JOURNAL OF BACTERIOLOGY, Oct. 1985, p. 70–77
0021-9193/85/100070-08$02.00/0

Vol. 164, No. 1

Promoters Recognized by *Escherichia coli* RNA Polymerase Selected by Function: Highly Efficient Promoters from Bacteriophage T5

REINER GENTZ† AND HERMANN BUJARD†*

Molekulare Genetik der Universität Heidelberg, D-6900 Heidelberg, Federal Republic of Germany

Received 6 March 1985/Accepted 24 June 1985

Highly efficient promoters of coliphage T5 were identified by selecting for functional properties. Eleven such promoters belonging to all three expression classes of the phage were analyzed. Their average AT content was 75% and reached 83% in subregions of the sequences. Besides the well-known conserved sequences around −10 and −33, they exhibited homologies outside the region commonly considered to be essential for promoter function. Interestingly, the consensus hexamers around −10 (TAT AAT) and −35 (TTG ACA) were never found simultaneously within the sequence of highly efficient promoters. Several of these promoters compete extremely well for *Escherichia coli* RNA polymerase and can be used for the efficient in vitro synthesis of defined RNA species. In addition, some of these promoters accept 7-mGpppA as the starting dinucleotide, thus producing capped mRNA in vitro which can be utilized in various eucaryotic translation systems.

Promoters of the *Escherichia coli* system start synthesis of functional RNAs with vastly different efficiencies. Little is known, however, about the rules by which functional parameters are implemented within a promoter sequence. Despite our knowledge of more than 150 promoter sequences (10) and a wealth of genetic and biochemical data (19), we are still unable to make reasonable predictions on functional properties of a promoter from structural information alone. Consensus sequences of *E. coli* promoters derived from sequence compilations have elucidated some important general features. However, synthesis of consensus promoters (5) have resulted in signals which are, at most, average in function (U. Deuschle and M. Kammerer, personal communication). This is not surprising if one considers the complexity of the process programmed by a promoter sequence as well as the fact that in the derivation of consensus sequences there is usually no value describing functional parameters given to individual sequences.

We approached this problem in a different way. By selecting for the most efficient unregulated promoters in the *E. coli* system, we expected to reveal sequences which would exhibit pertinent structural features most clearly. The selection principles utilized for the identification of efficient promoters were the determination of (i) the rate of complex formation between RNA polymerase and promoter in vitro, (ii) the relative efficiency of RNA synthesis in vitro under competitive conditions, and (iii) the relative promoter strength in vivo.

The in vitro analysis of promoter-carrying DNA fragments has been described previously (6, 7). For the in vivo study of promoters we developed cloning systems which allow the stable integration of strong promoters as well as the precise determination of their in vivo function (9, 21; U. Deuschle, M.S. thesis, University of Heidelberg, 1984). Of about 60 promoters tested (including those of coliphage T7, fd, and λ) some of the most efficient signals were found in the genome of coliphage T5. Here we describe the application of the pDS1 vector system (21; Fig. 1) for the selective cloning of strong promoters, the identification and structural analysis of 11 promoters of the phage T5 genome, and some of the functional properties of these promoters. As can be seen from the results of this and previous studies (9), several promoters described here appear especially useful for the efficient in vitro synthesis of defined RNA species, and as some of the promoters accept 7-mGpppA as the starting dinucleotide capped RNAs can be directly obtained in vitro. This transcription-coupled capping allows an efficient and selective expression of cloned DNA sequences in vitro which has been found to be especially useful in studying the translocation of proteins into or through membranes (11, 23).

MATERIALS AND METHODS

Enzymes and chemicals. Restriction enzymes, T4 DNA ligase, calf intestinal alkaline phosphatase, and RNase T1 were purchased from Bethesda Research Laboratories, Gaithersburg, Md.; New England Biolabs, Inc., Beverly, Mass.; or Boehringer Mannheim Biochemicals, Indianapolis, Ind.; and T4 DNA kinase was obtained from H. Schaller (University of Heidelberg). Reactions were carried out as recommended by the supplier. The isolation of bacteriophage T5 DNA and *E. coli* RNA polymerase has been described previously (7). *Xho*I synthetic linkers were obtained from Collaborative Research, Inc., (Waltham, Mass.) and were present in ligation assays in a 20-fold molar excess relative to that of the various DNA fragments. [γ-^{32}P]ATP and [α-^{32}P]UTP were from Amersham & Buchler (Braunschweig, Federal Republic of Germany) and 7-mGpppA was obtained from P-L Biochemicals, Milwaukee, Wis.

Plasmids and their nomenclature. The basic pDS1 vector system has been described previously, and here we follow previously proposed nomenclature (21). The identity of the promoters and terminators which have been integrated can be derived from the designation of the plasmid: pDS1/p_{H207},$t_0$1 describes a plasmid-carrying promoter p_{H207} in front of the coding sequence (dhfr) for dihydrofolate reductase (DHFR) and terminator t_0 from phage lambda at site 1 (Fig. 1). Another terminator used was t_{fd} from coliphage fd (9). For the correct in-frame positioning of the

* Corresponding author.
† Present address: F. Hoffmann-La Roche & Co. A.G., ZFE CH 4002 Basel, Switzerland.

translational start sites within cloned promoter-carrying fragments with respect to the dhfr sequence, *Bam*HI linkers of 8, 10, and 12 base pairs (bp) in length were inserted into pDS1. The length of the *Bam*HI linker relating to the reading frame can also be derived from the designation of a plasmid. pDS1-8/p_{H207},$t_0$1 denotes the same plasmid as described above which carries a *Bam*HI linker of 8 bp.

