In vivo gene expression directed by synthetic promoter constructions restricted to the −10 and −35 consensus hexamers of *E.coli*

Marie-Ange Jacquet*, Ricardo Ehrlich+ and Claude Reiss

Institut Jacques Monod, CNRS and Université Paris VII, Tour 43, 2 Place Jussieu 75251, Paris cédex 05, France

Received February 2, 1989; Revised and Accepted March 23, 1989

ABSTRACT

Two synthetic DNA sequences, carrying no other known *E.coli* promoter element than the consensus hexamers (CH) TTGACA (CH−35) and TATAAT CH(−10), spaced by 17 bp, were inserted in pBR329, in a position enabling transcription of the complete Cm^r gene. The region upstream of the Cm^r transcription start was carefully cleared of w.t. promoter elements (full deletion of the wild type (w.t.) Cm^r promoter upstream +2 and large portion of an upstream coding sequence). Both synthetic promoters, which differ only by the sequences of the spacers (non consensus, constrained in AT or GC) support *in vivo* high level Cm^r gene expression. The GC rich spacer is associated with transcription start at the usual +1 position, but with the AT rich spacer, transcription starts at several places, mainly in CH(−10). Rearranged promoter sequences derived from the synthetic ones upon transformation with partly ligated plasmids, yield new insights on the role of the standard CH pair, the size of the spacer and the sequence downstream of CH(−10).

INTRODUCTION

Sequence comparison among over 250 promoters recognized by *E.coli*. RNA-polymerase (RNAP) reveals (1) a baffling diversity, except for the so-called, −10 and −35 consensus hexamers (CH) (transcription starts at address +1). However, the sequences of the CH can deviate appreciably from the standard CH sequences, TTGACA CH(−35) and TATAAT CH(−10); in particular, no known natural promoter exhibits the standard CH set. On the other hand, sequences matching reasonably well the CH sequences and spaced by 17±2 base pairs (bp) are found at many places in the genome where no promoter function is detected (2).

Outside the CH sequence, conservation is of poor statistical significance, and point mutations usually have unpredictable effects. It could well be that the classical 'vertical' averaging of sequences aligned 'horizontally' is not adequate. Synergic signal elements could be deposited at given places within an individual promoter, modulating for instance the effect(s) of the CH pair. These elements could differ, or occur at different places, for different promoters and hence 'vertical' averaging of the promoters would not allow their identification (see (3) for examples).

Indications as to the presence of signal elements outside the CH are puzzling. For certain promoters, sequences upstream of −35 (up to −160) (4) or downstream of −10 (as far as +20) (5) have been shown to have important effects on promoter strength. In other promoters, deletion of the region upstream of −35 (6), or even the CH(−35) itself (7) has little effect on transcription efficiency. Clearly, elements contributing to the promoter signal can reside throughout the region footprinted by RNAP (typically from −50 to +10 (8); in some cases footprints extending to −100 or +20 have been observed (9)).

To investigate the promoter signal, synthetic promoter sequences were tested which conform to the best average sequences available from a variety of promoter compilations. They were ligated into plasmids, at places where the core of the w.t. promoter had been deleted, and tested for their ability to promote transcription. The two largest synthetic promoter sequences were constructed by Dobryrinin *et al.* (10) and by Rossi *et al.* (11); both were introduced in place of the *tetR* promoter in pSC101, from CH(−35), either to the transcription start (11) or to the translation start (10), downstream. Transcription efficiencies ranked from moderate (Dobrynin's construction, tested by Deuschle *et al.* (5)), to excellent (11). However, both constructions were tested after insertion into the genome context of the excised w.t. promoter, and thus unidentified, w.t. promoter signal elements upstream and downstream of the insert could still be present. Furthermore, both inserts carried, outside the CH pair, 'best' consensus sequences (although the scores for consensus never exceeded 40% for any base).

To avoid ambiguities connected with uncontrolled, possibly ubiquitous signal elements, we decided to test synthetic promoter sequences restricted to the standard CH pair spaced by 17 bp, surrounded by deliberately non-consensus sequences. In particular, both CH were framed by unique restriction sites, enabling various combinations of upstream, downstream of spacer or hexamer sequences to be tested. These were inserted into a genome site devoid of any known promoter activity. We report here the *in vivo* behaviour of the first series of such constructs, aimed primarily at the relation of a standard CH pair to promoter activity, spacer sequence, and length.

MATERIALS AND METHODS

Bacterial strains and plasmid vector

E.coli strain HB101 was used as host in transformation experiments ; the plasmid vector was pBR329 (12).

Synthetic promoters

Both direct and complementary strands of the synthetic promoters (Fig.1) were synthesized by the phosphotriester method in a SAM1 (Biosearch) DNA synthesizer, and phosphorylated with ^{32}P using polynucleotide kinase. Both strands were anealed at 37°C overnight, in 220 mM sodium phosphate, 400 mM KCl, 1 mM EDTA.

Plasmid construction

Restriction endonucleases and other enzymes used for nucleic acid engineering were from commercial sources and used according to the supplier's instructions ; standard molecular cloning methods were employed (13).

