Transcription *in vivo* directed by consensus sequences of *E.coli* promoters: their context heavily affects efficiencies and start sites

Marie-Ange Jacquet* and Claude Reiss

Institut Jacques Monod, CNRS-Université Paris VII, tour 43, 2 Place Jussieu, 75251 Paris Cedex 05, France

Received December 14, 1989; Revised and Accepted January 29, 1990

ABSTRACT

We studied *in vivo* transcription and gene expression directed by a series of synthetic sequences, bearing the consensus hexamer (CH) pair of *E.coli* promoters in various contexts. The results demonstrate that, for the contexts tested, the CH pair supports transcription activity and gene expression, whether the spacer linking them is AT or GC rich, or is as short as 14 bp or as large as 26 bp (standard size 17 bp). However, we find that the context influences transcription efficiency by as much as an order of magnitude, and is able to scatter transcription start sites over a region of as much as 30 bp, including start sites within a CH or even between the two sequences of the CH pair. The results demonstrate that, although the CH pair can be sufficient for directing transcription by *E.coli* RNAP, important determinants for promoter activity are at least in part contained in the context of the consensus sequences; they advocate a synergic interplay of signals borne by the CH pair and its context, extending over all parts of the promoter sequence. A two-step model is proposed, in which properly located consensus sequences provide RNAP with facilities required for stereospecific docking along the promoter sequence; the result would be a sharp change in the local environment of the double helix inducing local isothermal unwinding. The size of the loop (related to the AT constraint in the promoter) and the extend of the environmental change required for unwinding would determine the rate of transcriptionally competent complex formation, positioning and grouping of start sites.

INTRODUCTION

The primary step in gene expression is transcription initiation, a process controlled by a series of instructions carried by a DNA sequence—the promoter—to which the transcription machinery (RNA polymerase and factors) binds specifically. The instructions specify a *minima* complexation of RNAP to DNA, activation of

the complex, the site(s) at which transcripts start, and the frequency of transcription initiation.

The basic mechanisms of transcription initiation are generally assumed to be similar in prokaryotic and eukaryotic systems. In the latter, however, several in-depth investigations have revealed highly sophisticated, often multi-level regulation loops, so that the comparatively simpler bacterial systems are more likely to disclose fundamental aspects of the process.

E.coli is especially well-suited for this purpose, as (i) one and the same enzyme, σ^{70} RNA Polymerase (RNAP), transcribes almost all *E.coli* genes; RNAP is a well-characterized multimeric enzyme, which can be purified in large amounts and initiates transcription *in vitro* with high fidelity and efficiency; and (ii) about one-tenth of the promoters of the *E.coli* system have been identified and sequenced.

Sequence comparison (1) disclosed two hexameric homology regions, complying more or less flexibly to the so-called consensus hexamers (CH), TTGACA (CH(-35)) and TATAAT (CH(-10)), located respectively about 35 and 10 bases upstream of the transcription start (+1). Although no base in any CH is strictly conserved, wild type (wt) promoters usually carry at least half of the consensus bases at the appropriate places; also, as a rule, properly located mutations, increasing homology to the CH are usually of the promoter—'up' type and vice-versa (2). Furthermore, transcription is supported by artificial promoters specified by a CH pair spaced by 17 bases, obtained either by fusion of fragments of wt promoters (tac promoter (3)) or by appropriate mutations of some wt promoter (4,5); fully synthetic sequences conforming to these specifications are also capable of directing good gene expression *in vivo* (6).

These observations raise three main questions:

- (i) what precise promoter instruction(s) is (are) carried by each CH?
- (ii) what role is played in promoter activity by the sequences surrounding CHs?
- (iii) to what extent can the CH sequences be altered and still sustain promoter activity, i.e. the role of place and type of base(s) altered, and of the context of the altered CH? From the answers, we expect to understand how promoter instructions are encoded and how they are executed.

* To whom correspondence should be addressed

We designed a flexible tool to investigate these points *in vivo* (6). A sequence was drawn, devoid of any homology to *E. coli* promoters, bearing a series of restriction sites not found in pBR329, bounded by *Hind III* and *Bam HI* sites. The latter enable insertion in pBR329, where *Hind III* overlaps the transcription start site of the chloramphenicol acetyl transferase (*cat*) gene (7). Location of the restriction sites was selected such that, on the non-template strand relative to the *cat* gene, sequences could be introduced or removed at a chosen distance from the wt *cat* transcription start.

In the present contribution, we address questions (i) and (ii) above. To this end, the pair of CHs was introduced in the sequence, with the characteristic orientation relative to the *cat* transcription start (CH(−10) closest to the latter). Constructs differing in size and sequence of the spacer between the CH pair, as well as in the distance between CH(−10) and the wt *cat* transcription start, were tested *in vivo* for transcription activity (relative amount of transcripts and *cat* protein per plasmid copy) transcription initiation site and phenotype (chloramphenicol resistance of the host cell).