Cloning of promoter-carrying T5 DNA fragments. Wild-type T5 DNA was digested with *Hin*dIII and separated by polyacrylamide gel electrophoresis (PAGE; 6%, 14 h, 5 V/cm, TBE buffer [90 mM Tris-hydrochloride, 90 mM sodium borate, 3 mM EDTA, pH 8.3]), and the fragments A through H, IJ, KL, M, and N (8) were isolated by electroelution (1 h, 150 V, buffer [5 mM Tris-hydrochloride, 0.5 mM EDTA, pH 7.6]). After the fragments were concentrated by ethanol precipitation, they were digested with either *Hae*III or *Alu*I and *Rsa*I and then fused to synthetic *Xho*I linkers by ligation. Upon digestion with *Xho*I, the fragments were cloned into pDS1 vectors by standard procedures. Transformation of *E. coli* M15 (9) was carried out as described by Morrison (17). Selection of transformants was for a high level of resistance to chloramphenicol (CM) (100 to 400 μg/ml). Selected colonies were analyzed by isolating their plasmid DNA from 1.5-ml cultures (1) and by examining the restriction patterns of the DNAs by PAGE; fractions of the culture were subjected to sodium dodecyl sulfate (SDS)-PAGE, and the presence of chloramphenicol acetyl transferase (CAT) was monitored after the protein pattern was visualized with Coomassie blue. Plasmids were isolated in preparative amounts by the method of Clewell and Helinski (3) and Radloff et al. (18).

In vitro transcription. Standard in vitro transcriptions were carried out in a volume of 100 μl containing 20 mM Tris-hydrochloride (pH 8), 10 mM $MgCl_2$, 0.1 mM EDTA, 120 mM KCl, 5% glycerol, 2 mM dithioerythritol, 400 μM ATP, 400 μM GTP, 200 μM CTP, 50 or 100 μM UTP, and 0.5 μM [α-^{32}P]UTP (3,000 Ci/mmol). For the synthesis of 5′-labeled RNA the concentration of ATP was lowered to 12 μM, i.e., 9 μM unlabeled and 3 μM labeled ATP [γ-^{32}P]ATP; ≧3,000 Ci/mmol). The concentration of UTP in these assays was 150 μM. The reactions were initiated by the addition of approximately 0.5 U of RNA polymerase per pmol of DNA (corresponding to five enzyme molecules per plasmid) and terminated by the addition of EDTA (final concentration, 40 mM) and phenol. Subsequent to phenol extraction, the nonincorporated nucleotides were separated by chromatography on Sephadex G57 (2-ml column); and the RNA-containing fractions were collected, precipitated by ethanol suspended in TE buffer, and, after the addition of three volumes of 90% formamide, heated to 90°C for 30 s and analyzed by PAGE in 7 M urea (8% 20 V/cm). Transcription-coupled capping was achieved by lowering the concentration of ATP in the standard assay to 5 μM and by including 250 μM 7-mGpppA. To demonstrate the incorporation of the dinucleotide, short transcripts were produced by replacing CTP with methyl-CTP (200 μM final concentration). After the assays were incubated for 5 min at 37°C, the concentration of ATP was raised to 1 mM and the reaction was allowed to proceed for another 10 min. The controls contained 400 μM ATP and no 7-mGpppA.

Analysis of in vitro transcripts with RNase T1. In vitro transcripts end labeled with [γ-^{32}P]ATP were eluted from 8% polyacrylamide gels containing 7 M urea. After precipitation with ethanol, the transcripts were suspended in a buffer containing 25 mM sodium citrate (pH 5.0), 7 M urea, and 1 mM EDTA (final volume, 5 μl). Upon the addition of 1 μg of tRNA per assay the transcripts were incubated with 1 U of T1 RNase for 15 min at 55°C. The reactions were stopped by freezing the samples in dry ice and then storing at −20°C. Immediately before the gel was loaded, the samples were heated to 90°C for 30 s and chilled on ice. For standardization of the electrophoretic pattern, fractions of the various transcripts were subjected to limited alkaline hydrolysis. A typical assay contained the transcript together with 5 μg of tRNA in 50 mM sodium bicarbonate-carbonate (pH 9.2; volume, 10 μl). The samples were incubated at 90°C for 7 min and then chilled on ice. Just prior to loading on a sequencing gel, an equal volume (10 μl) of 10 M urea–1.5 mM EDTA containing 0.05% each of xylene cyanol and bromphenol blue was added, and the samples were heated for 30 s at 90°C and chilled on ice.

DNA sequencing. DNA was sequenced by the method of Maxam and Gilbert (14, 15). Fragments were dephosphorylated (calf intestinal alkaline phosphatase), labeled at the 5′ termini with T4 polynucleotide kinase and [γ-^{32}P]ATP, and then subjected to secondary restriction endonuclease digestion to generate fragments labeled at one end only. Separation of the fragment by PAGE and elution of the radioactive material from the gels resulted in end-labeled DNA, the

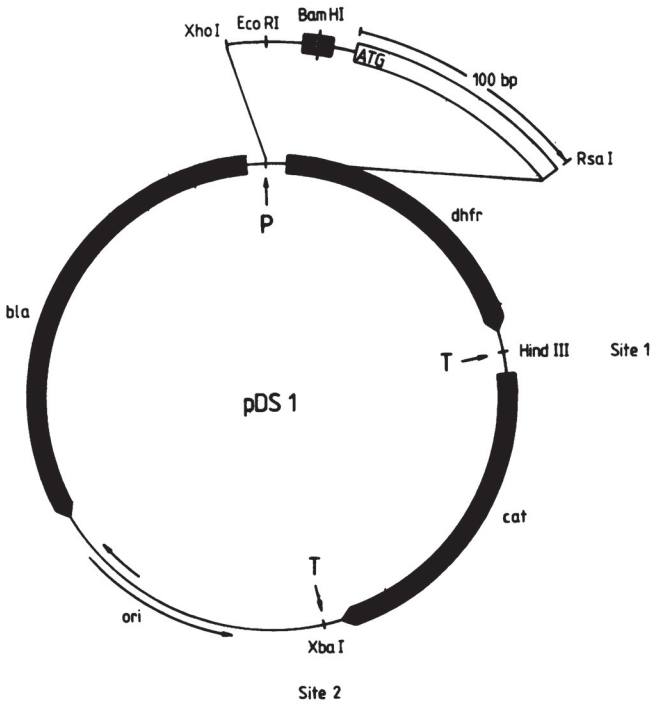

FIG. 1. pDS1 cloning system. The principle of this vector system has been described previously (21). The plasmid contains the ColE1 replicon (*ori*; arrows indicate the direction of transcription of RNA I and RNA II; 12) and, as a selectable marker, ampicillin resistance (conferred by the *bla* gene). Two indicator functions, encoded by the sequence for *dhfr* and *cat*, are fused in such a way that they can be brought under the control of one promoter (P) but can also be separated by inserting a terminator (T) at site 1. A second terminator at site 2 can prevent transcriptional readthrough into the replication region. Whereas the *cat* gene contains its genuine ribosomal binding site, the *dhfr* sequence is not preceeded by such a signal. The *Bam*HI site was created by inserting synthetic linkers of 8, 10, and 12 bp. The resulting plasmids allow the in-frame positioning of the *dhfr* sequence to any upstream translational initiation site. The first *Rsa*I cleavage site within the *dhfr* sequence which was utilized for the production of runoff transcripts is indicated. The nomenclature used with this vector system is described in the text.

