EEC program 'Capital Humain et Mobilité' to M.D. P.J.L. and I.I. are recipients of long-term fellowships from the Ministère de la Recherche et de l'Espace.

REFERENCES

- Chamberlin, M. and Ryan, T. (1982) *The Enzymes* **15**, 87–108.
- Tabor, S. and Richardson, C.C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1074–1078.
- Rosenberg, A.H., Lade, B.N., Chui, D.-S., Lin, S.-W., Dunn, J.J. and Studier, F.W. (1987) *Gene* **56**, 125–135.
- Studier, F.W. and Moffatt, B.A. (1986) *J. Mol. Biol.* **189**, 113–130.
- Chamberlin, M. and Ring, J. (1973) *J. Biol. Chem.* **248**, 2235–2244.
- Golomb, M. and Chamberlin, M. (1974) *J. Biol. Chem.* **249**, 2858–2863.
- Iost, I., Guillerez, J. and Dreyfus, M. (1992) *J. Bacteriol.* **174**, 619–622.
- Jacquet, M. and Kepes, A. (1971) *J. Mol. Biol.* **60**, 453–472.
- Silhavy, T.J. and Beckwith, J.R. (1985) *Microbiol. Rev.* **49**, 398–418.
- Bremer, H. and Dennis, P.P. (1987) In F. C. Neidhart, J. L. Ingraham, B. K. Low, B. Magasanik, M. Schaechter and H. E. Umbarger (ed.), *Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology*. American Society for Microbiology, Washington, pp. 1527–1542.
- Chevrier-Miller, M., Jacques, N., Raibaud, O. and Dreyfus, M. (1990) *Nucleic Acids Res.* **18**, 5787–5792.
- Christie, G.E., Farnham, P.J. and Platt, T. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4180–4184.
- Dreyfus, M. (1988) *J. Mol. Biol.* **204**, 79–94.
- Maloy, S. and Nunn, W.D. (1981) *J. Bacteriol.* **145**, 1110–1112.
- Raibaud, O., Mock, M. and Scharz, M. (1984) *Gene* **29**, 231–241.
- Neidhart, F.C., Bloch, P.L. and Smith, D.F. (1974) *J. Bacteriol.* **119**, 736–747.
- Gorman, C.M., Moffat, L.F. and Howard, B.F. (1982) *Mol. Cell Biol.* **2**, 1044–1051.
- Yarchuk, O., Jacques, N., Guillerez, J. and Dreyfus, M. (1992) *J. Mol. Biol.* **226**, 581–596.
- Laemmli, U.K. (1970) *Nature* **227**, 680–685.
- Harlow, E. and Lane, D. (1988) In E. Harlow and D. Lane (ed.), *Antibodies, A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 471–511.
- Uzan, M., Favre, R. and Brody, E. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8895–8899.
- Emilsson, V. and Kurland, C.G. (1990) *EMBO J.* **9**, 4359–4366.
- Milligan, J.F., Groebe, D.R., Witherell, G.W. and Uhlenbeck, O.C. (1987) *Nucleic Acids Res.* **15**, 8783–8798.
- Sampson, J.R. and Uhlenbeck, O.C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1033–1037.
- Komine, Y., Adachi, T., Inokuchi, H. and Ozeki, H. (1990) *J. Mol. Biol.* **212**, 579–598.
- Emilsson, V., Näslund, A.K. and Kurland, C.G. (1993) *J. Mol. Biol.* **230**, 483–491.
- McClain, W.H., Foss, K., Jenkins, R.A. and Schneider, J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9260–9264.
- Macdonald, L.E., Zhou, Y. and McAllister, W.T. (1993) *J. Mol. Biol.* **232**, 1030–1047.
- Belasco, J.G. and Brawerman, G. (1993) In J. G. Belasco and G. Brawerman (ed.), *Control of messenger RNA stability*. Academic Press, Inc., San Diego, California, pp. 475–493.
- Smith, J.D. (1976) *Prog. Nucleic Acid Res. Mol. Biol.* **16**, 25–73.
- Crothers, D.M., Cole, P.E., Hilbers, C.W. and Shulman, R.G. (1974) *J. Mol. Biol.* **87**, 63–88.
- Altman, S., Kirsebom, L. and Talbot, S. (1993) *FASEB J.* **7**, 7–14.
- Studier, F.W. (1972) *Science* **176**, 367–376.
- Platt, T. (1986) *Annu. Rev. Biochem.* **55**, 339–372.
- Jürgensen, F. and Kurland, C.G. (1990) *J. Mol. Biol.* **215**, 511–521.
- Deuschle, U., Kammerer, W., Gentz, R. and Bujard, H. (1986) *EMBO J.* **5**, 2987–2994.
- DeFranco, C. and Schottel, L. (1989) *Nucleic Acids Res.* **17**, 1139–1157.
- Bacher Reuven, N. and Deutscher, M.P. (1993) *FASEB J.* **7**, 143–148.
- Davanloo, P., Rosenberg, A.H., Dunn, J.J. and Studier, F.W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 2035–2039.
- Moffatt, B.A. and Studier, F.W. (1987) *Cell* **49**, 221–227.
- Schneider, T.D., Stormo, G.D., Gold, L. and Ehrenfeucht, A. (1986) *J. Mol. Biol.* **188**, 415–431.
- Morse, D.E. and Yanofsky, C. (1969) *Nature* **224**, 329–331.
- Petersen, C. (1992) *Mol. Microbiol.* **6**, 277–282.
- Lewicki, B.T.U., Margus, T., Remme, J. and Nierhaus, K.H. (1993) *J. Mol. Biol.* **231**, 581–593.