The larger fragment (3804 bp) in the BamH I−Hind III digest of pBR329 (Fig.2) was isolated by electrophoresis in 2 % agarose. This fragment was mixed with the synthetic promoter duplex. Ligation with T4 DNA ligase was partial only, as enzyme concentration and reaction time conditions did not allow completion of the ligation reaction. The mixture of fully and partly ligated plasmids was introduced into HB101 cells made competent by $CaCl_2$ treatment. Transformed cells were isolated on LB-agar plates containing 100μg/ml ampicillin, after the different recombinant colonies containing the insertion were identified by screening on chloramphenicol/agar plates at different antibiotic concentrations (100 and 400 μg/ml) and by hybridization on nitrocellulose filters. Alternatively, minipreparations of plasmids were tested for the unique restriction sites which had been introduced into the synthetic promoter sequence (see Fig.1).

The nucleotide sequences of the Nar I−Pvu II restriction fragment (236 bp) which

Fig.1. Sequences of synthetic and rearranged promoters
The consensus hexamers TTGACA and TATAAT (bold faced) and unique restriction sites of the designed sequences are indicated. Rearranged bases are underlined, dashes link contiguous bases after deletion.
pBRGC-0 and pBRAT-0 are the sequences chemically synthesized and cloned into the pBR329 plasmid deleted of the small Hind III-BamH I fragment.
pBRGC-2,-4,-6,-8 and pBRAT-1,-2,-3,-4,-5,-6,-8,-10 are rearranged sequences.
▲ insertion sites. ⇓ Arrows indicate *in vivo* transcription start sites, accurate to within ± 1 base (results from at least 4 independent experiments).

contains the cloned promoter, was obtained by the Maxam and Gilbert procedure (14).

In vivo expression of chloramphenicol resistance

The ability of plasmids to confer resistance to increasing concentrations of chloramphenicol in liquid cultures was examined by exposing aliquots of a rapidly growing culture to chloramphenicol (up to 1 mg/ml). Cell growth was monitored by OD_{550}. Efficiency of plating was measured on agar containing 50μg/ml ampicillin and 100 μg/ml chloramphenicol; incubation was overnight at 37°C (Fig.3).

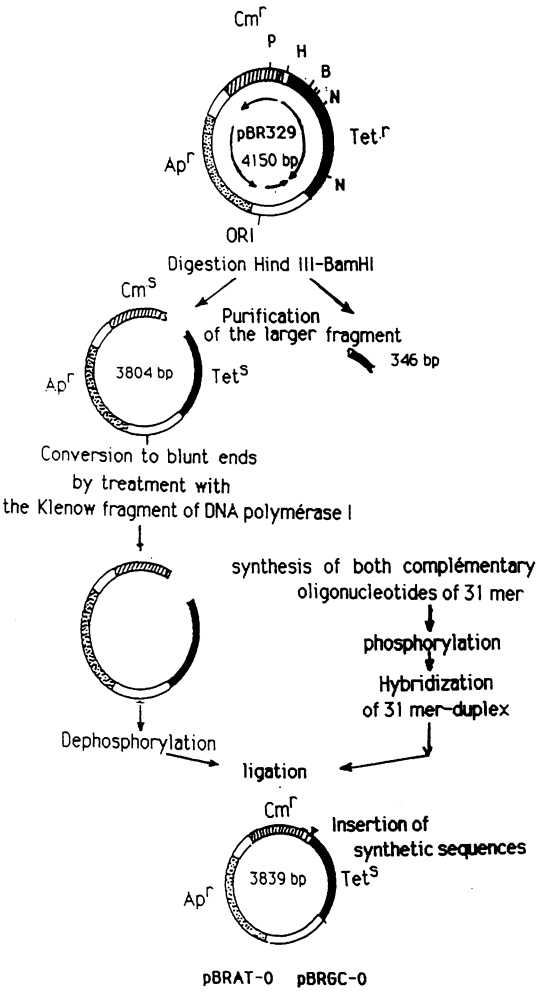

Fig.2. Plasmid construction
Strategy used to construct the synthetic promoter sequences and their insertion within pBR329 deleted of Hind III-BamH I small fragment.
The circular map of pBR329 is shown with the localisation and direction of Ap^r, Tc^r, Cm^r genes ; restriction sites used for the sequencing : Nar I (N), Pvu II (P) ; for cloning : Hind III (H), BamH I (B) ; ORI, origin of DNA replication.

Mapping transcription start site in vivo.
A synthetic 20-mer DNA primer (sequence +27 to +47, assuming transcription start of w.t. Cm^r gene in pBR329 in the middle of the Hind III restriction site (12)) was prepared as described above. The HPLC-purified primer was end-labeled with $\delta^{32}P$ -ATP using T4 polynucleotide kinase (Biolabs).