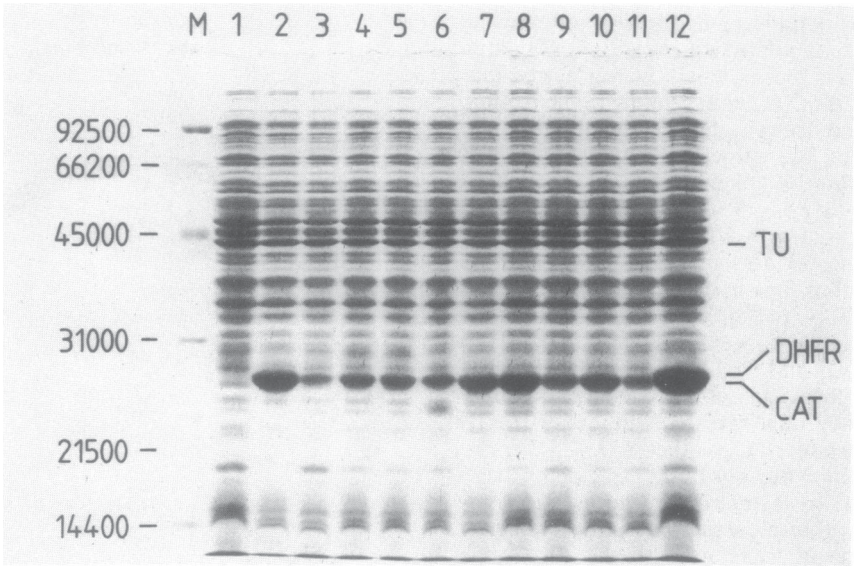

FIG. 2. Analysis of total protein from *E. coli* M15 cells harboring T5 promoter-carrying pDS1 plasmids. DNA fragments identified in the first cloning experiments were reintegrated into plasmids pDS1- (8,10,12)/$t_0$2 (Fig. 1). Colonies highly resistant to CM were grown in LB medium, and fractions of the culture were analyzed by SDS-PAGE as described previously (21). The positions of CAT, TU (elongation factor of translation), and a fusion product of DHFR are indicated. Molecular weights are shown on the left. Lanes M and 1 show a size marker and the pattern of plasmid-free M15 cells, respectively, whereas lanes 2 through 11 exhibit the proteins from clones carrying the following T5 DNA fragments or promoters, respectively: F20, F33, p_{H207}, p_{N25}, F81 (carrying p_{N26}), F30, F41, F22, F5, F25*. Lane 12 shows the effect of positioning the *dhfr* sequence in-frame with a translational start signal located in fragment F25*. Excessive amounts of a DHFR fusion protein were produced. Although in this particular case the fusion product is obscuring the CAT protein, the presence of both proteins has been demonstrated (see text).

sequence of which was determined by standard procedures. In all cases both strands were sequenced.

Analysis of cloned DNA fragments by Southern hybridization. Wild-type T5 DNA, digested with *Hin*dIII, was separated on a 0.5% agarose gel (14 h, 3 V/cm, 1× Loening buffer [36 mM Tris-hydrochloride, 30 mM NaH_2PO_4, 1 mM EDTA, pH 7.6]). After being stained with ethidium bromide, the gel was photographed together with a measuring tape. To ensure good transfer of large fragments from the gel, it was soaked for 1 min in 100 mM HCl, neutralized (two times for 10 min each in 0.5 M Tris-hydrochloride, pH 7.5), and then denatured in 0.5 M NaOH for 20 min. Upon neutralization (see above), the DNA was transferred from the gel to nitrocellulose filter paper (0.45 μm; Schleicher & Schuell, Inc.) by the method of Southern (20). The filters were marked to indicate the positions of the fragments, rinsed in 2× SSC (1× SSC is 0.15 M NaCl plus 15 mM sodium citrate), and baked in vacuo (2 h at 80°C). For hybridization, filters were pretreated for 1 h at 68°C in hybridization buffer (5× SSC, 50% formamide, 1× Denhardt reagent [4]). Hybridizations were carried out in 12-ml plastic vials containing 1 ml of hybridization buffer, approximately 10^5 cpm (Cerenkov) of the denatured DNA probe (1 min at 95°C in 0.5 × TE buffer), and the filter strip. Hybridizations were allowed to proceed at 42°C for 8 to 14 h on a rotary shaker. For some fragments (e.g., p_{H22} and p_{F30}) the conditions were optimized to reduce cross-hybridizations (hybridization temperature, 42, 44, 46, or 48°C). Subsequently, the filters were washed three times at 42°C in 5× SSC buffer–50% formamide and finally twice in 2× SSC at room temperature. Filters were then air dried, and autoradiographs were prepared with Kodak XR-5 films.

RESULTS

Cloning of promoter-carrying fragments. For the cloning of the promoters of the different expression classes of bacteriophage T5, the DNA fragments of various digests of the 120-kilobase phage genome were fused to synthetic *Xho*I linkers and cloned into the *Xho*I site of pDS1-10/t_{fd}1 (Fig. 1; see above for nomenclature). Selection of transformed *E. coli* M15 was carried out on agar plates containing increasing amounts of CM, and colonies resistant to 100 μg/ml or more were grown and analyzed. All isolates showed at least one cloned fragment in the size range between 130 and 1000 bp. These fragments were isolated and recloned into pDS1-(8, 10, and 12)/$t_0$1 or pDS1-(8, 10, and 12)/$t_0$2. All of the more than 20 fragments gave rise to colonies highly resistant to CM (i.e., ≧100 μg/ml). Cultures prepared from such colonies (10 ml of Luria broth, 30 μg of CM per ml; 37°C overnight) were analyzed with respect to their protein pattern by SDS-PAGE. Considerable amounts of CAT protein were observed in all cases (Fig. 2, lanes 2 to 11). The effect of positioning a translational start signal of a promoter-carrying fragment in frame with the *dhfr* sequence is shown in Fig. 2, lane 12. The presence of the DHFR fusion protein obscuring the CAT protein (Fig. 2, lane 12) has been confirmed by dilution of the sample (thereby generating two separable bands), as well as by immunoblotting with DHFR- and CAT-specific antibodies (data not shown). In this way we not only identified efficient promoters but also interesting translational start signals which will be described elsewhere.

