JOURNAL OF VIROLOGY, Feb. 1981, p. 673-682
0022-538X/81/020673-10$02.00/0

Vol. 37, No. 2

Isolation and Mapping of Small Cauliflower Mosaic Virus DNA Fragments Active as Promoters in *Escherichia coli*

T. D. McKNIGHT AND R. B. MEAGHER*

Department of Microbiology and Department of Molecular and Population Genetics, University of Georgia, Athens, Georgia 30602

Small *Eco*RI* fragments of cauliflower mosiac virus DNA (strain CM4-184), which act as promoters for the tetracycline resistance gene on the promoter probe plasmid pBRH4 in *Escherichia coli*, have been isolated and mapped on the viral genome. Two regions of the viral genome contain DNA sequences with promoter activity in *E. coli*. Two independent cloned fragments from one region direct a high level of tetracycline resistance (up to 38 μg of tetracycline per ml). Two independent fragments from the second region of the viral genome also direct tetracycline resistance, but at lower levels. The activity of the two fragments with the strongest promoter activity in *E. coli* may direct transcription of the viral genome in a clockwise direction. This is consistent with the direction of transcription predicted from sequence analysis of the viral DNA (Franck et al., Cell **21**: 285-294, 1980). One of these fragments maps at the start of a large open translational reading frame which is predicted to contain the coding sequence for the viral coat protein. Each promoter-active fragment is located in the 5'-terminal portion of one of the six open reading frames predicted from the DNA sequence.

Bacterial promoters have been used to control the expression of a number of foreign genes molecularly cloned in *Escherichia coli*: for example, the synthetic genes for somatostatin (21), for insulin (16), for complementary DNA for human leukocytic interferon (34), for human growth hormone (H. L. Heyneker, personal communication), and for native DNA for hepatitis B virus antigens (9). The advantages of promoter manipulation to genetic engineering technology and its potential application to plant molecular cloning systems have been briefly reviewed (30, 31). Our research is directed at the isolation and characterization of highly active plant promoter sequences which can be used to control foreign DNA cloned in plant cells.

A plant DNA virus should be a good source of strong plant promoters. Cauliflower mosaic virus (CaMV) DNA, strain CaMV(B), has a molecular weight of 4.9×10^6 (32). If the translation of CaMV occurs in only one frame, it can code for 260,000 daltons of protein. When fragments of the virus are cloned in plasmid vehicles, three polypeptides with molecular weights of 42,000 (42K), 40K, and 37K are expressed in *E. coli* at high levels (33). Expression of these polypeptides was independent of the insertion orientation of the fragments. In other words, the information for the initiation and termination of these polypeptides is contained entirely within the viral DNA fragments. Therefore, there must be sequences in the virus which function as transcriptional promoters in *E. coli*. It is not known whether similar initiation sites are used to synthesize viral polypeptides in plant cells. Furthermore, the relationship of procaryotic promoters to sequences which function as promoters in eucaryotes is unknown.

The structure and function of *E. coli* ribosomal and structural gene promoters have been studied in detail (37, 39, 42, 50). The majority of information about procaryotic promoter structure comes from studying the DNA sequences bound and protected from nuclease digestion by *E. coli* holo-RNA polymerase. This work has been aided by the fact that *E. coli* holo-RNA polymerase binds true promoter DNA sequences several orders of magnitude more strongly than transcribed regions. In recent studies on the tetracycline resistance promoter of pBR322, Rodriguez et al. (39) have been able to separate the binding and initiation functions of a promoter. It is clear from this work that the "Pribnow box" sequence is sufficient for the binding of *E. coli* RNA polymerase but will not direct the initiation of transcription. A second sequence found in most bacterial promoters, roughly 25 base pairs from the Pribnow box and upstream relative to the direction of transcription, is required for initiation (39).

Very little is known about the structure of eucaryotic promoters. This is in part due to the small differences between the binding of eucaryotic RNA polymerase to specific, correct pro-

moter sequences and the binding to nonspecific DNA sequences. Without strong binding, a major advantage in procaryotic studies, these eucaryotic promoter sequences have proved difficult to identify and isolate. The specific sequences protected by eucaryotic RNA polymerase have not been identified.

Most of the data concerning the location and structure of eucaryotic RNA polymerase promoters is the result of locating the 5′ termini of in vivo eucaryotic transcripts on the purified DNA substrates (for example, see references 4, 5, and 12). When the sequences of these eucaryotic promoter regions are compared with the consensus sequence for procaryotic promoters (41), striking similarities are commonly seen. The majority of eucaryotic transcription units contain a short guanine-cytosine-rich sequence, followed by a substantial adenine-thymine-rich sequence which resembles TATAAATA. These sequences are usually found 30 base pairs upstream from the initiation of transcription (3, 41). This eucaryotic sequence resembles the Pribnow box found in procaryotic promoters, which is of the form 5′-TATAATG. The procaryotic sequence is usually found 10 base pairs upstream from the start of transcription (40, 41).

A few exceptions to the common eucaryotic structure have been reported (3). These exceptions may represent a second eucaryotic promoter structure. A eucaryotic sequence comparable to the procaryotic -35 initiation sequence has not been identified.

Indirect evidence from molecular cloning experiments suggests that some eucaryotic sequences may be functional as promoters in procaryotic cells. For example, several lower eucaryotic DNA sequences direct the synthesis of functional polypeptide products when cloned in *E. coli* (1, 2, 10, 33, 38, 47). Furthermore, it has been shown that the information for the initiation of transcription and translation of most of these polypeptides in *E. coli* comes entirely from within the cloned eucaryotic fragment. The reverse is also true; a procaryotic DNA fragment has been found which synthesizes a functionally correct polypeptide in yeasts (11).

In vitro eucaryotic RNA polymerase II and procaryotic RNA polymerase bind related sequences on polyoma virus DNA (26, 27). However, as was pointed out above, binding is in no way sufficient for initiation. These authors did show that two of the five binding sites identified also could direct the initiation of transcription in vitro for eucaryotic and procaryotic polymerases. In a recent study by Hale et al. (18), it was shown that *E. coli* holo-RNA polymerase binds supercoiled simian virus 40 DNA upstream from the AUG translational start codon for all but one of several simian virus 40 structural genes. With one exception, the polymerase binding sites occur before open reading frames with potential initiation codons. These studies are further evidence that part of the procaryotic promoter structure which interacts with RNA polymerase may also be found in the 5′ regions of eucaryotic structural genes.