Preparation of mRNA synthesized *in vivo* in *E.coli* strains HB101 carrying the plasmids under investigation was performed as described in (15). 20 μg of total mRNA were

coprecipitated with 1.10^7 cpm of the primer. The precipitate obtained after spinning in an Eppendorf centrifuge (4°C) was resuspended in 20 µl of hybridization buffer (80 % formamide, 0.4 M NaCl, 1 mM EDTA, 0.04 M Pipes, pH 6.4). Incubation was at 47°C overnight, after preincubation for 2' at 85°C. Hybrids confined in the aqueous phase following phenol treatment were precipitated with ethanol. Primer extension was carried out (42°C, 60 minutes) with 30 µl AMV reverse transcriptase (Boehringer Mannheim), 300 units/ml, tris HCl, pH 8.3, KCl 140 mM, $MgCl_2$ 10 mM, mercaptoethanol 20 mM ; extended products were migrated in 10 % polyacrylamid gel, 7 M urea together with size markers, and analyzed by autoradiography (autoradiograms not shown ; as reported in (16), the size of the markers were shifted 1−2 bp upwards with respect to the extended products).

RESULTS

Synthetic promoter sequences

Two synthetic promoter sequences were designed (Fig.1), sharing the same length (31bp), both standard CH (TTGACA at −35 and TATAAT at −10) spaced by 17 bp, and restriction sites BamH I, just upstream of CH(−35), and Hind III, just downstream of CH(−10). They differ in their spacer sequence : GC-0 has a spacer of 76.5 % GC which carries unique restriction sites for Kpn I, Sma I and Xho I ; AT-0 has a spacer of 76.5 % AT which carries unique restriction sites for Xba I and Bcl I. The spacer sequences were chosen to avoid possible formation of particular structures (cruciform, bends, etc...). In the two constructions, both CH are tightly framed by unique restriction sites ; these synthetic promoters are therefore convenient construction sets for building promoters from a variety of sequence elements.

Promoter insertion into pBR329

The synthetic promoters were cloned into pBR329, replacing the 346 bp long Hind III−BamH I fragment. The BamH I site in pBR329 (address 606 in the sequence numbering of Covarrubias and Bolivar (12)) covers codon 96 in the Tc^r gene (Fig.2) ; Hind III (address 260) cleaves the transcribed sequence of the Cm^r gene at position +2. Consequently, the removed Hind III−BamH I fragment of 346 bp carries the promoter and transcription initiation site controlling expression of Cm^r on one strand, and the Tc^r gene from CH(−10) of the Tc^r promoter to about 1/4 of the Tc^r coding sequence (N-terminal side), on the other strand. The larger (3804bp) Hind III−BamH I circularized fragment of pBR329 was indeed found to confer Ap^r/Cm^s and Tc^s to HB101 (data not shown).

Both synthetic promoters were ligated into the Hind III/BamH I gap of pBR329 (ligation reaction not completed) and the resulting plasmids used to transform *E.coli* strain HB101. Tc^s/Ap^r transformants (carrying the plasmids) were identified by colony hybridization and screened for the Cm^r phenotype. The presence of the synthetic promoters in the transformants was determined by mapping the unique restriction sites they carry.

Two transformants were recovered, carrying the designed promoter sequences. pBR GC-1 has exactly the planned GC-0 sequence ; pBR AT-10 has the planned AT-0 insert, except for the deletion of an A in the Hind III site downstream of CH(−10) ; the larger Hind III−BamH I fragment of pBR329 is found unmodified in both plasmids.

Promoter sequence rearrangements.

As stated in Materials and Methods, the ligation reactions with both designed sequences AT-0 and GC-0, were not brought to completion and hence transformation was attempted

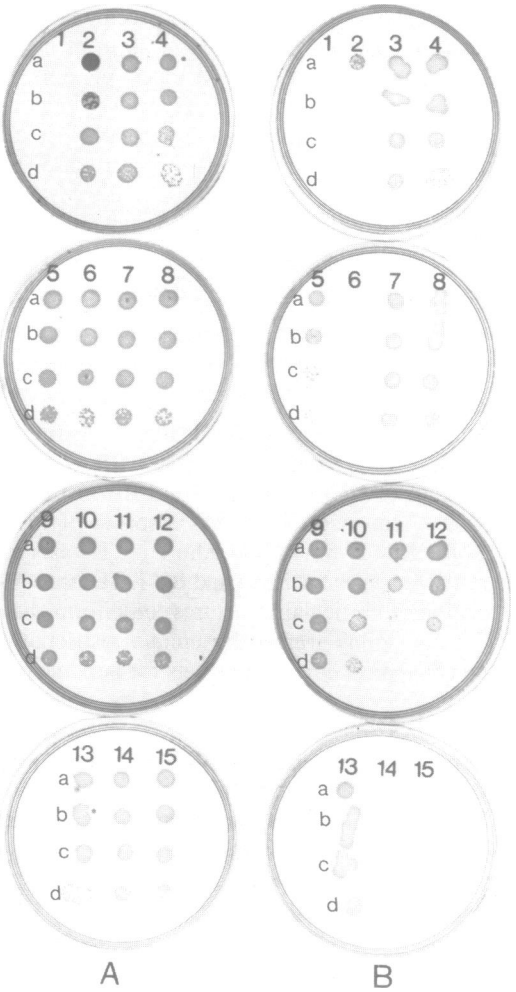

Fig.3. Efficiency of plating
Growth on LB-agar plates of HB101 cells carrying (1) no plasmid, ; (2) pBR329 deleted of small fragment Bam HI−Hind III and ligated ; (3) pBR329, (4) pBRGC-1, (5) pBRGC-2, (6) pBRGC-4, (7) pBRGC-6, (8) pBRGC-8, (9) pBRAT-1, (10) pBRAT-2, (11) pBRAT-3, (12) pBRAT-4, (13) pBRAT-5, (14) pBRAT-6, (15) pBRAT-8.
A-Control plates with 50 μg/ml ampicillin
B-Plates containing 100 μg/ml chloramphenicol respectively. a) to d) correspond to dilutions of cell stock (from stationary culture, 10^0, 10^{-2}, 10^{-4} and 10^{-6} times, respectively; 10 μl of cell suspension were deposited in each case.