The results indicate that a sequence including the CH pair in the proper orientation and relative position can indeed act as a promoter, but the amount of transcript can be modulated by as much as a factor of 10 by the surrounding sequences, and transcription start sites displaced from the conventional place, 8 bases downstream of CH(−10), by as much as 14 bases upstream to 27 bases downstream of the wt transcription start site.

MATERIAL AND METHODS

Bacterial strains

E. coli strain HB101 (7) was used as host for transformation experiments. Transformed cells were selected on LB-agar plates, containing 50 µg/ml ampicillin and 30 µg/ml chloramphenicol.

Plasmids

The plasmids used were wild-type pBR329 and the synthetic plasmid series pBRAT and pBRGC derived from pBR329 (sequences shown) (fig.1). The construction of the synthetic plasmids was described in (6).

Their stability was controlled periodically by the Maxam and Gilbert sequencing procedure (8).

Extraction and purification of total RNA synthesized *in vivo*

Total RNA was prepared from exponentially growing cells subjected to quick chilling as described in (9).

Mapping and quantification of the transcription starts of synthetic promoters

Two synthetic 20-mer DNA primer sequences +27 to +47 and +82 to +102 (assuming transcription start (+1) of wt *cat* gene in pBR329 in the middle of the *Hind III* restriction site (7)) were synthesized by the phosphotriester method in a SAM1 (Biosearch) DNA synthesizer. The HPLC purified primers were end-labeled with γ^{32} -P-ATP using T4 polynucleotide kinase (Biolabs).

The determination of the transcription start directed by wild-type and synthetic promoters was performed as described in (6) by extension of the synthetic primer with AMV reverse transcriptase (Boehringer Mannheim).

Extended products were separate in polyacrylamide gels and visualized by autoradiography. Exposure times were chosen such that the FUJI X ray film response remained approximately linear. Autoradiograms were quantified using a model CS930

Fig. 1: Quantitative determination of transcription start sites *in vivo* by primer extension with reverse transcriptase. a: location and orientation of *cat*, *tet* and *bla* gene in pBR329; the strategy used for construction of the pBR -GC and -AT plasmids is sketched schematically. b: sequences of the synthetic promoters with aligned consensus hexamers (bolt); deletions are marked by (−), inserted or substituted bases are underlined; insertion point is indicated by vertical arrow (↓). Asterisk (*) indicates transcription start site observed *in vivo*; horizontal arrow (→) gives the direction of transcription. The number beneath indicates the % transcript starting at each site (bolt for major site).

Shimadzu scanning densitometer which provided the area under each scanned peak (total band intensity), taken as a relative measure of the transcription efficiency of the promoters at the various start sites. For each plasmid, the primer extension experiments were repeated several times (up to 10 times for those promoters initiating transcription at 'unconventional' places). the

percentage of transcripts starting at a given place are averages from estimations made in individual experiments.

Quantification of *cat* mRNA synthesized *in vivo* by filter binding assay

We followed the method given in (10). The labeled DNA primer used was the 20-mer, sequence +27 to +47 described in the previous section. The nitrocellulose membranes were exposed to Fuji X-ray film at -70°C for 48 h to localize the spots. The relative concentration of specific mRNA was measured by liquid scintillation counting of the excised membrane areas.

Plasmid copy-number determination

Two different methods were used to prepare and quantify plasmid DNA.

a) Determination from unfractionated cell lysates.

The assay method was modified from (11). Cultures were grown exponentially to an optical density at 600 nm, $A_{600} = 0.6$; 10 ml samples were centrifuged (5 min; 7500 rpm) and the bacterial pellets were resuspended in 0.12 ml RNase-lysosome-TE buffer (0.2 mg/ml RNase, 2 mg/ml lysozyme, 10 mM Tris HCl, pH 8, 20 mM $\text{Na}_2\text{-EDTA}$, pH 8) and incubated 15 min at room temperature. The bacteria were then lysed by addition of 0.12 ml 2% SDS, followed by incubation for 15 minutes at 65°C and vigorous shaking. 20 μl of this lysate were added to 10 μl of loading solution (60% sucrose, 0.25% bromophenol blue, 2% SDS in 20% ethanol).

5 μl (for quantify chromosomal DNA) and 25 μl (for quantify plasmid DNA) of this mixture were loaded in two wells on horizontal 1% agarose gels (10 cm long, 5 mm thick). After staining with 0.5 $\mu\text{g/ml}$ ethidium bromide for 15 min and destaining in distilled water for 60 min, the fluorescent DNA bands were photographed over a transilluminator (300 nm wavelength), using Polaroid film type 55 (Polaroid Corp., Cambridge, Mass.) for negatives.