In summary, it is suggested from (i) the comparisons of eucaryotic and procaryotic sequences, (ii) the expression of some cloned eucaryotic DNAs in *E. coli*, and (iii) the binding of polymerase and initiation of transcription for two mammalian viruses that homology of structure and function may exist between procaryotic promoters and the 5′ regions of eucaryotic genes. Therefore, it should be possible to use some *E. coli* promoter assays in the examination of eucaryotic promoter sequences.

To speed the isolation of DNA fragments which are active as promoters in *E. coli* (i.e., sequences which both bind RNA polymerase and direct initiation of transcription), promoter probe plasmids have been constructed (35, 39, 49). These plasmids are tetracycline-sensitive derivatives of pBR322. The sequence of the promoter required for the initiation of transcription of a tetracycline resistance gene has been deleted and replaced by a synthetic *Eco*RI site. Tetracycline resistance can be regained when a fragment containing promoter activity is inserted in the correct polarity into this site. These vehicles have been used to demonstrate that eucaryotic DNA fragments can have some promoter activity in *E. coli* (35).

A physical map of the DNA from the wild-type CaMV(B) strain has been published (32). A more complete physical map of CaMV(CM4-184) has been determined (unpublished data). This strain is deficient in aphid transmission and therefore can be manipulated in recombinant plasmids in *E. coli* at the P1-EK1 level. It contains a small deletion relative to CaMV(B) and has a molecular weight of 4.8×10^6. In this manuscript, we describe the isolation, characterization, and mapping of small promoter-active DNA fragments isolated from CaMV(CM4-184) DNA.

MATERIALS AND METHODS

Chemicals and media. Cultures were grown on Luria agar plates or in Luria broth (23). Polycillin N (Bristol Laboratories, Syracuse, N.Y.) was used at 20 μg/ml in Luria plates to select for ampicillin resistance. Tetracycline hydrochloride (Sigma Chemical Co., St. Louis, Mo.) was added to Luria agar at concentrations from 5 to 50 μg/ml to select for tetracycline resistance. All other chemicals were standard reagent grade.

CaMV(CM4-184) is derived from wild-type CaMV(B) strain (29), is deficient in aphid transmission

properties, and contains a small deletion and several altered restriction cleavage sites relative to wild type (20; unpublished data).

Plasmids and strains. Plasmid pIM4 consists of the entire genome of CaMV(CM4-184) joined to pBR322 (7) through the single *SalI* site. This plasmid was the source of all viral DNA used in this work. The *SalI* linear form of the cloned viral DNA was prepared in milligram quantities by digesting pIM4 with *SalI*, separating the viral DNA from the vehicle by preparative agarose electrophoresis, and banding in cesium chloride-propidium iodide gradients (9a, 32).

The vehicle pBRH4 is a tetracycline-sensitive derivative of pBR322 which has had the promoter for the tetracycline resistance gene deleted and replaced by a synthetic *Eco*RI linker. This vehicle was constructed and supplied by H. L. Heyneker (Fig. 1). A description of this plasmid was published by West et al. (49).

E. coli SK1592 is a phage T1-resistant derivative of SK1590 and was used for the initial transformations because of its high transformation frequency and long-term stability on drug plates (24). Plasmids showing tetracycline resistance in SK1592 were transformed into *E. coli* strain HB101 (8), where the tetracycline resistance levels were determined, and large amounts of amplified DNA can be produced (32).

Enzymes and buffers. T4 DNA ligase was a gift from H. L. Heyneker. *SalI* was prepared as described by Greene et al. (17). *Eco*RI was a gift from Patricia J. Greene. *Eco*RI to be used for *Eco*RI* digests was dialyzed overnight against 10 mM Tris-hydrochloride (pH 8.0)–50 mM NaCl–0.1 mM EDTA–50% glycerol (36). This dialysis raised the pH of the original storage buffer from 7.5 and lowered the NaCl concentration from 200 mM. Other restriction endonucleases were prepared within the Genetics Department at the University of Georgia.

Restriction enzyme digests were carried out in the buffers used by Bolivar et al. (6) and Greene et al. (17). Ligations were done in the buffer recommended by Dugaiczyk et al. (13). *Eco*RI* digests were performed, using 3 to 5 U of dialyzed enzyme per μg of plasmid DNA in 30 μl of 20 mM Tris-hydrochloride (pH 8.8)–2 mM $MgCl_2$. After 16 h at 37°C, the reactions were about 95% complete.

Electrophoresis. Agarose and polyacrylamide gels were run in the Tris-borate system of Greene et al. (17). The ratio of acrylamide to bisacrylamide was 30:1. *Hin*dIII digests of phage λ DNA gave a series of useful molecular weight standards ranging from 15.8 × 10^6 to 1.31 × 10^6. *Alu*I digests of pBR322 gave a series of molecular weight standards with a useful range from 0.61 × 10^6 to 0.042 × 10^6 (44).

Construction and selection of promoter-active plasmids. A total of 5 μg of pBRH4 was cleaved with *Eco*RI and ligated to 2 μg of *Eco*RI* fragments from the cloned *SalI* linear viral DNA. The ligation was carried out in 70 μl of ligase buffer (13) with 0.5 U of T4 ligase at 12°C overnight. After the ligated DNA was precipitated with 3 volumes of ethanol, it was transformed into *E. coli* SK1592 by the $CaCl_2$–RbCl method (24). Transformants were selected by plating onto Luria agar plates containing 5 μg of tetracycline per ml 1 h after the transformation. All of the resultant colonies were then patched onto plates with 5, 10, 15, or 20 μg of tetracycline per ml to further select for tetracycline-resistant recombinants. Recombinants resistant to 10 μg of tetracycline per ml or higher concentrations of the drug were analyzed further.