with a mixture of linear and circular plasmids in each case. Transformation with linear DNA is known to be inefficient, and to generate high mutation rates in the vicinity of the DNA extremities, in contrast to circular DNA (17). Indeed, for both pBR AT and pBR GC series, the transformation yield, as deduced from the Amp^r phenotype, was low (10^{-4}). Among the transformants, a large spectrum of Cm resistance was observed. For both series, a set of colonies was isolated, each displaying a different level of Cm resistance

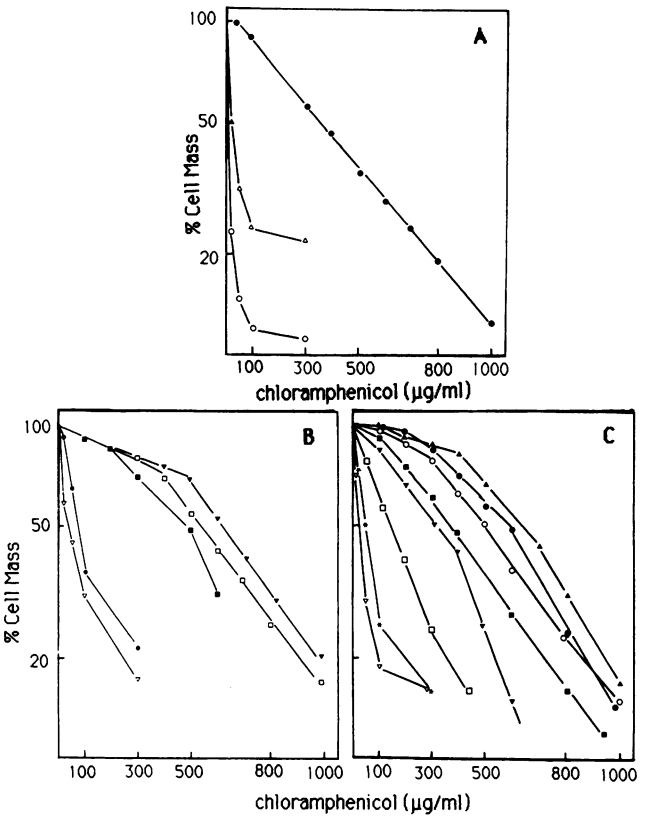

Fig.4. Chloramphenicol resistance in liquid medium for HB101 cells containing different plasmids.
Top: A-Controls : with HB101 cells no plasmid ; HB101 carrying pBR329 deleted of Hind-Bam HI (346 bp) △ and pBR329 ●.
Bottom : B-Cells containing the plasmids of pBRGC series : pBRGC-1:▼; pBRGC-6:□; pBRGC-8:■; pBRGC-2:●; and pBRGC-4:▽.
C-Cells containing the plasmids of pBRAT series : pBRAT-10: ▲; pBRAT-5:●; pBRAT-1:○; pBRAT-2:■; pBRAT-4:▼; pBRAT-3:□; pBRAT-6:* and pBRAT-8:▽.
Ordinate : % of cell mass with respect to untreated cells, mesured by optical density at 550 nm after 2 hours growth in LB medium at 37° of an initial 0.05 OD_{550} culture.

and carrying at least one of the unique restriction sites introduced into the synthetic promoter; plasmids were sequenced between Pvu II and Nar I.

A great variety of sequences rearranged in the vicinity of the ligation sites was obtained. Those we have sequenced (Fig.1) are likely to be a small sample of the rearranged sequences actually produced. For the pBR GC series, among the variants analyzed, only pBR GC-8 keeps both CH, but its spacer is increased to 18 bp and a C to T transversion had occurred in the Kpn I/Sma I site. The 3 other plasmids isolated, pBR GC-2,-4 and -6 carry CH(−35), but only the latter has a remnant (5/6) of CH(−10). None of the pBR GC series had rearranged in the promoter area upstream of Kpn I or downstream of Hind III.

For the pBR AT series, 8 promoter sequences were characterized ; 7 members of the series carry CH(−10), 5 have in addition CH(−35), one has none (pBR AT-3) ; three

Table 1. Chloramphenicol resistance of selection plasmids
Cm concentrations at which 50% of the HB101 cells carrying plasmids as indicated survive in LB liquid medium. Data are average over 5 independent experiments at least.