0.1 μg purified supercoiled DNA was loaded on the gel as calibration standard. The photographic negatives from the electrophoresis gel containing different DNA samples and standard were scanned. The surfaces of peaks in the densitometry profile (corresponding to DNA bands) were measured with the aid of an electronic integrator (Shimadzu scanner mod. CS930) and calibrated with DNA bands containing known amounts of DNA.

The amount of plasmid DNA was calculated from the ratio of the plasmid peak area to the chromosomal DNA peak in the same lane.

b) Determination from plasmid minipreps.

The method of Holmes and Quigley (12) was used with some modifications: 30ml, 15ml, 10ml and 7.5ml of bacteria were grown at 37°C to $A_{600}=0.2$; 0.4; 0.6 and 0.8. Bacteria were pelleted at 3000g for 5 min, washed with 1ml TE buffer and resuspended in 500 μl of freshly prepared buffer: 8% sucrose V/V, 50mM EDTA, 50mM Tris-HCl, pH 8.0; 0.5% Triton X100, 4 mg/ml of lysozyme.

The solution was placed in a boiling water bath for 1 min and centrifuged immediately at 12000g for 10 min at room temperature. The supernatant was precipitated at -20°C for 10min with an equal volume of cold isopropanol, then centrifuged at 12000g for 15 min at 4°C . The pellet was dried under vacuum, resuspended in 250 μl (0.4M sodium acetate) and reprecipitated at -20°C for 30 min with 750 μl of ethanol.

After sedimentation by centrifugation, the DNA precipitate was dried, dissolved in 15 μl of water and 5 μl of loading solution (50% glycerol, 0.1% bromophenol blue in 80 mM Tris-acetate, 40 mM EDTA, pH 8.0) and electrophoresed.

cat protein assay

Cell-free extracts of *E. coli* strain HB101 harboring the different plasmids were prepared after growth in 7 ml Luria broth supplemented with 0.2% glucose and containing 30 $\mu\text{g/ml}$ chloramphenicol. At $A_{600}=6$, cells were harvested by centrifugation, washed and suspended in 600 μl of 30 μM DTT, 50 mM Tris hydrochloride at pH 7.8. Extracts were homogenized by three freezing/thawing cycles.

Samples corresponding to 4–1 μg and 200–40–4–0.4 ng of total protein were incubated in 0.25 M tris-hydrochloride pH 7.5, 1 μCi of ^{14}C chloramphenicol (25 $\mu\text{Ci/ml}$, Amersham) and 20 μl of 4 mM acetyl-coenzyme A (final volume 180 μl).

All reagents except coenzyme A were incubated together at 37°C for 5 min, coenzyme A was added and incubation resumed for 5, 10, 15, 20 and 30 minutes at 37°C . The reaction was stopped by ethyl acetate extraction. The organic layer was dried and taken up again in 3 μl of ethyl acetate, spotted on thin-layer silica gel plates and run with chloroform methanol (95:5, ascending). The chromatography separates acetylated from unreacted ^{14}C chloramphenicol. Chromatograms were autoradiographed and the identified spots were cut out and counted for radioactivity in a scintillation counter, yielding the relative amount of acetylated chloramphenicol.

RESULTS

Synthetic promoter design and insertion into plasmid pBR329

The general strategy was outlined in (6); here, we briefly describe those sequences carrying the CH pair. Two series of constructs were tested, bearing no resemblance to any 'consensus' *E. coli* promoter sequence, except the CH pair, and avoiding whatever sequence feature might give rise to structural or conformational particularities (palindromes, runs of purines or of alternating purine-pyrimidine, etc...). The constructs end with *HindIII* and *BamHI* cleavage sites, enabling insertion into the *HindIII*/*BamHI* gap of pBR329.

One construct, pBR GC-1, carries a 17 bases spacer sequence 76.5% GC, linking the CH pair. pBR GC-6 and pBR GC-8 derive from pBR GC-1, by deletion of one base (G upstream CH-10) and insertion of one base (T upstream (CH-10)) in the spacer, respectively. In pBR GC-6, the first T of CH(–10) underwent transition to C. A second construct, pBR AT-10, carries a 17 bp spacer sequence 76.5% AT (fig.1). pBR AT-4 and pBR AT-1 derive from pBR AT-10 and correspond to a 3 base deletion (TAT) exactly in the middle of the 17 bp sequence, and a 9 bp insertion at the same place, respectively. pBR AT-5 and pBR AT-2 derive from pBR AT-10 by insertions flanking CH-10 downstream, of 12 and 19 bases, respectively. It may be important to recall that pBR GC-6 and -8, as well as pBR AT-1, -2, -4 and -5, are spontaneous mutants derived *in vivo* from constructs pBR GC-1 and pBR AT-10, respectively, under chloramphenicol selection (see (6)).