Determination of the levels of tetracycline resistance. Overnight cultures of HB101 containing re-

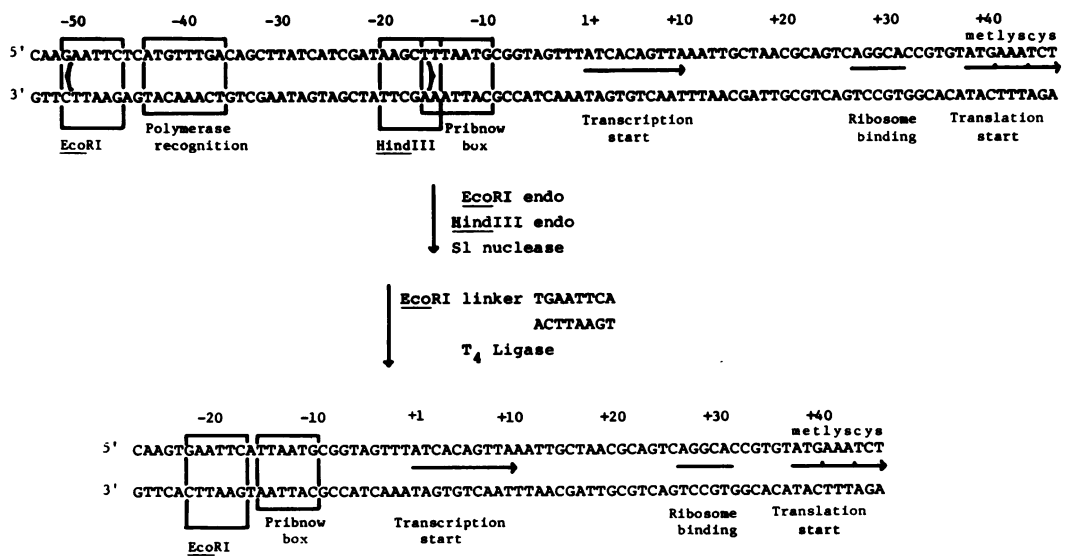

FIG. 1. *Sequence, in vitro construction, and proposed molecular functions of the promoter cloning vehicle pBRH4. A small EcoRI-HindIII endonuclease fragment containing the tetracycline resistance gene promoter was removed from pBR322 as shown. In place of this deletion there now exists a synthetic EcoRI site in pBRH4, the promoter cloning plasmid. This plasmid was supplied by H. L. Heyeneker.*

combinant plasmids were inoculated into 25 ml of Luria broth. When cultures reached 100 Klett units (Klett-Summerson colorimeter, no. 54 green filter), they were diluted 10^6-fold in 10 mM potassium phosphate buffer (pH 7.2). From the final dilution, 0.4 ml was spread onto Luria agar plates containing 0, 5, 10, 15, 20, 30, and 50 μg of tetracycline per ml, respectively. Colonies arising on the plates were counted at 24 and 48 h after plating. Results at 48 h were determined to be more reproducible than those at 24 h. Therefore, all numbers reported are those determined at 48 h. Each plating was prepared in duplicate. The two sets of plates usually yielded counts within 10% of each other. Platings which gave greater errors were discarded and repeated. In determining the efficiency of plating on various concentrations of tetracycline, the number of colonies on the control plates lacking drugs was taken as 100% efficiency of plating (EOP_{100}). The level of tetracycline resistance of various recombinants is reported as the concentration of tetracycline which allowed 50% of the number of colonies contained on the control to develop (EOP_{50}) (45, 46).

Hybridization mapping of promoter-active fragments. The *SalI* linear form of the cloned viral genome was digested with various restriction enzymes and electrophoresed on 1.5% agarose gels. Transfers of these gels to nitrocellulose filter paper (Schleicher & Schuell Co., Keene, N.H.; BA85) were performed as described by Southern (43). These filters were then probed with promoter-active plasmids labeled with ^{32}P by nick translation (29). Autoradiograms were made, using Du Pont Cronex 2DC film with exposure times of 2 to 4 h. The smallest region of the genome which could account for hybridization to all of the restriction fragments showing up on the autoradiogram was the position assigned to the promoter-active fragment.

RESULTS

When the *Eco*RI* fragments of CaMV(CM4-184) DNA were ligated into the *Eco*RI site of pBRH4 and transformed into *E. coli*, strains with a wide range of tetracycline resistance levels were found. Tetracycline resistance levels of 74 transformants which could grow on Luria agar with 5 μg of tetracycline per ml were measured semiquantitatively by streaking them onto a series of plates with tetracycline concentrations ranging from 5 to 20 μg/ml.

Twenty recombinants showing good growth on 10 μg of tetracycline per ml were investigated further in *E. coli* strain HB101. More accurate determinations of their tetracycline resistance levels were made by determining the EOP_{50} on a wider range of tetracycline concentrations. The EOP_{50} for these plasmids fell into four groups, with values clustered around 13, 16, 24, and 38 μg of tetracycline per ml. In Fig. 2, a graph of the plating efficiencies of representative plasmids from these groups is shown, along with the pBRH4 vehicle and its parent plasmid pBR322.

To determine the size and number of viral *Eco*RI* fragments responsible for the newly acquired tetracycline resistance levels, the plasmids were purified and digested with *Eco*RI*. For several plasmids resistant to the highest levels of tetracycline ($EOP_{50} > 30$, pMM128; Fig. 2), there were no viral bands visible on 1.8% agarose gels. To examine these recombinants for the exact sizes of their inserted DNA fragments and not to miss very small fragments, two other methods were used.

