MOLECULAR AND CELLULAR BIOLOGY, May 1985, p. 1170–1183
0270-7306/85/051170-14$02.00/0

Vol. 5, No. 5

DNA Primase-DNA Polymerase α from Simian Cells: Sequence Specificity of Initiation Sites on Simian Virus 40 DNA

MASAMITSU YAMAGUCHI, ERIC A. HENDRICKSON, AND MELVIN L. DEPAMPHILIS*
Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

Received 19 November 1984/Accepted 4 February 1985

Unique single-stranded regions of simian virus 40 DNA, phage M13 virion DNA, and several homopolymers were used as templates for the synthesis of (p)ppRNA-DNA chains by CV-1 cell DNA primase-DNA polymerase α. Intact RNA primers, specifically labeled with an RNA capping enzyme, were typically 6 to 8 ribonucleotides long, although their lengths ranged from 1 to 9 bases. The fraction of intact RNA primers 1 to 4 ribonucleotides long was 14 to 73%, depending on the template used. RNA primer length varied among primers initiated at the same nucleotide, as well as with primers initiated at different sites. Thus, the size of an RNA primer depended on template sequence. Initiation sites were identified by mapping 5′ ends of nascent RNA-DNA chains on the template sequence, identifying the 5′-terminal ribonucleotide, and partially sequencing one RNA primer. A total of 56 initiation events were identified on simian virus 40 DNA, an average of 1 every 16 bases. Some sites were preferred over others. A consensus sequence for initiation sites consisted of either 3′-dCTTT or 3′-dCCC centered within 7 to 25 pyrimidine-rich residues; the 5′ ends of RNA primers were complementary to the dT or dC. High ATP/GTP ratios promoted initiation of RNA primer synthesis at 3′-dCTTT sites, whereas low ATP/GTP ratios promoted initiation at 3′-dCCC sites. Similarly, polydeoxythymidylic acid and polydeoxycytidylic acid were the only effective homopolymer templates. Thus, both template sequence and ribonucleoside triphosphate concentrations determine which initiation sites are used by DNA primase-DNA polymerase α. Remarkably, initiation sites selected in vitro were strikingly different from initiation sites selected during simian virus 40 DNA replication in vivo.

Simian virus 40 (SV40) provides a simple, well-characterized model for the replication of a single replicon in mammalian chromosomes (6, 7). Initiation of SV40 DNA replication requires a unique, *cis*-acting, 65-base-pair (bp) viral DNA sequence (*ori*), the virus-encoded protein T-antigen, and one or more factors found in the nuclei of permissive monkey cells. The interaction of these three components results in bidirectional replication which originates from nucleotides 5223 and 5224 (strain wt800 [13]) located within *ori* (12). DNA synthesis on the forward arm of a replication fork occurs in the same direction as fork movement and is a relatively continuous process, but DNA synthesis on the retrograde arm must occur in the opposite direction and therefore is a discontinuous process involving the repeated initiation of Okazaki fragments (5). Thus, as a replication fork advances, its forward arm is maintained as double-stranded DNA, while its retrograde arm will contain a short stretch of single-stranded DNA template that provides an Okazaki fragment initiation zone (6, 7).

RNA primers, 9 to 11 bases long, are initiated in vivo within the initiation zone at one of several possible sites. The primary sites are 3′-deoxypurine-deoxyribosylthymine (3′-dPuT) and secondary sites are 3′-deoxypurine-deoxyribosylcytidine (3′-dPuC), with the 5′ end of the RNA primer complementary to either dT or dC (13). Although these sites occur, on average, once every 7 nucleotides, initiation events only occur once every 135 bases (the average size of Okazaki fragments [1]). Therefore, once synthesis of an RNA primer is initiated, elongation of the resulting nascent chain appears rapid enough to prevent additional initiation events downstream. Neither the template sequence encoding the RNA primer nor the transition point in the template where RNA synthesis changes to DNA synthesis shows any sequence preference (13). Accordingly, the resulting RNA primers begin with either pppA or pppG and exhibit a heterogeneous internal base sequence and a near-random composition of ribonucleoside-p-deoxynucleoside junctions (1, 4).