Upon insertion into the *BamHI*/*HindIII* gap of pBR329, these constructs yield our pBR-AT and pBR-GC series (fig.1). In pBR329, the *HindIII* site begins 3 bases downstream of the published transcription initiation site (address 268 in (7)) of the chloramphenicol acetyl transferase (*cat*) gene (taken as +1 below). Thus, except for pBRAT-2 and -5, CH(–10) is at the

normal position relative to the wt transcription start site of the *cat* gene.

Bam HI cleaves pBR329 far inside the coding sequence of the *tet*^r gene (codon 96). Thus both pBRGC and pBRAT series not only lack the *tet*^r and *cat* promoter—at least the parts upstream transcription start of the *cat* gene—but any unidentified promoter element which could exist upstream of −35 in the wt *cat* promoter would also have been removed. pBR329, deleted for the smaller *Hind* III–*Bam* HI fragment (pBR HB in fig.1), supports only very marginally *cat* gene transcription or expression (see below). We are therefore confident that spurious wt *cat* promoter elements located upstream the site of wt transcription start are absent in our pBRGC and AT constructions and that transcription through the *cat* gene, directed by some other promoter carried by pBR329, is negligible.

pBRAT and GC constructions, carrying the CH pair, support *in vivo* efficient *cat* gene transcription and only slightly affect plasmid copy number

To quantify transcription from a plasmid-borne promoter, plasmid copy number must be monitored. Work from Bujard's laboratory (13, 14) demonstrates that plasmid copy number not only varies with bacterial growth rate, but also can depend on the transcription efficiency of plasmid-borne promoters.

Fig.2: Amount of plasmid DNA vs. cell density during exponential growth. For each density (A_{600} 0.2, 0.4, 0.6, 0.8) (A_{600} = 0.4 is approximately 5×10^7 cells/ml), nucleic acids were separated by gel electrophoresis; bands were quantified by densitometry (the area under the peak corresponding to a band was taken as a relative measure of the plasmid copy number). — Comparison between HB101 transformed with a) the plasmids pBRGC-1 $\square$; pBRGC-6 $\triangle$; pBRGC-8 $\blacktriangle$ and pBR329 $\blacksquare$; b) pBRAT-10 $\bullet$; pBRAT-1 $\square$; pBRAT-2 $\circ$; pBRAT-4 $\triangle$; pBRAT-5 $\blacktriangle$ and pBR329 $\blacksquare$.

We measured carefully by two different methods the plasmid copy number for pBR329 and all pBR-AT and -GC mutants studied, harboured in HB101 cells growing exponentially, from A_{600} =0.2 to 0.8 in LB medium supplemented with chloramphenicol. The results are reported in fig.2. The plasmid copy number per cell increased linearly with A_{600} in each case, but it can be seen that the rate of the increase is specific for each plasmid type. The measurements reported in the following sections have been made on material extracted from HB101 cells, grown to A_{600} =0.6 in LB medium. Numbers of plasmids per cell under these conditions, relative to pBR329 taken as 1, are given in table I, first line. We observe that, at a given A_{600} , the plasmid copy number for each of our constructs exceeds that found for pBR329 by 30–55%; the reason for this is unknown.

Quantification of transcripts

The quantitative primer extension assay was used with two 20-mer primers, complementary to the transcribed sequence of the *cat* gene of pBR329, +27 to +47 and +82 to +102 (wt transcription start at +1, see fig.1). The assay yields both the amount of transcripts at individual start sites, and the addresses of the latter. Because of the unusual location of some of the start sites (see below), the experiments were repeated independently at least five times (up to ten times for certain plasmids) and gave consistent results, whatever the primer used. Typical examples are shown fig. 3. Quantification of the transcripts relative to those from pBR329 (taken as 1) is given in table I, line 2 (together with the error estimate), consistent with the results obtained by the filter binding assay (fig. 4).

For the plasmids tested, the amount of transcript per plasmid differed by a factor of ten, the strongest being pBRGC-1 (6.5 times pBR329), the weakest pBRAT-4 (0.6 times pBR329). The GC series is consistently stronger than the AT series (average 4.6 and 2.1 times the number of *cat* transcripts of pBR329, respectively). The AT/GC content of the spacer linking consensus hexamers is clearly involved in determining the *cat* transcription efficiency from these constructions. For the pBRGC series, the efficiency drops by a factor of two on changing the spacer size by ± 1 base pair. For the pBRAT series, the three constructions pBRAT-10, -2 and -5, identical from the *Bam* HI site down to and including CH(−10), also differ in their efficiencies by a factor of two (clearly a consequence of their sequence diversities downstream CH(−10). pBRAT-10, which carries a 9 bp insert in its spacer, is nevertheless among the most efficient pBRAT constructions.