*Hinc*II cleaves on either side of the *Eco*RI insertion site in pBRH4 and gives rise to fragments with molecular weights of 2.17×10^6 and 0.73×10^6. The smaller fragment contains the *Eco*RI site. Alterations in molecular weight of this band were used to measure the size of inserted DNA (Table 1). For example, by this technique the promoter plasmids with the highest activity which show no new bands (Fig. 2) were shown to have the same size insert, about 280 base pairs. When the recombinant plasmids are digested with *Eco*RI* and separated by elec-

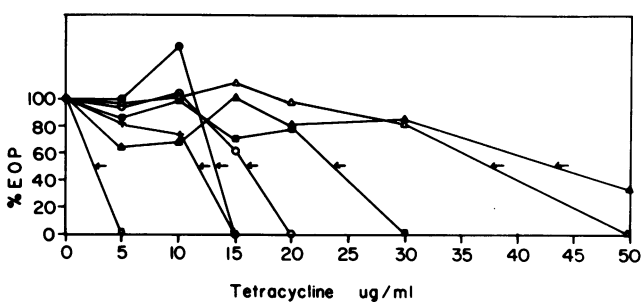

FIG. 2. *Determination of the EOP of strains containing a tetracycline resistance gene controlled by CaMV DNA fragments. The number of colonies on Luria control plates represents EOP_{100}. The number of colonies on the various concentrations of tetracycline are given relative to this number from Luria plates. Symbols represent strains containing the different CaMV inserts controlling tetracycline resistance as follows: □, pBRH4, tetracycline-sensitive plasmid vehicle; ▲, pBR322, tetracycline-resistant control and parent of pBRH4. Symbols for plasmids containing the CaMV(CM4-184) inserts: ●, pMM12; +, pMM40; ○, pMM116; ■, pMM140; △, pMM128.*

TABLE 1. *Summary of promoter-active plasmids containing CaMV DNA*

Plasmid	EOP_{50}[a]	CaMV genome position of inserted promoter (kb)	Size of insert (base pairs)[b]	No. of independent isolations of active fragment[c]
pBRH4	2	—	—	—
pMM40	11	4.45–4.70	120	3
pMM12	13	4.70–5.25	370	4
pMM116	17	3.52–3.83	610	1
pMM140	23	3.15–3.52	420	1
pMM128	38	3.52–3.83	280	7
pBR322	43	—	—	—

[a] EOP values are expressed as micrograms of tetracycline per milliliter.

[b] These sizes are calculated from the mobilities of the 322-*Hae*III and 322-*Alu* fragments on the 7.5% acrylamide gel in Fig. 3.

[c] For example, the fragment contained in pMM128 was found in six other independently isolated tetracycline-resistant transformants.

trophoresis on 7.5% polyacrylamide gels, the extra DNA inserts can be detected (Fig. 3). The most active fragment contained in pMM128, which usually ran as a doublet on agarose, migrated just above the lowest *Eco*RI* band of the vehicle at approximately 280 base pairs (Fig. 3, track h). This recombinant and other most highly resistant plasmids contained a 280-base-pair insert, and all contained one of the two viral *Bam*HI sites, confirming that the inserted fragments are identical in all isolates. In all seven isolates of this fragment, the viral *Bam*HI site is the same distance from the *Bam*HI site in the tetracycline gene (data not shown), indicating that only one orientation of the inserted fragment is found.

Also present on this promoter-active fragment is a bona fide *Eco*RI site at one end. The distance between this site and the *Bam*HI site of the insert is 0.165×10^6 daltons, the distance between a *Bam*HI site and an *Eco*RI site on the native viral DNA reported by Meagher et al. (32) for the CaMV(B) strain and by Volovitch et al. (48) for the PV147 strain. For the CM4-184 strain we have also demonstrated that these sites are in the same position, suggesting that the viral end of the *Eco*RI site in pMM128 is half of a genuine viral *Eco*RI site and not an artifact of cloning the *Eco*RI* fragments. Therefore, these restriction sites on pMM128 allow accurate positioning of the promoter fragment on the viral genome. The *Eco*RI site defines one end, and digestion of the plasmid with *Eco*RI* followed by *Bam*HI shows that the promoter fragment ends less than 20 base pairs counterclockwise from the *Bam*HI site (Fig. 4). The positions of the *Eco*RI site, the viral *Bam*HI site, and the *Bam*HI site of the vehicle give the relation between the orientation of the fragment in the plasmid and its orientation on the viral genome. Since the distance between *Bam*HI sites is greater than the distance between the *Eco*RI site and the *Bam*HI site of the vehicle in the tetracycline gene, transcription on the fragment in *E. coli* occurs in the direction shown by the arrow in Fig. 4 relative to a map of the viral genome.

The other promoter-active fragments were mapped on the viral genome by Southern hybridizations. The *Sal*I linear form of the viral DNA purified from recombinant plasmids was digested with various restriction enzymes. These fragments were separated by agarose gel electrophoresis and transferred to nitrocellulose. The filters were probed with nick-translated, promoter-active plasmids. Control nick-translated pBRH4 does not hybridize to any fragments of

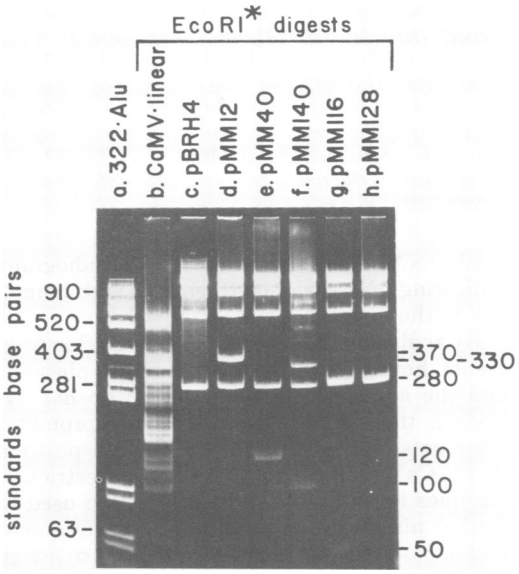

FIG. 3. *Electrophoretic separation of the EcoRI* endonuclease fragments of the five promoter-active clones on 7.5% acrylamide. The following samples were digested with EcoRI* endonuclease and run in the lettered tracks: (b) purified SalGI linear form of CM4-184 strain of CaMV; (c) pBRH4; (d) pMM12; (e) pMM40; (f) pMM140; (g) pMM116; (h) pMM128. Track (a) contains molecular weight standards, pBR322 plasmid digested with AluI. The HaeIII digest of pBR322 also was a standard in this gel (data not shown). Molecular sizes of the pertinent DNA fragments are indicated on the sides of the figure in base pairs. The fluorescent bands have been photographed, using a yellow filter, after transillumination with short-wave UV light.*