In vitro transcription of the cloned fragments and analysis of the transcripts. To verify the presence of promoters within the cloned material, the various chimeric plasmids were digested with *Xho*I, and the liberated fragments were, after separation by PAGE and elution from the gel used as templates for in vitro transcription. Alternatively, the plasmids were digested with *Rsa*I, and the mixtures of fragments were transcribed, producing run-off transcripts from the cloned promoter to the first *Rsa*I site within the *dhfr* sequence (Fig. 1). At limiting amounts of RNA polymerase,

the run-off transcripts were labeled with either [γ-^{32}P]ATP or [α-^{32}P]UTP. In all but one case (fragment F 22) transcripts of the same length were obtained from a given template independent of the method of labeling. No label was introduced into the RNA originating from the fragment F 22 if [γ-^{32}P]ATP was used. Consequently, this transcript does not start with an adenosine. Efficient transcription can be initiated, however, with the dinucleotide uridylyl-(3',5')uridine (data not shown); this promoter is therefore most likely an U-starter (see below). Two transcripts of different lengths were produced in comparable amounts from fragment F 41, indicating the presence of two promoters within this template separated by about 100 bp (Fig. 3). In all cases the RNA populations obtained originated almost exclusively at the phage promoters (Fig. 3; see below), demonstrating their ability to favorably compete with the plasmid promoters for RNA polymerase.

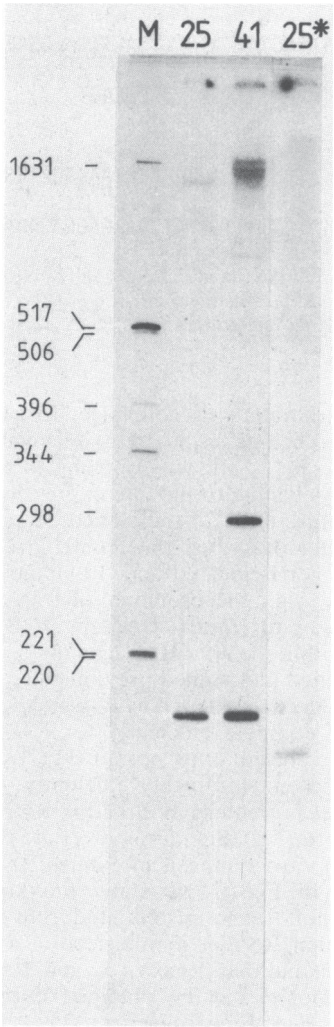

FIG. 3. In vitro transcription of chimeric plasmids. Plasmids of the type pDS1-8/p,$t_0$2 containing different promoter-carrying fragments were digested with *Rsa*I and transcribed in vitro (molar ratio of RNA polymerase to plasmid, 5:1). The labeled transcripts were separated by 7 M urea PAGE (8%, 20 V/cm, 3 h) and visualized by autoradiography. As indicated, the different plasmids contained fragments F25 (25), F41 (41), and F25* (25*), respectively; M denotes markers whose sizes are given in nucleotides on the left. In each case the dominating transcripts (≧95% of hybridizable counts) originated from the cloned promoters. The size of the transcripts (designated on the left) allows an estimation of the position of the promoter within the cloned fragment.

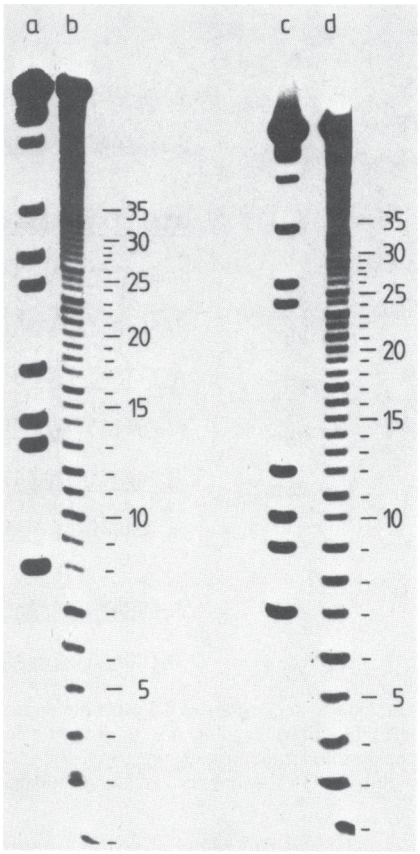

FIG. 4. Examples for the sequence analysis of transcripts. Transcripts subjected to either partial RNase T1 digestion or to limited alkaline hydrolysis were analyzed by PAGE (20%, 8.3 M urea, 50 V/cm, 2 h). Here the pattern originating from transcripts initiated by p_{K28a} and p_{G25} are shown in lanes a and c, respectively. The corresponding alkaline hydrolysates are in lanes b and d. Numbers on the right are distances to the 5' end of the transcripts in nucleotides.

Eleven different promoters were characterized within the 20 fragments originally cloned. Ten of these initiated their transcripts with adenosine and gave rise to relatively short RNAs. The lengths of these RNAs allowed us to estimate the positions of the promoters within the cloned fragments, which facilitated their further analysis.