In previous experiments (M. Yamaguchi, E. A. Hendrickson, and M. L. DePamphilis, J. Biol. Chem., in press), DNA primase-DNA polymerase α purified from CV-1 cells (a permissive host for SV40 DNA replication) was shown to synthesize RNA-primed DNA chains on natural DNA templates that were similar to Okazaki fragments in replicating SV40 DNA. RNA primers were initiated with A or G, and most of the primers were pppA/G$(pN)_{5-7}$. However, the composition of the 5′-terminal nucleotide on RNA primers synthesized in vitro depended on the relative amounts of ATP and GTP in the reaction mixture, whereas the lengths of RNA primers were affected by the relative amounts of CTP and UTP. These data suggest that the structure of RNA primers synthesized by DNA primase in vivo could be modulated by the relative concentrations of ribonucleoside triphosphates, perhaps by altering the selection of initiation sites on the template.

In this study, the precise nucleotide locations of 56 RNA primer initiation sites were mapped on 1,102 nucleotides of SV40 DNA. Site selection in vitro depended on the relative concentrations of ATP and GTP as well as on template sequence. Template sequence also affected the size of RNA primers synthesized. The composition and location of these sites were then compared with those of sites selected in vivo when the same DNA sequences were replicated in SV40-infected CV-1 cells.

MATERIALS AND METHODS

Enzymes. T4 DNA polymerase was purchased from P-L Biochemicals; DNA polymerase I Klenow fragment, restric-

* Corresponding author.

tion enzymes *Bam*HI, *Sph*I, *Hin*dIII, and *Msp*I were purchased from New England Biolabs; nuclease P1 was purchased from Boehringer-Mannheim Chemicals; and vaccinia guanylyltransferase, calf thymus terminal deoxynucleotidyl transferase, RNase T1, and RNase PhyM were purchased from Bethesda Research Laboratories. DNA primase-DNA polymerase α was purified from African green monkey CV-1 cells (Yamaguchi et al., in press). Fraction VIII-A (200,000 U of primase activity per mg of protein), which was free of ATPase and both endo- and exonuclease activities, was used throughout this study.

Chemicals. Nucleotides and GpppNs were purchased from P-L Biochemicals, polyethyleneimine (PEI)-cellulose was purchased from Brinkmann Instruments, Inc., and radioactive nucleotides were purchased from New England Nuclear Corp. Sepharose CL-4B and CL-2B were purchased from Pharmacia Fine Chemicals, Inc., Biogel P-60 was purchased from Bio-Rad Laboratories, and low-melting-point agarose was purchased from Bethesda Research Laboratories. Phenol was prepared as described previously (18).

Templates. Polydeoxyadenylic acid [poly(dA)], polydeoxyguanylic acid [poly(dG)], polydeoxythymidylic acid [poly(dT)], polydeoxycytidylic acid [poly(dC)], polyuridylic acid [poly(U)], and polycytidylic acid [poly(C)] ($S_{20,w}$ = 5 to 9) were purchased from P-L Biochemicals. Single-stranded circular DNA from M13mp7 (7,238 nucleotides) or mSV clones was isolated as described previously (12). The self-complementary sequence in the polylinker region of M13mp7 DNA allowed the single-strand SV40 DNA inserts to be released by digesting 300 μg of recombinant phage DNA with *Bam*HI at 37°C for 20 h (20) under the conditions recommended by the supplier. Samples were then extracted once with phenol and chromatographed on a Sepharose CL-4B column (0.9 by 20 cm) equilibrated with 10 mM Tris-hydrochloride (pH 8.0)–1 mM EDTA–20 mM NaCl to separate insert DNA from M13 DNA which eluted in the void volume. Fractions containing the insert were identified by subjecting an aliquot to electrophoresis in a 1% agarose gel (21). DNA was visualized by ethidium bromide staining, fractions containing the SV40 inserts were pooled and adjusted to 0.3 M sodium acetate (pH 7.0), and the DNA was concentrated by precipitation with 75% ethanol for 16 h at −20°C. Three cycles of Sepharose CL-4B column chromatography were required to obtain SV40 insert DNA free of vector DNA as judged by ethidium bromide staining of agarose gels.