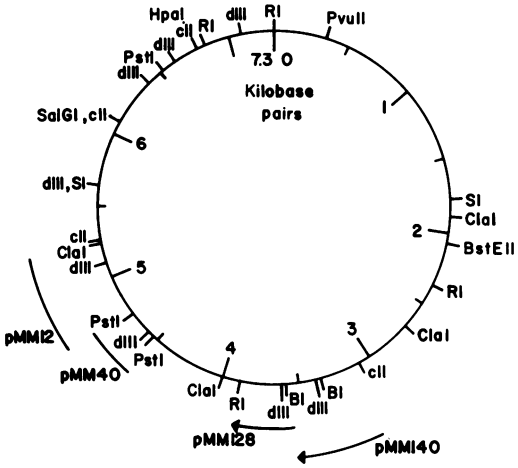

FIG. 4. *Location of promoter-active fragments in the CaMV(CM4-184) genome. Fragments contained in pMM128 and pMM140 allow the highest level of tetracycline resistance, whereas those in pMM12 and pMM40 allow only moderate tetracycline resistance. The positions of the promoter fragments are shown relative to the following endonuclease cleavage sites: EcoRI (RI); HincII (cII); S1 nuclease (S1); ClaI; HindIII (dIII); BstI or BamHI (BI); PstI; SalGI; PvuII; BstEII; and HpaI. The direction of transcription from the fragments in pMM128 and pMM140 is indicated by the arrow. The direction of transcription was not determined for pMM40 and pMM12. The EcoRI site at the top of this map was defined as the origin (32).*

viral DNA. Figure 5 shows an autoradiogram, indicating the restriction fragments with homology to the insert in pMM12.

By analyzing the overlapping restriction fragments homologous to the promoter plasmids and the adjacent fragments which do not hybridize, the general location of the promoter fragments can be determined. Such a position on the CaMV map also allowed other restriction enzymes to be chosen, which could be used to further pin down the promoter-active fragment locations. These techniques were used to assign the locations of the promoter fragments shown in Fig. 4. The sizes of the promoter regions, determined by hybridizations and restriction enzyme analysis, are not much larger than the promoter fragments themselves (Fig. 3; Table 1).

The promoter-active plasmids pMM116 and pMM140 hybridize to the same region of the genome as pMM128. *Eco*RI* digestion of pMM116 shows that it contains two copies of the promoter fragment found in pMM128 and a third smaller viral fragment of 50 base pairs (Fig. 3, track g). The positions of the *Bam*HI sites on this plasmid and the lack of the *Eco*RI site found in pMM128 can be used to determine the positions and orientations of these three fragments. Plasmid pMM116 contains two head-to-head copies of this fragment plus a small 50-base-pair fragment between them. pMM116 is resistant to

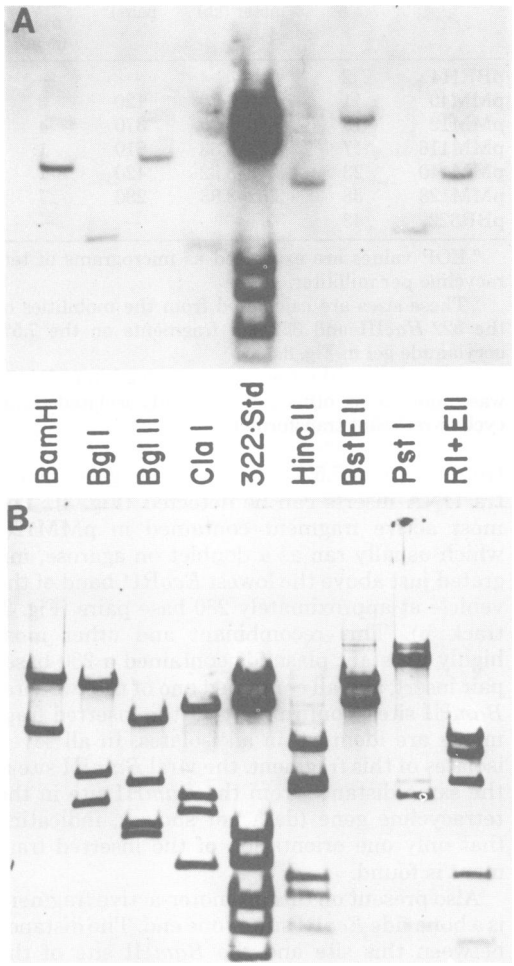

FIG. 5. *Southern hybridization of CaMV DNA fragments to nick-translated promoter-active plasmid pMM12. (B) Native 1.0% agarose gel stained with ethidium bromide; (A) autoradiograph of the Southern hybridization of this gel. Gel tracks contain digestions of the SalI linear form of CM4-184 with the following restriction endonucleases from left to right: BamHI, BglI, BglII, ClaI, standards, HincII, BstEII, PstI, EcoRI + BstEII. The central track contains the following combined molecular weight standards: the linear form of pBR322, 2.9 × 10⁶; the HincII fragments of pBR322, 2.17 × 10⁶ and 0.74 × 10⁶; and the AluI fragments of pBR322, 0.60 × 10⁶, two at 0.44 × 10⁶, 0.35 × 10⁶, and 0.27 × 10⁶. Promoter plasmid pMM12 was nick translated with [³²P]ATP to a specific activity of 2 × 10⁷ cpm/μg, and 2 × 10⁶ cpm was used to hybridize with the filter.*

less than half the amount of tetracycline (Fig. 2) as pMM128. Presumably, initiation in pMM116 occurs on the 280-base-pair fragment in the same orientation as in pMM128, but transcription must proceed through an extra 330 base pairs of DNA.

pMM140 also hybridizes to the same general region of the genome as pMM116 and pMM128 (Fig. 4), but contains two *Eco*RI* fragments of 330 and 100 base pairs not found in any other promoter-active plasmid (Fig. 3, track f). An *Eco*RI* fragment corresponding to the one in pMM128 is not found in pMM140. There is a *Bam*HI site present in the largest of these fragments. However, the viral *Bam*HI site in pMM128 and pMM116 is on an *Eco*RI* fragment with a different size. The extra *Bam*HI site in pMM140 must be the other viral *Bam*HI site. The distance between the *Bam*HI site on pBRH4 and the *Bam*HI site in the cloned viral DNA is approximately 20 base pairs greater then the distance from the *Hin*dIII site to the *Bam*HI site of the vehicle (data not shown). These data indicate that the transcription from the largest viral *Eco*RI* fragment in pMM140 is clockwise (Fig. 4).

pMM40 contains a 120-base-pair *Eco*RI* fragment (Fig. 3, track e) and has an EOP_{50} of 12 μg of tetracycline per ml (Fig. 2). The position assigned to it by Southern hybridization as shown in Fig. 4 is confirmed by the presence of both a *Pst*I site and a *Hin*dIII site on the inserted fragment (data not shown). The pMM40 promoter fragment was found in two other plasmids, pMM39 and pMM139. pMM39 is identical to pMM40. pMM139 contains three other viral *Eco*RI* fragments in addition to the one in pMM40 and pMM39. pMM139 was not characterized further.