Table 1: Quantification of plasmid copy number, *cat* gene transcripts, *cat* protein and cellular chloramphenicol resistant from pBR329 (taken as 1) and the pBR-AT and -GC plasmid constructions in *E.coli* HB101 cells growing exponentially (A_{600} =.6). pBR HB is the plasmid derived from pBR329 by deletion of the small (346 bp) *Bam* HI–*Hind* III fragment carrying the wt Cm^r promoter. Data are averages of 5 (relative transcripts amounts and *cat* protein) and 10 (plasmid copy number and antibiotic resistance) independent experiments at least. Bold faced numbers are values corresponding to one plasmid copy.

	Plasmid									
	pBR329	pBRGC-1	pBRGC-6	pBRGC-8	pBRHB	pBRAT-10	pBRAT-1	pBRAT-2	pBRAT-4	pBRAT-5
DNA plasmid	1 $\pm$ 0.3	1.45 $\pm$ 0.4	1.3 $\pm$ 0.5	1.55 $\pm$ 0.4	1.75 $\pm$ 0.5	1.4 $\pm$ 0.4	1.55 $\pm$ 0.6	1.35 $\pm$ 0.7	1.5 $\pm$ 0.5	1.3 $\pm$ 0.2
Transcripts	1 $\pm$ 0.2	9.5 $\pm$ 1.6	4.8 $\pm$ 2	5.5 $\pm$ 1.3	0.21 $\pm$ 0.06	3.5 $\pm$ 1.2	4.6 $\pm$ 0.3	1.6 $\pm$ 0.6	0.95 $\pm$ 0.5	3.9 $\pm$ 1.2
	1	6.5	3.7	3.5	0.1	2.5	3	1.2	0.6	3
<i>cat</i> activity	1 $\pm$ 0.5	12.8 $\pm$ 0.7	8.2 $\pm$ 0.2	8.8 $\pm$ 0.2	0.03 $\pm$ 0.01	11.3 $\pm$ 0.3	13.5 $\pm$ 0.5	1.6 $\pm$ 0.5	1 $\pm$ 0.4	5.4 $\pm$ 3
	1	8.8	6.3	5.7	0.02	8.1	8.7	1.2	0.6	4.1
Resistance <i>in vivo</i>	1 $\pm$ 0.2	1.5 $\pm$ 0.1	1.3 $\pm$ 0.05	1.1 $\pm$ 0.2	0.05 $\pm$ 0.01	1.7 $\pm$ 0.05	1.3 $\pm$ 0.1	0.8 $\pm$ 0.1	0.7 $\pm$ 0.1	1.5 $\pm$ 0.2

Transcription start site mapping

In a previous report (6), we mapped the main transcription start sites of various constructions, including those studied here, using reverse transcriptase extension of a primer complementary to the transcribed sequence +27 to +47 (+1 transcription start in pBR329). To be sure not to miss transcripts possibly starting further downstream, a series of assays was in addition performed with a second primer, +82 to +102. Our results, displayed fig.1, fully confirm and in addition extend and quantify those given in (6). The average percentage of transcripts starting at a given

Fig. 3: Map of transcription initiation *in vivo* of *cat* gene. Examples of primer extension assay from *in vivo* transcripts. An autoradiogram of extended products is shown (10% polyacrylamide, 7M urea gel). For each plasmid, the assay was repeated several times. Lanes 1 and 2 show results with total RNA from HB101 transformed with pBRGC-1 and pBRAT-1, respectively. Lanes M: 32 P- labeled *Hpa* II -cleaved pBR322 as size markers—Size is in nucleotides.

site is indicated fig.1 (it is usually very close to the percentages found in individual experiments).

Each pBRGC construct started transcription primarily at a single site; pBRGC-1 started almost all its transcripts 6 bases downstream of the last T of CH-10, the same holding for 80% of the transcription from pBRGC-6. 9 out of 10 transcripts from pBRGC-8 also started at a single site, but here located right in CH(-10), a consequence of the spacer length increased by 1bp. In each case, one minor second site was seen, less than 8 bases from the major one.

cat transcripts from the pBRAT series, in contrast started at many places, sometimes scattered over as much as a 30 bp sequence. In AT-10 (17 bp spacer) as well as in AT-1 (26bp spacer), 3 out of 4 transcripts started at two places in CH(-10), the remaining, 4–5 bases downstream. AT-2 and AT-5 share with AT-10 the same sequence down to and including CH(-10), differing only by sequences inserted just downstream of CH(-10), 20 and 12 bp long, respectively. 2 out of 3 transcripts from AT-2 started in CH(-10) and four additional started sites were found in the following 30 bp sequence. In AT-5, 3 out of 4 transcripts started at the conventional place 7–12 bp downstream of CH(-10), but 15% started right in the middle of the spacer sequence, and a minor start site was found 17bp downstream of CH(-10).