Plasmid	Cm concentration 50% HB101 survival	Plasmid	Cm concentration 50% HB101 survival
pBR GC -1	650 μg/ml	pBR AT -10	650 μg/ml
pBR GC -6	530 μg/ml	pBR AT -5	600 μg/ml
pBR GC -8	490 μg/ml	pBR AT -1	490 μg/ml
pBR GC -2	70 μg/ml	pBR AT -2	360 μg/ml
pBR GC -4	34 μg/ml	pBR AT -4	320 μg/ml
		pBR AT -3	140 μg/ml
		pBR AT -8	50 μg/ml
		pBR AT -6	26 μg/ml
pBR 329	350 μg/ml	HB101 cells, no plasmid	10 μg/ml
pBR 329, deleted of small BamHI–Hind III fragment and ligated	20 μg/ml		

of the five promoters which have both CH also maintained the synthetic spacer sequence (pBR AT-10, 2, 5) ; one has 4 bases deleted from the spacer (pBR AT-4) and one has inserted a repeat of 9 bases (pBR AT-1). pBR AT-4 is the only promoter not having a rearranged synthetic sequence downstream of CH(−10) ; all others have lost the Hind III site flanking CH(−10). Four of them have inserted next to CH(−10) a repeat (9 to 19 bp) taken from the spacer sequence and part of CH(−10). Although the main focus in the present contribution is on the designed sequences, pBR GC-1 and pBR AT-10, we include some relevant results obtained with rearranged sequences.

In vivo activity mediated by the selected promoters.

Fig. 3 and 4 show the growth of HB101 cells transformed by the two selected series of plasmids in solid agar and liquid medium containing increasing chloramphenicol concentrations. 50 % inhibition is reached at chloramphenicol concentrations indicated in Table 1.

In both series, promoters with synthetic CH(−10) and CH(−35) display *in vivo* Cm resistances 1−2 times that of w.t. pBR 329. The levels of resistance conferred by the corresponding plasmids are not sensitive to spacer sequence constraints (76 % GC or 76 % AT) and poorly to spacer length (14 to 26 bp). A 5−10 fold reduction of Cm resistance is observed with promoter sequences having lost one CH, either CH(−10), apparently specific of the pBR GC series, or CH(−35), which occurs in the pBR AT series.

The behaviour of pBR AT-3, which has no CH, is surprising, as it still confers appreciable Cm resistance to its host (40 % that of w.t. pBR 329).

In vivo transcription start sites for selected promoters.

Mapping transcription start sites was attempted by the primer extension assay using AMV reverse transcriptase. The synthetic primer was complementary to the 20 base sequence, positions +27 to +47 of the Cm^r mRNA in pBR329 (assuming transcription start (address +1) in the middle of the Hind III restriction site (12)) ; the initiation codon of the Cm^r gene starts at +48.

Transcription start sites could be mapped by this method only for the class of promoter constructions conferring moderate to high chloramphenicol resistance to the host bacteria, i.e. maps for pBR GC-2 and -4, and pBR AT-6 and -8, could not be obtained. The results are displayed on fig.1. pBR GC-1 starts transcription *in vivo* 6 bp downstream of the CH(−10), so is similar to w.t. pBR329, which initiates transcription at two sites located 5 (major site) and 7 bp downstream of the CH(−10) hexamer. pBR GC-6 has its main start site also 6 bp downstream of the CH(−10). pBR GC-8 is unconventional, as its transcription starts mainly within CH(−10). Still more unexpected, pBR AT-1 and -10 initiate transcription at three sites, one 4bp downstream of the CH(−10), but the two others, including the major middle one, are in CH(−10). The other rearranged sequences of the pBR AT series display mainly single transcription start sites, pBR AT-2 in CH(−10), pBR AT-3 7bp, pBR AT-4 12bp, and pBR AT-5 7bp, downstream of CH(−10). For each plasmid, the primer extension assay was repeated at least 4 times and gave consistent results. No major start site was found downstream of the Cla I site.

All Cm^r mRNA genes transcribed *in vivo* by the class of promoter constructions considered here carry most—if not all—untranslated sequences of w.t. Cm^r mRNA, involved in ribosome attachment and assembly. It is therefore very likely that mRNA species produced by this class of promoter constructions are translation templates of equivalent efficiency. This is further supported by our preliminary results of mRNA quantitation *in vivo* , which show for this class of promoter constructions an almost constant ratio of the amount of mRNA to antibiotic resistance as given in table I (M.A.Jacquet, unpublished).

DISCUSSION AND CONCLUSIONS

Design and test of 'modular' promoters.

A promoter can be envisioned as a series of instructions needed for transcription : tight binding of RNAP, closed, then open complex formation, start address, initiation frequency, regulatory instructions, etc... It is likely that each instruction is best expressed by a specific sequence, but slight alterations of the latter would not supress the instructions, only reduce its effect to some extent.

Control and regulation of expression of a particular gene sets specific demands on its promoter, which could be met by selecting a given set of promoter instructions. The set could be hierarchized by suitable alterations of the sequences encoding the instructions. We believe that this, and the squeezing due to the finite promoter size, are the main reasons for the observed promoter sequence diversity.