Figure 4 shows the position of the weak promoter-active fragment in pMM12 determined from Southern hybridization. This 370-base-pair *Eco*RI* fragment was found in *Eco*RI* digests of four independent isolates. *Hin*cII cleavage patterns of three other weak promoter plasmids were identical, suggesting that they may also have this promoter fragment.

Table 1 summarizes the size, location, and EOP_{50} of the promoter-active fragments cloned from CaMV. The promoter-active fragments may be divided into two classes: high-level promoter activity fragments which map between 3.15 and 3.85 kilobases (kb) and low-level activity fragments which map between 4.50 and 5.25 kb.

DISCUSSION

Small DNA fragments from CaMV which function as promoters for the tetracycline resistance gene of pBRH4 were identified. Based on the sizes of the fragment inserts, their EOP at various tetracycline concentrations, and their location on the CaMV genome, the clones were placed into two categories.

The two clones resistant to the lowest level of tetracycline, pMM40 and pMM12, contained different *Eco*RI* inserts (from approximately 4.45 to 4.80 and 4.8 to 5.3 kb on the CaMV map, respectively). These two inserts lie adjacent to each other in the genome. In the Southern blot hybridization used to map these fragments, pMM40 showed a low level of homology, with a short segment of the viral genome at roughly 2.8 kb. We have not been able to show that this is due to the presence of any contaminating second fragment in this recombinant. It may represent true homology between two segments of viral genome.

Clones with the highest levels of tetracycline resistance map between 3.1 and 3.8 kb. Recombinants pMM116 and pMM128 share a common *Eco*RI* fragment of 280 base pairs which maps at 3.52 to 3.83 kb. The plasmid most active in promoter activity, pMM128, contains this fragment alone, whereas the other clone contains additional DNA fragments. Another highly active fragment is contained in pMM140 and maps adjacent to but not overlapping the fragment in pMM128.

When a wild-type CaMV(B) DNA *Eco*RI fragment which maps from 0 to 3.9 kb was cloned in *E. coli* minicells, it directed the synthesis of a 37K polypeptide at high levels (see above and reference 33). It is clear from the data presented that promoter activity for the polypeptide expression observed in minicells came from within this DNA fragment. In the CM4-184 strain used in this study, there is an extra *Eco*RI site within this fragment and a small 200-base-pair deletion at about 3.1 kb relative to the wild-type DNA used in minicells. The most active promoter fragment is contained in this general region; however, its direction of transcription is opposite to what would be required for synthesis of any reasonably sized protein from the *Eco*RI fragment in minicells. The fragment of high-level activity contained in pMM140 could not serve as the promoter for the 37K polypeptide observed in minicells because its orientation is incorrect (Fig. 4).

The 1.9×10^6-dalton *E. coli* fragment running from 3.85 to 6.8 kb is contained in both wild-type CaMV(B) and CM4-184 strains of the virus. When the fragment from the wild-type strain was cloned in *E. coli* minicells, it directed the synthesis of high levels of 40K and 42K proteins. It was shown that the information for initiation of transcription and translation of these proteins

came entirely from within the cloned DNA fragment (33). The fragments in pMM12 and pMM40 could contain the promoter(s) for expression of these polypeptides.

The levels of tetracycline resistance conferred by the recombinant plasmids were taken as an indication of the relative activities of the promoters in *E. coli*. Resistance levels could be changed by other means. For example, if an inserted promoter caused transcription to begin upstream from its usual starting site, the new mRNA may have altered stability and allow more or less translation from the same amount of message.

The most active fragment contained in pMM128 was obtained more often than any other. Of the 16 independent clones, 7 contained this fragment (Table 1). The fragments in pMM140, pMM40, and pMM12, although less active, are contained in the genome in the same stoichiometry as the fragment in pMM128. However, these three less active fragments were obtained less frequently. All of these recombinant plasmids direct tetracycline resistance at greater than 10 μg/ml. The initial screening was performed at 5 μg/ml. These data suggest that there is a selection against the less active fragments during the initial molecular cloning and drug selection.

It is also likely that some true viral promoters were not functional in *E. coli*. They might have an altered structure at which *E. coli* polymerase cannot bind and initiate, or these sequences might require a positive acting viral or plant gene product for expression.

The validity of the technique for isolating the true CaMV viral promoters awaits confirmation by other techniques, such as mapping of the viral transcripts on the physical restriction map. Initial northern blot mapping data demonstrates that the fragment contained in pMM128 has a sequence in common with polyadenylated viral specific transcripts (Hagen, Condit, and Meagher, unpublished data).

The direction of transcription determined in *E. coli* for the fragments contained in pMM128 and pMM140 is consistent with the direction of transcription of the native CaMV genome, as suggested by the data of Hull et al. (19). No fragments with promoter activity in *E. coli* were obtained which contained either of the single-strand gap regions of the viral genome, as might be expected from the data of Hull et al. (19).

After this manuscript was completed, the nucleotide sequence of the CaMV cabbage B-S genome was published (14). Six extensive open translational reading frames were found in one strand of the virus. These data suggest that if splicing of transcripts occurs, it is very limited. The promoter-active fragment in pMM128 can be located in this viral sequence, and it extends from nucleotide 2143 to nucleotide 2417. This fragment contains sequences upstream and including the start of translational reading frame IV, which is free of termination codons for 1,500 base pairs in a clockwise direction. Franck et al. (14) suggest that this reading frame codes for the viral coat protein on the basis of codon frequency in the DNA and the amino acid content of the viral coat protein.