Finally, AT-4 with its 14bp spacer started its transcripts at a rather conventional distance, 6 (38%) and 12 bp (62%) downstream of CH(-10).

AT-10, -2 and -5 provide a clear demonstration that sequences downstream of CH(-10) are involved in setting transcription starts and, compared to AT-10, AT-4 shows that spacer size is a relevant factor also.

cat expression and chloramphenicol resistance

The amount of *cat* protein produced *in vivo* under control of the synthetic promoters can be estimated from the specific *cat* activities, in a given amount of total protein extract from cells harbouring the plasmids. *cat* activity in the pBR -AT and -GC series, relative to pBR329, is given in table I; we observe that pBR HB, the plasmid derived from pBR329 by deletion of the smaller *Hind* III -*Bam* HI fragment which carries the *cat*

Fig.4: Determination of relative mRNA concentration measured by filter hybridization. For each plasmid assayed, decreasing concentrations of total RNA (from left to right : 15, 10, 5 and 2.5 µg) were spotted and hybridized with end-labeled 20-mer primers. Autoradiograms of the nitrocellulose membranes after 48 h of exposure.

promoter, showed no significant *cat* activity. In all cases, transcription efficiency paralleled *cat* activity qualitatively, but its ratio to the amount of transcripts in a given plasmid was not constant. In particular, per plasmid copy, pBR AT-1 and -10 had *cat* activities almost as strong as pBR GC-1, despite the fact that the latter produced more than twice as many transcripts. We believe this is due to the more extended (9 bases) untranslated 5' end of mRNA transcribed from both pBR AT-1 and -10, relative to pBR GC-1, which could facilitate translation initiation of the *cat* gene.

Chloramphenicol resistance conferred by each plasmid to the host cell is also given in table I; the values are slightly different from those given in (6), because the averages were taken over more measurements (10 at least).

DISCUSSION AND CONCLUSIONS

In the present contribution, we focus on the only identified *E. coli* promoter elements, the pair of consensus hexamers and their location relative to the wt transcription start site of the marker gene *cat* (termed 'wt(+1)' below). Eight constructs were tested, all carrying the CH pair (with the exception of one transition), but differing in their context, i.e. (i) 17 bp spacer, constraint in AT or GC; (ii) spacer sizes varying from 14 to 26 bp; and (iii) distances from CH(−10) to wt(+1) between 6 and 27 bp. The plasmids carrying the constructs enabled RNAP to initiate transcription of the *cat* gene. Since the promoterless plasmid, pBRHB, deleted for the smaller *Hind III*–*Bam HI* fragment, transcribes *cat* at a very basal level only, we conclude that the constructs indeed contribute in potent, specific promoter activities. The only common promoter element shared by the constructs is the CH pair: in the contexts tested, the latter appear *sufficient* to provide RNAP with the minimal 'instructions' required for the series of steps leading to transcription initiation. (By 'instructions', we mean the set of non-covalent interactions enabling the promoter and its surroundings to trigger in a chronological order the steps enabling RNAP to initiate transcription. Incidentally, these instructions can be encoded also by sequence elements bearing no resemblance with the CH pair (6, 15), i.e. they are not directly related to a symbolic (literal) code).

In the eight constructs, the CH pair is sufficient to specify a promoter structure; is it also sufficient to specify the two fundamental characteristics of promoters: efficiency of transcription, as determined by transcription initiation frequency, and selection of transcription start address? From our results, it is clear that it is the context of the CHs which plays a major role in the specification of these 'executive' instructions.

Transcription efficiencies of our constructions vary by a factor of ten. Since the region upstream of CH(−35) and downstream of wt(+1) are common to all of them, it follows that important determinants of promoter efficiency are with sequences in between, that is in the spacer and/or in the sequence from CH(−10) to wt(+1).

Spacer size and sequence of the spacer modulate transcription efficiencies by as much as a factor of ten in our constructs. pBR GC-1 and pBR AT-10 differ only in the sequence of their 17 bp spacer (and a deletion of A in the *Hind III* site just downstream of CH(−10)). Assuming that the latter has little influence, the almost threefold higher efficiency of pBR GC-1 demonstrates that the sequence of the 17 bp spacer, and especially its AT/GC base composition constraint, can be a potent determinant of the transcription initiation rate. Because of the irregular

purine/pyrimidine sequences selected in both AT and GC rich spacers, we do not expect conformations of the double helix in the spacer region to deviate from a conventional 'B' form in either construct (C. Reiss, unpublished results from molecular modelling calculations). Therefore, we do not believe, in the present case, that the differences in efficiency are linked to altered spacer conformations (16). Several years ago, we observed that wt-*E. coli* promoters in general are moderately GC-constrained in their spacer sequence (17), suggesting that the enhanced stability or rigidity, and/or the reduced order in the hydration shell, associated with the GC-rich spacer, could be relevant for enhancing promoter efficiency.