Adopting this view, one could try to take apart the set of instructions of a natural promoter and study individually each instruction for its role in that promoter. We took a complementary approach, by introducing, in a sequence environment as neutral as possible, one or a few putative promoter instructions and check for their possible ability to express a gene *in vivo*, then to study the molecular mechanism of their activity, if any. To this end, we designed sequences carrying both standard CHs, framed with unfrequent restriction sites enabling easy handling of these or of other known or anticipated promoter elements.

Once ligated upstream of a selected test gene, the construction can be tested for its ability to promote gene expression. To be conclusive, testing of synthetic promoters or promoter elements should be performed in a DNA context devoid of uncontrolled promoter activity ; this requirement is not necessarily met—and difficult to control—if the synthetic sequence is inserted in, or in the context of, a w.t. promoter, as was the case for most studies of synthetic promoter or promoter elements published so far. Since it is at present difficult

to recognize promoter elements outside of the -10 to -35 promoter core, the safest is to remove as much sequence up and downstream of the w.t. promoter as compatible with *in vivo* expression of the gene, prior to introducing the construction to be tested.

The primary goal of the present study was twofold : is a pair of standard CH spaced by 17bp sufficient information to promote transcription ? What is the role of the sequence of the 17bp spacer ? Unambiguous answers are provided by the planned constructions, pBR GC-1 and pBR AT-10.

A pair of CH, spaced by 17 bp, is sufficient for strong gene expression in vivo.

When ligated in front of the Cm^r gene, at its transcription start site, both pBR AT-10 and pBR GC-1 support indeed very efficient gene expression *in vivo*, almost twice that of the w.t. promoter they replace. This confirms and quantitates previous findings, for instance those obtained by inserting, in place of the core of the w.t. promoter of Tet^r, a sequence carrying the standard CH pair, spaced by a 17 bp 'consensus' sequence (11).

Our results go however one step further, as our constructions deliberately avoid 'consensus' in the spacer sequence, and operate in a sequence context from which w.t. promoter elements have been removed to the best of our knowledge, at least upstream of +1. We conclude that a pair of CH sequences, spaced by 17 bp, is *sufficient* by itself for strong gene expression *in vivo*. This statement does not exclude that putative late promoter elements (5), which could be deposited in the transcribed 5′ end of the Cm^r gene and hence would be present in our constructions, could contribute to the expression (see below).

The sequence of the 17 bp spacer and sequence downstream of CH(−10) involved in setting transcription start.

Our laboratory had been previously active in analyzing composition constraints in natural DNA sequences (AT/GC, or purine/pyrimidine (Pu/Py) constraints). The constraints specify well delimited domains displaying dedicated physical properties (18). When applied to *E.coli* promoter sequences, a consensus domain map is found, in which the two domains harboring the CH pair (obviously rich in AT) were separated by a spacer domain higher in GC (19). We therefore decided, as a first application of our modular promoter concept, to synthesize promoter sequences carrying both standard CH, separated by a spacer of 17 bp of either high GC or high AT contents, to see the effect of these composition constraints on activity. Particular features such as runs of Pu (in particular A), Pu/Py alternance (specially GC) or palindromic sequences were avoided because of associated unusual conformations.

Both pBR AT-10 and pBR GC-1 support the same high level of gene expression *in vivo*. Thus, for a spacer of 17 bp, AT or GC constraints do not affect expression efficiency *in vivo*, a conclusion holding at least for our spacer sequences. However, this conclusion may be valid only for promoters bearing the standard CH pair, as it was shown by Auble et al. (20) that the sequence of a 17 bp spacer, linking *non*-standard CH pairs, can modulate gene expression *in vivo*.

A striking difference between pBR GC-1 and pBR AT-10 is the location and homogeneity of their Cm^r transcription start sites. We observe that close to 100 % of the Cm^r transcripts in pBR GC-1 start at the conventional place, 6 bp downstream of CH(−10). In pBR AT-10, Cm^r transcription start is scattered over a 9 bp sequence, and occurs at 3 main sites; about 2/3 of the transcripts start in CH(−10) (result not shown, our unpublished observations). This surprising location has been consistently found in 10 separate primer extension experiments carried out with Cm^r transcripts from pBR AT-10. Systematic