Hybridization mapping and restriction enzyme analysis of the largest viral *Eco*RI* fragment in pMM140 indicates that it contains the region just before the start of open translational reading frame III. This reading frame begins at nucleotide 1812 and extends clockwise to nucleotide 2219 (14).

Apparent differences between the restriction maps of the CM4-184 and cabbage B-S strains of CaMV (25) make it difficult to locate precisely the promoter-active fragments of pMM40 and pMM12 in the DNA sequences. The *Hin*dIII site at nucleotide 3251 corresponds to the *Hin*dIII site contained in pMM40, but we could not determine the boundaries of this promoter-active fragment in the sequence. The *Hin*dIII site on the CM4-184 genome found in pMM12 is not present on the cabbage B-S DNA (14). The promoter-active fragment in pMM12 maps adjacent to the pMM40 fragment (Fig. 5). Therefore, both of these fragments are located at or just before the start of the longest open translational reading frame on the cabbage B-S genome. Designated region V, this sequence extends from nucleotide 3591 to nucleotide 5672 (14).

Sequence analysis of the promoter-active fragments described and a comparison with the sequence of the whole viral genome will be very helpful in analyzing and interpreting transcription of the viral genome. A true CaMV promoter fragment should be most useful in constructing genes to be functionally expressed in plant cells.

ACKNOWLEDGMENTS

We thank Gina Eckenrode and Tom Hagen for technical assistance and Daniel Vapnek for his initial reading of the manuscript. This work was supported initially by U.S. Department of Agriculture Science and Education Administration SEA grant 5901-0410-8-0068-0 and, in the main part, by National Science Foundation grant PCM-7900833. T.D.M. was supported in part by Public Health Service Genetics Training grant 1011RH194069 from the National Institutes of Health.

LITERATURE CITED

1. **Alton, N. K., J. A. Hautala, N. H. Giles, S. R. Kushner, and D. Vapnek.** 1978. Transcription and translation in *E. coli* of hybrid plasmids containing the catabolic dehydroquinase gene from *Neurospora crassa*. Gene 4: 241-259.

2. **Bach, M. L., F. Lacroute, and D. Botstein.** 1979. Evidence for transcriptional regulation of orotidine-5'-phosphate decarboxylase in yeast by hybridization of mRNA to the yeast structural gene cloned in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **76**:386–400.
3. **Baker, C. C., J. Herisse, G. Courtois, F. Galibert, and E. Ziff.** 1979. Messenger RNA for the Ad2 DNA binding protein: DNA sequence encoding the first leader and heterogeneity at the mRNA 5' end. Cell **18**:569–580.
4. **Battey, J., and D. A. Clayton.** 1978. The transcription map of mouse mitochondrial DNA. Cell **14**:143–156.
5. **Berk, A. J., and P. M. Sharp.** 1977. Sizing and mapping of early adenovirus in RNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
6. **Bolivar, F., R. L. Rodriguez, M. C. Betlach, and H. W. Boyer.** 1977. Construction and characterization of new cloning vehicles. 1. Ampicillin resistant derivatives of the plasmid pMB9. Gene **2**:75–93.
7. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Crassa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. II. A multiple cloning system. Gene **2**:95–113.
8. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41**:459–472.
9. **Burrell, C. J., P. Mackay, P. J. Greenaway, P. H. Hofschneider, and K. Murry.** 1979. Expression in *Escherichia coli* of hepatitis B virus DNA sequences cloned in plasmid pBR332. Nature (London) **279**:43–47.

9a. **Carreira, L. H., B. C. Carlton, S. M. Bobbio, R. T. Nagao, and R. B. Meagher.** 1980. Construction and application of a modified "gene machine": a circular concentrating preparative gel electrophoresis device employing discontinuous elution. Anal. Biochem. **106**: 455–468.