The size of the spacer linking the CH pair was shown earlier to be a negative modulator for chloramphenicol resistance, when departing from the 17 bp optimum (6). pBR GC-1, -6 and -8, which differ in spacer size (17, 16 and 18 bp, respectively), confirm this trend at the transcriptional level, since promoter efficiencies of pBR GC-6 and -8 are slightly over half that of pBR GC-1. In the pBR AT series, the same trend is shown by pBR AT-4 (spacer of 14 bp), four-fold less active than pBRAT-10 (17 bp spacer), but, in contrast, pBR AT-1 with a 26 bp spacer matches the transcription efficiency of pBR AT-10. It is possible that a stereospecific relationship holds between CH(−10) and CH(−35) which is restored upon adding 9 bp (just short of one helix turn) to the 17 bp spacer optimum, reminiscent of findings reported recently for eukaryotic promoters (18).

The fact that pBR AT-10 and -1 display comparable, high promoter activities despite the fact that their spacers differ in size by as much as 9bp, could indicate not only that RNAP may not require simultaneous contacts with both CH(−10) and CH(−35) at some step in the formation of the active complex, but also that the spacer size does not necessarily play a major role in the kinetic pathway leading to transcription initiation. These suggestions assume that some 'cryptic' consensus sequence has not taken over the role of the standard CH pair present in pBR AT-1 and -10, a reasonable assumption (see below).

The sequence between CH(−10) and wt(+1) is also actively involved in controlling promoter strength. The constructs in pBR AT-10, -2 and -5 share identical sequences upstream and including CH(−10), and downstream wt(+1); pBR AT-2 carries a 20 bp insert, pBR AT-5 a 12 bp insert, both flanking CH(−10) downstream. Compared to pBR AT-10, the efficiency of pBR AT-2 is reduced by a factor of two, but pBR AT-5, in contrast, is slightly more active. We observe that the 20 bp insert is 80% AT, but the 12 bp insert is only 60% AT. This suggests that modulation of promoter strength by the sequence downstream of CH(−10) could depend on base composition constraints, again consistent with our earlier observation that, downstream of the −10 consensus sequence, wt *E. coli* promoters carry on the average a domain 52% ± 10% AT (17). Kammerer *et al.* (19) have shown that the sequence of the 20 first transcribed base is capable of modulating transcription activity by more than one order of magnitude. Our results show the presence of additional promoter strength instructions between CH(−10) and wt(+1), hence argue for the extension of the downstream regulatory region (DSR) defined by Kammerer *et al.* (19) to include the sequence abutting against CH(−10).

The context of the CHs is also clearly involved in specifying position and grouping of transcription start sites, with distinct behaviours for AT- or GC-rich spacers. pBRGC constructs initiate transcription at a few sites only, tightly clustered around a distinct major site; a rather opposite trend is observed with most pBRAT constructs. pBRAT-2, -5 and -10 share a common

sequence upstream (and including) CH(−10), yet start transcription at many different places, from 10 bp upstream CH(−10) (pBRAT-5) to 27 bp downstream (pBRAT-2). The unconventional location of major start sites inside or upstream CH(−10) is shared by pBRGC-8 and all pBRAT constructs, except pBRAT-4.

These results support an unambiguous conclusion : in our constructs, transcription starts are not unconditionally linked to the position of the CHs; the contexts of the latter bear important determinants for position, multiplicity and grouping of transcription start; therefore, a quest for 'cryptic' consensus sequences associated with the observed transcription start sites appears meaningless.

As a naive but consistent rationale for our findings, and as a working hypothesis for further studies on promoter structure and functioning, we tentatively offer the following scheme: properly located consensus hexamers (or their synonymous variants bearing similar instructions) enable stereospecific docking of RNAP on the promoter sequence. As a result, the local environment (mainly ionic density and hydration shell) of the double helix would undergo a sharp change and induce *isothermal* unwinding of promoter domain(s) having suitable stabilities (AT/GC constraints, see (21 and 23) for the underlying thermodynamics). We envision the apparent rate of unwinding and the extend of the unwound domain(s), to be essential in determining transcription initiation frequency as well as start site location and grouping.