Sequence analysis of promoter-carrying fragments and their major transcripts. DNA sequence determinations were carried out by the method of Maxam and Gilbert (14, 15). Fragments of 500 bp and less in size were directly end labeled, and after cleavage with a second restriction endonuclease they were separated by electrophoresis and sequenced. Large fragments were first digested with various restriction enzymes, and the promoter-carrying segments were determined by binding RNA polymerase by the nitrocellulose filter binding technique as described previously (data not shown; 7). Fragments defined in this way were end labeled, sequenced, and recloned into the *Xho*I or *Eco*RI site of pDS1/$t_0$1 (Fig. 1). In all cases the sequences for both strands were determined. The functional orientation of the various promoter sequences was confirmed by relating cleavage sites within the promoter-carrying fragments to the

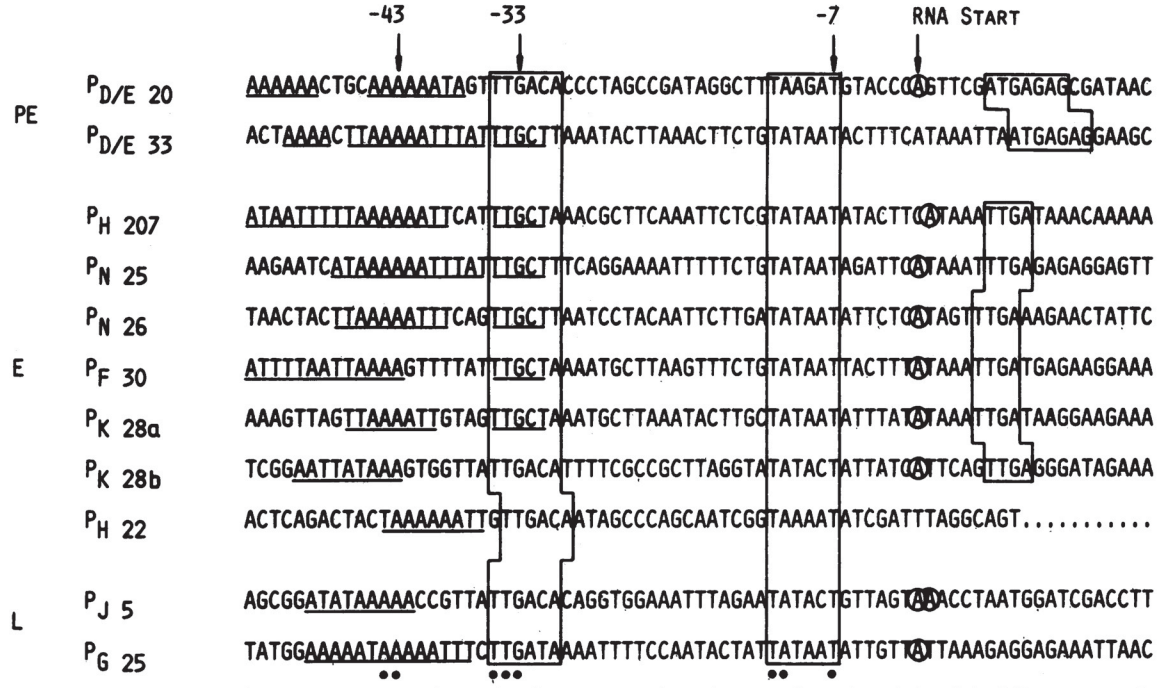

FIG. 5. Nucleotide sequences of T5 promoters. Regions of major sequence homology are boxed, and the eight fully conserved positions are indicated by filled circles. The AT-rich regions centered around −43 are underlined, and the starting nucleotide of the RNA is circled. Since p_{H22} can initiate transcription with uridylyl-(3′,5′)uridine (data not shown), the starting nucleotide of its RNA most likely must be positioned within the TTT sequence. Abbreviations: PE, preearly; E, early; L, late.

unique *Bam*HI site in the pDS1 plasmid (data not shown). The precise start point of transcription within a promoter sequence can be determined by comparing the DNA sequence coding for the 5′ terminal region of the transcript with the sequence of the corresponding RNA. We have therefore analyzed the G-pattern of transcripts labeled in vitro with [γ-^{32}P]ATP isolated from polyacrylamide gels. Examples for such analyses, which were carried out for all but two transcripts, are shown in Fig. 4. A compilation of promoter sequences characterized in this way, Fig. 5.

Mapping of the cloned promoters along the T5 genome. Forty-one major promoters of the T5 genome previously have been mapped with respect to their location within the different expression classes (preearly, early, late), as well as in relation to various restriction maps (22). Since the distribution of *Hin*dIII cleavage sites throughout the T5 DNA permits a rather good distinction between preearly, early, and late regions of the genome (8), we used *Hin*dIII digests of T5 wild-type DNA which were separated by PAGE and transferred to nitrocellulose for an analysis of various promoter-carrying fragments by the method of Southern (20). The ^{32}P-end-labeled fragments were hybridized to the immobilized T5 DNA under stringent conditions. Two fragments, F20 and F33 hybridized to both *Hin*dIII-D and -E (Fig. 6). Since these two *Hin*dIII fragments contain the terminally redundant preearly region, the promoters located in F20 and F33 must belong to that expression class. With the exception of F22, the remaining fragments hybridized to just one fragment from the *Hin*dIII digest of T5 DNA and can therefore be attributed rather precisely to their respective expression classes. Fragment F22 associated, even under the most stringent conditions, to *Hin*dIII-B, -D/E, -G, and -H. Whereas the strongest signal was clearly obtained with *Hin*dIII-H, considerable sequence homology must exist in the other regions as well. In conclusion, we identified and sequenced 11 promoters of coliphage T5. Two of these ($p_{D/E20}$ and $p_{D/E33}$) belong to the preearly, seven (p_{F30}, p_{H207}, p_{N25}, p_{N26}, p_{K28a}, p_{k28b}, and p_{H22}) belong to the early, and two (p_{J5} and p_{G25}) belong to the late expression classes.

The nomenclature used for the cloned T5 promoters is as follows. The letter described the *Hin*dIII fragment within which the promoter is located and the number defines the cloned fragment; e.g., the promoter of fragment F_5 is p_{J5} since F_5 hybridizes to *Hin*dIII fragment J. Exceptions are the promoters within the *Hin*dIII fragments N and K, which have been identified and named previously (2, 22).