10. **Clark, L., and J. Carbon.** 1979. Functional expression of cloned yeast DNA in *Escherichia coli*: specific complementation of arginosuccinate lyase (argH) mutations. J. Mol. Biol. **120**:517–532.
11. **Cohen, J. D., T. R. Eccleshall, R. B. Needleman, H. Federoff, B. A. Buchferer, and J. Marmur.** 1980. Functional expression in yeast of the *Escherichia coli* plasmid gene coding for chloramphenicol acetyltransferase. Proc. Natl. Acad. Sci. U.S.A. **77**:1078–1082.
12. **Dhar, R., K. N. Subramanian, J. Par, and S. M. Weissman.** 1977. Nucleotide sequence of a fragment of SV40 DNA that contains the origin of DNA replication and specifies the 5' ends of "early" and "late" viral RNA IV. J. Biol. Chem. **252**:368–376.
13. **Dugaiczyk, A., H. W. Boyer, and H. M. Goodman.** 1975. Ligation of Eco RI endonuclease-generated RNA fragments into linear and circular structures. J. Mol. Biol. **96**:171–184.
14. **Franck, A., H. Guilley, G. Jonard, K. Richards, and L. Hirth.** 1980. Nucleotide sequence of cauliflower mosaic virus DNA. Cell **21**:285–294.
15. **Gannon, F., K. O'Hare, F. Perrin, J. P. Le Pennac, C. Benoist, M. Cochet, R. Breathnach, A. Royal, A. Garapin, B. Cami, and P. Chambon.** 1979. Organization and sequences at the 5' end of a cloned complete ovalbumin gene. Nature (London) **278**:428–434.
16. **Goeddel, D. V., D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A. D. Riggs.** 1979. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Proc. Natl. Acad. Sci. U.S.A. **76**:106–110.
17. **Greene, P. J., H. L. Heyneker, F. Bolivar, R. L. Rodriguez, M. C. Betlach, A. A. Covarrubias, K. Backmann, D. J. Russell, R. Tait, and H. W. Boyer.** 1978. A general method for the purification of restriction enzymes. Nucleic Acids Res. **5**:2373–2380.
18. **Hale, P., R. S. Woodward, and J. Lebowitz.** 1980. *E. coli* RNA polymerase promoters on superhelical SV40 DNA are highly selective targets for chemical modification. Nature (London) **284**:640–644.
19. **Hull, R., S. N. Covey, J. Stanley, and J. W. Davies.** 1979. The polarity of the cauliflower mosiac virus genome. Nucleic Acids Res. **7**:669–677.
20. **Hull, R., and S. H. Howell.** 1978. Structure of the cauliflower mosaic virus genome. II. Variation in DNA structure and sequence between isolates. Virology **86**: 482–493.
21. **Itakura, K., T. Hirose, R. Crea, A. D. Riggs, H. L. Heyneker, F. Bolivar, and H. W. Boyer.** 1977. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science **198**: 1056–1063.
22. **Kozak, M.** 1978. How do eucaryotic ribosomes select initiation regions in messenger RNA? Cell **15**:1109–1123.
23. **Kushner, S. R.** 1974. In vivo studies of temperature-sensitive *recB* and *recC* mutants. J. Bacteriol. **120**: 1213–1218.
24. **Kushner, S. R.** 1978. An improved method for transformation of *Escherichia coli* with Col E1 derived plasmids, p. 17–23. *In* H. W. Boyer and S. Nicosia (ed.), Proceedings of the International Symposium on Genetic Engineering. Elsevier/North-Holland Biomedical Press, Amsterdam.
25. **Lebeurier, G., O. Whitechurch, A. Lesot, and L. Hirth.** 1978. Physical map of DNA from a new cauliflower mosaic virus strain. Gene **4**:213–226.
26. **Lescure, B., A. Chestier, and M. Yaniv.** 1978. Transcription of polyoma virus *in vitro*. II. Transcription of superhelical and linear polyoma DNA by RNA polymerase II. J. Mol. Biol. **124**:73–85.
27. **Lescure, B., C. Dauguet, and M. Yaniv.** 1978. Transcription of polyoma virus DNA in vitro. III. Localization of calf thymus RNA polymerase II binding sites. J. Mol. Biol. **124**:87–96.
28. **Lung, M. C. Y., and T. P. Prione.** 1973. Studies on the reason for differential transmissibility of cauliflower mosaic virus isolates by aphids. Phytopathology **63**: 910–914.
29. **Maniatis, T., A. Jeffrey, and D. G. Kleid.** 1975. Nucleotide sequence of the rightward operator of phage λ. Proc. Natl. Acad. Sci. U.S.A. **72**:1184–1188.
30. **Meagher, R. B.** 1977. The development of a molecular cloning system in higher plants, p. 129–158. *In* A. Hollander (ed.), Genetic engineering for nitrogen fixation. Plenum Publishing Corp., New York.
31. **Meagher, R. B., and T. D. McKnight.** 1980. A model for a molecular cloning system in higher plants: isolation of plant viral promoters, p. 63–75. *In* C. J. Leaver (ed.), Genome organization and expression in plants. Plenum Publishing Corp., New York.
32. **Meagher, R. B., R. J. Shephard, and H. W. Boyer.** 1977. The structure of cauliflower mosaic virus. I. A restriction endonuclease map of cauliflower mosaic virus DNA. Virology **80**:362–375.
33. **Meagher, R. B., R. C. Tait, M. Betlach, and H. W. Boyer.** 1977. Protein expression in *E. coli* minicells by recombinant plasmids. Cell **10**:521–536.
34. **Nagata, S., H. Taira, A. Hall, L. Johnsrud, M. Streuli, J. Ecsodi, W. Boll, K. Cantell, and C. Weissman.** 1980. Synthesis in *E. coli* of a polypeptide with human leukocyte interferon activity. Nature (London) **284**: 316–323.
35. **Neve, R. L., R. W. West, and R. L. Rodriguez.** 1979. Eukaryotic DNA fragments which act as promoters for a plasmid gene. Nature (London) **277**:324–325.
36. **Polisky, B., P. Greene, D. E. Garfin, B. J. McCarthy, H. M. Goodman, and H. W. Boyer.** 1975. Specificity of substrate recognition by the Eco RI restriction en-

donuclease. Proc. Natl. Acad. Sci. U.S.A. **72**:3310–3314.
37. **Pribnow, D.** 1975. Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. J. Mol. Biol. **99**:419–443.
38. **Ratzkin, B., and J. Carbon.** 1977. Functional expression of cloned yeast DNA in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **74**:487–491.
39. **Rodriguez, R. L., R. W. West, and H. L. Heyneker.** 1979. Characterizing wildtype and mutant promoters of the tetracycline resistance gene in pBR313. Nucleic Acids Res. **6**:3267–3288.
40. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
41. **Scherer, G. E. R., M. D. Walkinshaw, and S. Arnott.** 1978. A computer aided oligonucleotide analysis provides a model sequence for RNA polymerase promoter recognition in *E. coli*. Nucleic Acids Res. **5**:3759–3773.
42. **Simpson, R. B.** 1979. The molecular topography of RNA polymerase promoter interaction. Cell **18**:277–285.
43. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
44. **Sutcliffe, J. G.** 1978. pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. Nucleic Acids Res. **5**: 2721–2728.
45. **Tait, R. C., and H. W. Boyer.** 1978. On the nature of tetracycline resistance controlled by the plasmid pSC101. Cell **13**:73–81.
46. **Tait, R. C., R. L. Rodriguez, and H. W. Boyer.** 1977. Altered tetracycline resistance in pSC101 recombinant plasmids. Mol. Gen. Genet. **151**:327–331.
47. **Vapnek, D., J. A. Hautala, J. W. Jacobson, N. H. Giles, and S. R. Kushner.** 1977. Expression in *Escherichia coli* K-12 of the structural gene for catabolic dehydroquinase of *Neurospora crassa*. Proc. Natl. Acad. Sci. U.S.A. **74**:3508–3512.
48. **Volovitch, M., G. Drugeon, J. P. Dumas, A. L. Haenni, and P. Yot.** 1979. A restriction map of cauliflower mosaic virus DNA (strain PV147) mapping of the cleavage sites of Hha I, SacI, AvaI, PvuII, PstI, XbaI, EcoRI, BglII, HindII, Hap II and Hind II & III. Eur. J. Biochem. **100**:245–255.
49. **West, R. W., Jr., R. L. Neve, and R. L. Rodriguez.** 1979. Construction and characterization of *E. coli* promoter-probe plasmid vectors. Gene **7**:271–288.
50. **Young, R. A., and J. A. Steiz.** 1979. Tandem promoters direct *E. coli* ribosomal RNA synthesis. Cell **17**:225–234.