As an example, promoter pBRAT-2 and pBRGC-1 include each a domain of comparable, low stability (80% AT), extending 50 bp downstream CH(−35) and 23 bp downstream CH(−10), respectively. The apparent rate of unwinding combines the rate at which the environmental change required for isothermal melting sets up, and the rate of the actual unwinding of the domain. This apparent rate is expected to diminish as the size of the unwound domain increases, because the required environmental change must span a larger DNA sequence and because the stress imposed to the flanking sequences, which opposes unwinding, increases. Therefore, opening of the pBRAT-2 promoter would take longer than that of pBRGC-1; if this step is rate-limiting *in vivo* as it is *in vitro*, the efficiency of the latter promoter would overrank that of pBRAT-2, as observed. The position of the unwound domain would also set the site of transcription initiation; its size could determine the grouping of the start sites, as a small loop (of the order of 20 bp, about the size of the loop observed in the ternary complex) would precisely position RNAP and its active center relative to the template, whereas in a large loop, RNAP could be free to initiate transcription at several places.

The main result of our study is to show that the promoter signal can be broken (in part at least) into a structural component, primarily carried by the consensus hexamers, and an executive component, mainly born by their contexts. The hypothetical scheme just sketched suggests a mechanism of promoter action; the experimental system we have designed is a suitable workbench for its testing.

The role of the CH context in promoter activity is also stressed in a recent paper by H.R.Drew and colleagues (C.M.Collis *et al.*, Nucl. Acids Research (1989), 17, 9447–68) which became available after our manuscript was submitted. Their approach is complementary to ours, as they examine the influence of special DNA structures on promoter activities, whereas we have purposely avoided sequences known or suspected to yield 'unusual' conformation.

ACKNOWLEDGEMENTS

Catherine Collin provided skillful technical assistance during determination of plasmid copy number and *cat* protein assay. We thank Dick d'Ari for help with the manuscript, C.Michon-Dubucs for oligonucleotide synthesis, L.Corme for typing the manuscript and R.Schwartzmann for photographic work.

REFERENCES

1. Harley, C.B. and Reynold, R.P. (1987), Nucl. Acid Res., **15**, 2343–51.
2. Hawley, D.K. and McClure, W.R. (1983), Nucl. Acid Res., **11**, 2237–55.
3. Brosius, J., Erfle, M. and Storella, J. (1985), J. Biol. Chem., **260**, 3539–41.
4. Dobrynin, V.N., Korobko, V.G., Severtson, A.I., Bystrov, N.S., Chuvpilo, S.A. and Kolosov, M.N. (1986), Nucl. Acid Res. Symp. Ser., **7**, 365–75.
5. Rossi, J.J., Soberon, X., Marumoto, Y., Mc Mahon, J. and Itekura, K. (1983), Proc. Natl. Acad. Sci. USA, **80**, 3203–07.
6. Jacquet, M.A., Ehrlich, R. and Reiss, C. (1989), Nucl. Acid Res., **17**, 2933–45.
7. Covarrubias, L. and Bolivar, F. (1982), Gene, **17**, 79–90.
8. Maxam, A.M. and Gilbert, W. (1980), Methods in Enzymol., **65**, 499–560.
9. Jacquet, M.A. and Ehrlich, R. (1985), Biochimie, **67**, 987–97.
10. Mc Master, G.K. and Carmichael, G.G. (1977), Proc. Natl. Acad. Sci. USA, **74**, 4835–54.
11. Chiang, C.S. and Bremer, H. (1988), Plasmid, **20**, 207–220.
12. Holmes, D.S. and Quigley, M. (1981), Analytical biochemistry, **114**, 193–197.
13. Stueber, D. and Bujard, H. (1982), Embo J., **1**, 1399–1404.
14. Deuschle, U., Kammerer, W., Gentz, R. and Bujard, H. (1986), Embo J., **5**, 2987–94.
15. Horwitz, M.S.Z. and Loeb, L.A. (1988), J. Biol. Chem., **263**, 14724–14731.
16. Auble, D.T. and de Haseth, P.L. (1988), J. Mol. Biol., **202**, 471–482.
17. Ehrlich, R., Marin, M., Larousse, A., Gabarro-Arpa, J., Schmitt, B. and Reiss, C. (1984), Folia Biologica, 105–18.
18. Lennard, C., Matthes, H.W.D., Egly, J.M. and Chambon, P. (1989), Nucl. Acid Res., **17**, 6903–14.
19. Kammerer, W., Deuschle, U., Gentz, R. and Bujard, H. (1986), EMBO J., **5**, 2995–3000.
20. Matthew, J.B. and Ohlendorf, D.H. (1986), J. Biol. Chem., **260**, 5860–5862.
21. Gabarro-Arpa, J., Tougaard, P. and Reiss, C. (1979), Nature, **280**, 515–517.
22. Marcaud, H., Gabarro-Arpa, J., Ehrlich, R. and Reiss, C. (1986), Nucl. Acids Res., **14**, 551–558.
23. Gabarro-Arpa, J. and Michel, F. (1982), Biochimie, **64**, 99–112.