RNA synthesis in vitro initiated by p_{N25} and p_{N26}. Promoters of coliphage T5 compete efficiently for RNA polymerase binding in the presence of other promoters (7). Some of them not only bind the enzyme highly efficiently, but they also outcompete other promoters in directing RNA synthesis in vivo and in vitro (6, 7, 9). Such promoters are therefore good candidates for in vitro synthesis of defined RNAs in preparative amounts. In Fig. 7 we show the kinetics of the production of runoff transcripts initiated from p_{N25} and p_{N26}. It can be seen that RNA is synthesized at a high rate for more than 30 min and that between 12 and 25 pmol of RNA per pmol of template can be obtained during this time, depending on the size of the transcript. More than 90% of the RNA synthesized consists of the expected runoff transcripts, and more than 95% of the total RNA is complementary to the coding strand. The larger transcripts seen in the upper insert of Fig. 7 were not analyzed in detail, but we suspect that they originate from incompletely digested templates.

Transcription-coupled capping in vitro. Various promoters were tested for their ability to initiate in vitro transcription with 7-mGpppA. Successful incorporation can be monitored either by translating the resulting RNA in an eucaryotic translation system (23) or by analyzing short transcripts by urea PAGE. Short transcripts were obtained by replacing

CTP with methyl-CTP, which stops transcription usually at the position of the first and second cytosine, resulting in two major RNA species. Transcripts obtained in this way from promoters p_{K28a}, p_{N26}, and p_{N25} are shown in Fig. 8. The length of the major RNA species was between 15 and 41 nucleotides, and a shift in migration corresponding to about 2 nucleotides was introduced by the incorporation of 7-mGpppA. The three promoters described here were found to be among the most efficient ones for the in vitro synthesis of capped RNAs.

DISCUSSION

The family of promoters described in this report was discovered previously by examining specific in vitro properties of a variety of promoters (6, 7). The functions monitored were the relative rate of complex formation between *E. coli* RNA polymerase and promoters, and the capacity of RNA synthesis under competitive conditions. Using the pDS1 plasmid system (Fig. 1), strong promoters were also readily cloned under conditions which, again, allowed a selection for function but this time in vivo. CM resistance as well as CAT or CAT and DHFR protein production of the plasmid-containing cell were used as markers. The promoters isolated in this way exhibit remarkable structural and functional properties. There are striking commonalties among the 11 promoter sequences despite the fact that they originate from the three different expression classes of phage T5 (Fig. 6). (i) The AT content of the relevant sequences from +20 to −55, (the regions in contact with RNA polymerase; data not shown) is on the average 75%, with blocks of up to 83% AT between −56 and −36 as well as between −12 and +8. (ii) There are highly conserved regions around +7, +1, −10, −33, and −43. (iii) The distance between the −10 and the −33 region is 17 bp for all but one promoter, p_{H22}, which is only 16 bp. This latter promoter is also the only one which most likely initiates RNA synthesis with a uridine instead of

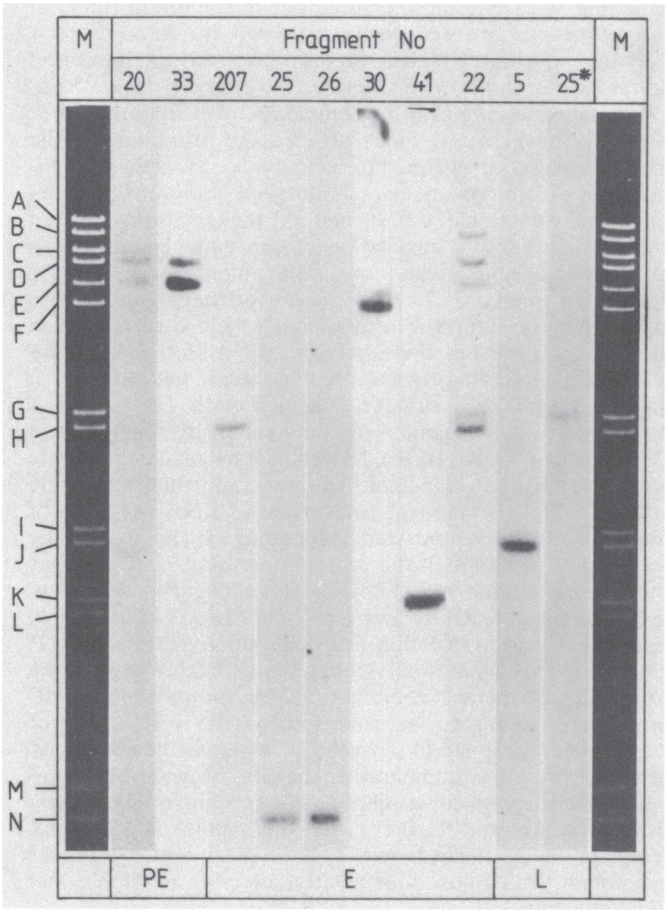

FIG. 6. Mapping of the positions of the cloned promoter fragments within the T5 genome. The autoradiogram depicts the result of Southern hybridizations of various labeled promoter-carrying fragments to a *Hin*dIII digest of T5 wild-type DNA separated by agarose gel electrophoresis. The ethidium bromide-stained pattern of the *Hin*dIII digest is shown in lanes M. The numbers above each lane designate the fragments. The correlation of the *Hin*dIII cleavage map with the genetic and functional organization of the T5 genome has been described previously (8, 22). PE, E, and L designate the preearly, early, and late expresion classes of phage T5.