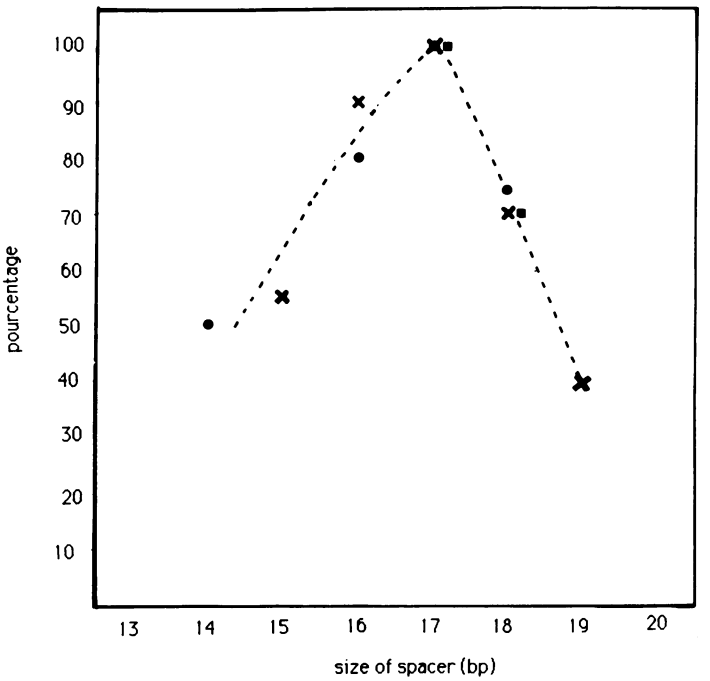

Fig.5 Gene expression
Level of gene expression conferred by a synthetic promoter construction bearing a standard CH pair, as a function of spacer size (relative to 17 bp spacer taken as 100 %).
● present work
× constructions of Aoyama et al. (24) tested in (26)
■ taken from Brosius et al. (25)

artifacts in these experiments are unlikely, as in other cases (PBR329 for instance), they gave the expected start sites. As pBR GC-1 and pBR AT-10 differ only in their spacer sequence (and by a 1 bp deletion downstream of CH(−10) in the latter), the results draw attention on contribution from the spacer sequence in setting transcription start.

Additional information on start site determinants is provided by the rearranged sequences, pBR AT-2 and -5, which differ from pBR AT-10 only by the sequences linking CH(-10) to the ClaI site. Their transcription starts differ from those found in pBR AT-10, as pBR AT-2 starts in CH(−10), but at a single site mainly, and pBR AT-5 at the conventional place, 7 bp downstream of CH(−10). We are left with the conclusion that for promoters carrying the standard CH pair, transcription start is not under exclusive control of the pair, in particular of CH(−10) ; the sequence of the 17 bp spacer, and/or that downstream of CH(−10) are also relevant elements. Contribution of sequences downstream of CH(−10) have been stressed earlier (5).

Recent findings reported by Horwitz and Loeb (21), who observe that modified promoter (with non-standard CH pairs) can have transcription start shifted towards CH(−10), ressemble our findings with pBR AT-10, -2 (and also pBR AT-1 and pBR GC-8). They attribute the effect to cryptic CH elements. Surrogate (−10) elements can be found in the AT rich spacer of pBR AT-10 like TAGATT, TTATAT or TATGAT, but the associated

(−35) sequence would match at best the GA pair of CH(−35), which is present also—but not active—in pBR GC-1. These surrogate elements are shared by pBR AT-5 but are not active there. It is furthermore difficult to concile the dominance of sequences with poor fit to consensus over a pair of standard CH, and the generally accepted (but lately challenged (22)) rule 'consensus is strength' (see the discussion in (21)).

Effects of size of spacers linking a pair of standard CH.

Transformation with partially ligated DNA proves to be a valuable trick for generating, in the vicinity of the ligation site, a series of randomly located mutations of variable extension in addition to the designed sequence. The rearranged sequences can be screened for a selected element, like for instance the size of the spacer linking a pair of standard CH. This is the case for rearranged plasmids pBR AT-1, -4 and pBR GC-8, -6 (the latter with a transversion in CH(−10)) which have spacer size ranging from 14 to 26 bp. The plot of *in vivo* gene expression *vs* spacer size (Fig.5) confirm that 17 bp is the optimum size, but evidences in addition three novel features : (i) efficiency drops steadily on both sides of the 17 bp optimum, more abruptly upon increasing than reducing the spacer size. (ii) for spacers of given size, the precise sequences seems not to affect relative effiency, which is consistent with our conclusion in the previous section ; (iii) most importantly, the drop is rather milde, since mutants with 14 to 26 bp spacers display a twofold variation of gene expression only ; this is in strong contrast with the effect observed in natural promoters (with non-standard CH), where departure from the natural 17 bp spacer by as little as ±1 bp may reduce gene expression by as much as an order of magnitude (20,23,24). As shown earlier, the spacer size is a negative promoter modulator ; our results suggest that its strength can be tuned by the fit of the consensus hexamers to the standard : the better the fit, the smaller the modulation effect.

Sequences of the CH type not necessary for efficient gene expression

Among the rearranged promoter sequences selected, fig.1, those having deleted one CH, i.e. pBR GC-2, pBR GC-4, pBR AT-6 and pBR AT-8, still confer Cm resistance, but their efficiencies have dropped one order of magnitude below that conferred by promoters bearing the CH pair. This result is consistent with earlier studies (27,28), although results reported recently (21) show that removal of one CH from a w.t. promoter does not systematically result in a drop of promoter strength, possibly through rescue from some cryptic CH element. A first example of a sequence displaying good promoter activity (40% w.t.), but missing discernable CH elements both 10 and 35 bp upstream of transcription start, is provided by the rearranged sequence pBR AT-3. Transcription of the Cm^r gene starts in pBR AT-3 in ClaI, at the boundary with HindIII. The corresponding −10 and −35 hexamers read TGTCAA and GTAGAG, respectively. The latter, completely out of consensus (TTGACA) is borrowed from the Tc^r coding sequence just downstream of BamH I. The former is of unknown origin ; it shares 3 bases with CH(−10) (TATAAT), but misses both most conserved ones (first A and last T). This suggests that the promoter signal(s) carried in place of the CH pair can be mimicked efficiently by sequences, or sequence combinations, altered from consensus beyond recognition. Cm resistance conferred by pBR AT-3 is stronger than that observed for inserts carrying one standard CH only, but less than that of inserts keeping both. This is strong indication that efficient supplementing of one CH depends stringently on the sequence occuring at the place of the other : whether consensus or not, these sequences seem to come in efficient constitutive promoters as matched pairs, indicative of the synergic action of these signals. The identity of the cryptic (synonymous) signal(s) involved is at present unknown.