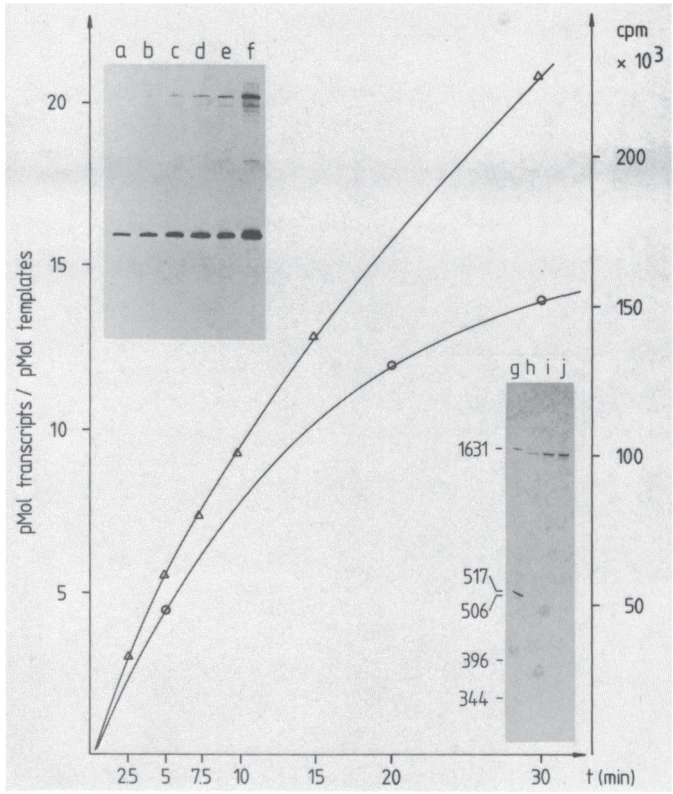

FIG. 7. Time course of in vitro RNA synthesis governed by p_{N25} and p_{N26}. RNA synthesis was allowed to proceed under standard conditions in 150 mM KCl and in presence of [α-^{32}P]UTP. The templates used were pDS1-10/p_{N25},$t_0$2, digested with *Rsa*I (△, upper insert), and pDS1-8/p_{N26},$t_0$2, digested with *Xba*I (○, lower insert). Samples of the reaction mixture were withdrawn at the times (t) indicated and divided into two fractions, of which one was analyzed by PAGE (inserts) and the other was used to monitor the acid-precipitable (5% trichloroacetic acid) counts (Cerenkov). It can be seen that RNA is produced at good rates for at least 30 min. The inserts show that the transcripts obtained are rather homogenous. More than 90% of the acid-precipitable counts can be hybridized to the single-stranded coding regions of the respective transcripts. Whereas the runoff RNA obtained with the *Rsa*I-digested plasmid is 192 nucleotides in length (upper insert), the size of the transcript shown in the lower insert is about 1,600 nucleotides (lane g, size marker). Lanes a through f in the upper insert show the PAGE analysis of the samples taken at times 2.5 through 30 min, and lanes h through j of the lower insert depict the corresponding analysis of samples taken at 5, 20, and 30 min, respectively, from transcription assays with *Xba*I-digested plasmid used as the template.

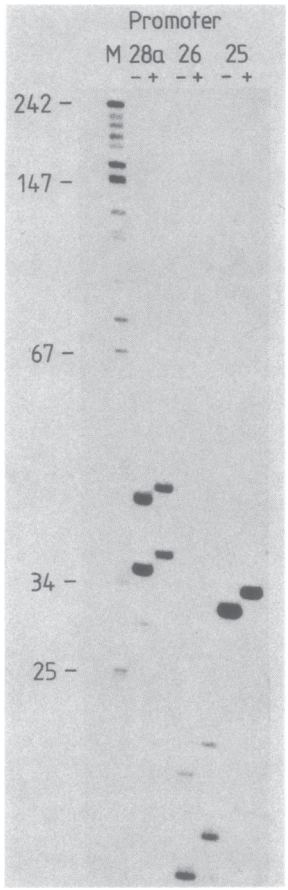

FIG. 8. Transcription-coupled capping in vitro. Plasmids of the type pDS1-10/p,$t_0$2 (form I DNA) containing the promoters p_{K28a}, p_{N26}, p_{N25}, respectively, were transcribed under standard conditions with (+) or without (−) 7-mGpppA in the presence of methyl-CTP instead of CTP. The labeled transcripts were separated by 7 M urea PAGE (20%, 50 V/Cm, 3 h) and visualized by autoradiography. Incorporation of 7-mGpppA clearly alters the electrophoretic mobility, permitting the identification of the capped transcription products. The positions of the first and second cytosine in the transcripts initiated from p_{28a}, p_{26}, and p_{25} are 33 and 40, 15 and 20, 30 and 36, respectively. M designates markers (*Hpa*II digest of pBR322) whose sizes in nucleotides are given on the left.

an adenosine. (iv) Perfect homology is found at five positions (−7, −11, −12, −34, and −43; Fig. 6). This number would increase to 8 (−33, −35, and −44) if one would allow a 17-bp distance between the −10 and −33 region of p_{H22}.

A closer look at these promoter sequences reveals additional interesting features. Promoters of the preearly and early expression class contain perfect homologies around +7 (preearly, ATGAGAG; early, TTGA) which is followed downstream by a block of purines (usually 10 of 12 bases). This, together with the homology near the RNA start position (T C/T ATA), suggests that in contrast to the common definition of a promoter sequence, regions around and downstream of +1 may be relevant for promoter function. Experiments to examine this hypothesis are currently in progress. Another striking feature of the sequences shown in Fig. 6 is the AT-rich block around −43. A clear selection against GC base pairs in this region is also seen in other promoters, including those of coliphage T7, the rRNA operons, and *lpp* promoter (for a review, see reference 10), all of which are efficient RNA initiation signals in vivo. Thus, as pointed out previously (2), regions outside of the classical promoter sequence which spans from +1 to −35 are clearly under selective pressure, a conclusion which becomes more obvious when highly efficient promoters are analyzed. It should also be noted that the homologies around −43, −33, and +7 are spaced in multiples of 10 bp; thus, in a DNA with a B configuration they can be recognized from one side. Finally, it is interesting that the so-called canonical sequences for the −10 and the −33 region, TATAAT and TTGACA, respectively, are never found simultaneously within the sequence of a strong promoter. As seen in Fig. 6, six of seven promoters with the TATAAT hexamer at −10 contain the sequence TTGCT and one contains the sequence TTGATA at −33; conversely, none of the promoters carrying the so-called canonical −35 hexamer TTGACA also contains the ideal −10 region. This holds also for other strong promoters such as p_L from phage lambda; promoters A_1, A_2, and A_3 from phage T7; and all the promoters of rRNA operons (for a review, see reference 10).

Although the promoters described in this report belong to the strongest transcriptional initiation signals of the *E. coli* system, they still differ in their in vivo efficiency two- to threefold (U. Deuschle, personal communication). Some of the sequences are recognized in vitro by *E. coli* RNA polymerase with the highest forward rate constant measured so far (M. Brunner, M.S. thesis, University of Heidelberg, 1984), whereas others form complexes with the enzyme at distinctly lower rates. Since at least the promoters of the early class can be assumed to have evolved together, possibly from a common ancestor, a detailed study of the structural and functional divergence of these closely related signals may allow a correlation of subsequences of a promoter with specific steps in promoter function.

As seen in Fig. 3, 7, and 8, some of the promoters are suitable for the synthesis of specific RNAs in vitro. The main advantages of using these signals are as follows: (i) By efficiently competing for the enzyme, more than 95% of in vitro RNA is T5 promoter specified with *E. coli* RNA polymerase and plasmid DNA isolated by the method Birnboim and Doly (1). (ii) The promoters readily reinitiate RNA synthesis at high ionic strength, and thus, in a time course of 20 to 30 min, 12 to 25 pmol of RNA per pmol of template can be synthesized, depending on the size of the transcript. (iii) Some of the promoters accept 7-mGpppA as the starting dinucleotide allowing the efficient in vitro synthesis of capped RNA; this RNA can directly be used in translation and translation-translocation experiments (11, 23). A comparison of this system with the widely known SP6 promoter-polymerase combination has shown that the T5 promoters appear to be superior for the production of capped and uncapped in vitro RNA. Whereas the specificity of both systems is comparable, the SP6 polymerase terminates transcription at sequences not recognized by *E. coli* RNA polymerase. Furthermore, the yields of RNA per template on a molar basis is severalfold better with T5 promoters, and, most importantly, capping is simple and most efficient.

A puzzling aspect remains if one considers the role of these promoters in the development of bacteriophage T5. Upon infection, this phage controls the expression of its genes at the transcriptional level in a well-defined temporal program. Distinct RNA species are synthesized at preearly, early, and late times. However, promoters of the different expression classes are active when integrated into plasmids and transformed into host cells. While the differential expression of the preearly and early class genes is controlled

by a two-step injection process of the DNA (13) it has so far been assumed that switching to late expression would require a modification of either the *E. coli* RNA polymerase (16) or the phage template. The properties of the late promoters described here (and additional evidence not included in this report) suggest that their activity at early times is suppressed either by negative control which would at least be partially specified by preearly genes or by physically withholding the late portion of the template from the transcriptional machinery until the onset of DNA replication.

ACKNOWLEDGMENTS

We thank D. Stüber for the various pDS1 constructs and Heidi Taatjes and John Hughes for excellent technical assistance. We also thank Ursula Peschke for help in localizing promoters in the T5 genome and Yvonne Kohlbrenner for help in preparing the manuscript.

This work was supported by grants Bu 338/12-15 from the Deutsche Forschungsgemeinschaft and by the Fonds der Chemischen Industrie Deutschlands.

LITERATURE CITED

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
2. **Bujard, H.** 1980. The interaction of *E. coli* RNA polymerase with promoters. TIBS **5**:274–278.
3. **Clewell, D. B., and R. Helinski.** 1969. Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. Proc. Natl. Acad. Sci. USA **62**:1159–1166.
4. **Denhardt, D. T.** 1966. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. **12**:641–646.
5. **Dobrynin, V. N., V. G. Korobko, J. V. Severtsova, N. S. Bystrov, S. A. Chuvpilo, and M. N. Kolosov.** 1980. Synthesis of a model promoter for gene expression in *Escherichia coli*. Nucleic Acids Res. Symp. Ser. **7**:365–376.
6. **Gabain v., A., and H. Bujard.** 1977. Interaction of *E. coli* RNA polymerase with promoters of coliphage T5. Molec. Gen. Genet. **157**:301–311.
7. **Gabain v., A., and H. Bujard.** 1979. Interaction of *Escherichia coli* RNA polymerase with promoters of several coliphage and plasmid DNAs. Proc. Natl. Acad. Sci. USA **76**:189–193.
8. **Gabain v., A., G. S. Hayward, and H. Bujard.** 1976. Physical mapping of the HindIII, EcoRI, SalI and SmaI restriction endonuclease cleavage fragments from bacteriophage T5 DNA. Molec. Gen. Genet. **143**:279–290.
9. **Gentz, R., A. Langner, A. C. Y. Chang, S. N. Cohen, and H. Bujard.** 1981. Cloning and analysis of strong promoters is made possible by the downstream placement of a RNA termination signal. Proc. Natl. Acad. Sci. USA **78**:4936–4940.
10. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
11. **Hurt, E. C., B. Pesold-Hurt, and G. Schatz.** 1984. The amino-terminal region of an imported mitochondrial precursor polypeptide can direct cytoplasmid dihydrofolate reductase into the mitochondrial matrix. EMBO J. **3**:3149–3156.
12. **Itoh, T., and J. Tomizawa.** 1980. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proc. Natl. Acad. Sci. USA **77**:2450–2454.
13. **Lanni, Y. T.** 1968. First-step-transfer deoxyribonucleic acid of bacteriophage T5. Bacteriol. Rev. **32**:227–242.
14. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
15. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
16. **McCorquodale, D. J.** 1975. The T-odd bacteriophages. Critical Rev. Microbiol. **4**:101–159.
17. **Morrison, D. A.** 1979. Transformation and preservation of competent bacterial cells by freezing. Methods Enzymol. **68**:326–331.
18. **Radloff, R., W. Bauer, and J. Vinograd.** 1967. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in Hela cells. Proc. Natl. Acad. Sci. USA **57**:1514–1521.
19. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
20. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
21. **Stüber, D., and H. Bujard.** 1982. Transcription from efficient promoters can interface with plasmid replication and diminish expression of plasmid specified genes. EMBO J. **1**:1399–1404.
22. **Stüber, D., H. Delius, and H. Bujard.** 1978. Electron microscopic mapping of promoters: localisation of promoters within the bacteriophage T5 genome. Molec. Gen. Genet. **166**:141–149.
23. **Stüber, D., I. Ibrahimi, D. Cutler, B. Dobberstein, and H. Bujard.** 1984. A novel *in vitro* transcription-translation system: accurate and efficient synthesis of single proteins from cloned DNA sequences. EMBO J. **3**:3143–3148.