Work is is progress to further characterize the promoter activity of BR GC-1 and pBR AT-10 *in vivo*, and of selected rearranged sequences. These extreme simplified but active promoters sequences could help disclosing the basic feature of promoter recognition, complex formation and activation (M.A.Jacquet, work in progress).

ACKNOWLEDGEMENTS

We wish to thank Prof. R.M.Hochstrasser for comments, Dr Peter Brooks for help with the manuscript, Prof. T.Aoyama for gift of the constructions described in (24)(see fig.5), C. Michon-Dubucs for oligonucleotide synthesis, L. Corme for typing the manuscript and R. Schwartzmann for photographical work.

*To whom correspondence should be addressed

+On leave of absence; present address: Instituto de Investigaciones Biologicas Clemente Estable, Avda. Italia 3318, Montevideo and Facultad de Humanidades y Ciencas, Tristan, Narvaja 1674, Montevideo, Uruguay

REFERENCES

1. Harley,C.B. and Reynolds,R.P. (1987), Nucl.Acid Res., **15**, 2343-51.
2. Stormo, G.D., Schneider, T.D., Gold, L. and Ehrenfeucht, A. (1982), Nucl. Acid Res., **10**, 2997-3011.
3. Lanzer,M. and Bujard,H. (1988), Proc.Natl.Acad.Sci. USA, **85**, 8973-77.
4. Bossi,L. and Smith,Y.M. (1984), Cell, **39**, 643-52
5. Deuschle, U., Kammerer, W., Gentz, R. and Bujard, H. (1986), Embo J., **5**, 2987-94.
6. Yu,X.M. and Reznikoff,W.S. (1986), J.Mol.Biol., **188**, 545-53.
7. Okamoto,T., Sugimoto,K., Sugisaki,H. and Takanami,M. (1977), Nucl. Acid Res., **4**, 2213-22.
8. Duval-Valentin,G. and Ehrlich,R. (1987), Nucl.Acid Res., **15**, 575-94.
9. Duval-Valentin,G. and Ehrlich,R. (1988), Nucl.Acid Res., **16**, 2031-44.
10. 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.
11. Rossi,J.J., Soberon,X., Marumoto,Y., Mc Mahon,J. and Itekura,K. (1983), Proc.Natl.Acad.Sci. USA, **80**, 3203-07.
12. Covarrubias,L. and Bolivar,F. (1982), Gene, **17**, 79-90.
13. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982), Molecular Cloning : A Laboratory Manual, Cold Spring Harbor Lab.
14. Maxam,A.M. and Gilbert,W. (1980), Methods in Enzymol., **65**, 499-560.
15. Jacquet,M.A. and Ehrlich,R. (1985), Biochimie, **67**, 987-97.
16. Ernoult-Lange,M. and May,E. (1983), J.Virol., **46**, 759.
17. Conley,E.C., Saunders,V.A. and Saunders,J.R. (1986), Nucl. Acid Res., **14**, 8905-17.
18. Marcaud,H., Gabarro-Arpa,J., Ehrlich,R. and Reiss,C. (1986), Nucl.Acid Res., **14**, 551-58.
19. Ehrlich,R., Marin,M., Larousse,A., Gabarro-Arpa,J., Schmitt,B. and Reiss,C. (1984), Folia Biologica, 105-18.
20. Auble,D.T., Allen,T.L. and DeHaseth,P.L. (1986), J.Biol.Chem., **261**, 11202-06.
21. Horwitz,M.S.Z. and Loeb,L.A. (1988), J.Biol.Chem., **263**, 14724-14731.
22. Knaus,R. and Bujard,H. (1988), EMBO J., **7**, 2919-2923.
23. Stefano,J.E.and Gralla,J.D. (1982), Proc.Natl.Acad.Sci.USA, **79**, 1069-72.
24. Aoyama,T., Takanami,M., Ohtsuka,E.,Tanaiyama,Y.,Marumoto,R., Sato,H. and Ikemara,M. (1983), Nucl.Acid Res., **11**, 5855-64.
25. Brosius,J. Erfle,M. and Storella,J. (1985), J.Biol.Chem., **260**, 3539-41.
26. Jacquet,M.A., (1987), Doctoral thesis, University ParisVII.
27. Ponnambalam,S., Webster,C., Bingham,A. and Busby,S. (1986), J.Biol.Chem., **261**, 16043-48.
28. Keilty,S. and Rosensberg,M. (1987), J.Biol.Chem., **262**, 6389-95.