Synthesis by DNA primase-DNA polymerase α. DNA synthesis was routinely carried out at 30°C for 20 min in 100 μl containing 50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*'-2-ethanesulfonic acid [pH 7.8]); 15% glycerol; 13 mM magnesium acetate; 10 mM sodium acetate; 100 μg of bovine serum albumin per ml; 1 mM dithiothreitol; 5 μg of DNA; 0.1 mM (4 μCi) each of [^{3}H]dATP, [^{3}H]dGTP, [^{3}H]dTTP, and [^{3}H]dCTP; 0.2 mM each of GTP, CTP, and UTP; 4 mM ATP; and 8 U of DNA primase-DNA polymerase α. Reactions were stopped with 15 mM EDTA and 0.5% sodium dodecyl sulfate, and then the reaction mixtures were incubated at 37°C for 30 min with 100 μg of proteinase K per ml. After phenol extraction, the products of the reaction were purified by gel filtration through a Biogel P-60 column (0.7 by 18 cm) with 10 mM Tris-hydrochloride (pH 8.0)–1 mM EDTA–20 mM NaCl. DNA in the pass-through was concentrated by ethanol precipitation.

Capping reaction. RNA-DNA chains containing a di- or triphosphate at their 5' ends were specifically radiolabeled in the capping reaction. Samples were denatured for 2 min at 100°C in 10 mM Tris-hydrochloride (pH 7.6)–1 mM EDTA and then rapidly cooled in ice water. The capping reaction was carried out in a 20-μl volume containing 50 mM Tris-hydrochloride (pH 7.9), 1.25 mM $MgCl_2$, 6 mM KCl, 2.5 mM dithiothreitol, 10 μM [α-^{32}P]GTP (2,600 to 3,000 mCi/μmol), and 6 U of vaccinia virus guanylyltransferase. After incubation at 37°C for 45 min, the reaction was terminated by the addition of 1 μl of 0.3 M EDTA. The products were purified by chromatography on Biogel P-60 as described above.

Nucleotide locations of initiation sites for DNA primase-DNA polymerase α: (i) linear mSV01 insert DNA as template. The template locations of 5'-end-labeled, capped nascent DNA chains were mapped at single-nucleotide resolution as described previously (12, 13) by annealing them to copies of their DNA template, cutting them at a unique restriction endonuclease site, and measuring the lengths of the resulting radiolabeled DNA fragments by gel electrophoresis. DNA synthesized on mSV01 insert DNA template was radiolabeled in the capping reaction, suspended in denaturing loading buffer (98% formamide, 1 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF), denatured at 90°C for 2 min, and then separated from template DNA by electrophoresis in a 10% polyacrylamide gel containing 8 M urea (17). [^{32}P]DNA in the 40-to-290-nucleotide range was recovered by extracting that portion of the gel in 0.5 M ammonium acetate–10 mM magnesium acetate–0.1% sodium dodecyl sulfate at 60°C for 16 h. mSV01 single-stranded, circular DNA (10 μg) was added to the sample, and DNA was precipitated in 75% ethanol for 16 h at −20°C. The sample was centrifuged in a Beckman Microfuge 12 horizontal rotor for 15 min. The pellet was suspended in 0.3 M NaCl–10 mM Tris-hydrochloride (pH 8.0)–1 mM EDTA, denatured at 100°C for 2 min, renatured at 65°C for 3 h, and then slowly cooled to room temperature. The 3' ends of hybridized fragments were extended by incubating the DNA at room temperature for 2 h with 1.5 U of DNA polymerase I (Klenow fragment) in 50 mM Tris-hydrochloride (pH 7.4)–10 mM $MgCl_2$–1 mM dithiothreitol–0.25 mM dTTP–0.25 mM dGTP–0.25 mM dCTP–0.25 mM dATP. DNA was purified by chromatography on a Sepharose CL-2B column (0.7 by 18 cm) with 100 mM NaCl–10 mM Tris-hydrochloride (pH 8.0)–1 mM EDTA. Material in the void volume was precipitated with ethanol as above, suspended in buffer, and digested with *Hin*dIII under the conditions specified by the supplier. SV40(I) DNA was included as an internal standard, and the reaction was terminated by extracting it once with an equal volume of phenol when all of the SV40(I) DNA was digested. Digestion was monitored by electrophoresis in 1% agarose minigels that were stained with ethidium bromide (18). DNA was precipitated with ethanol, suspended in denaturing loading buffer (see above), denatured at 90°C for 3 min, and then fractionated by electrophoresis in an 8% polyacrylamide–8 M urea gel (25).

The sequence of an appropriate DNA fragment was determined in parallel on the same gel. Double-stranded mSV01(RF) DNA (4 μg in 20 μl) was digested with 10 U of *Hin*dIII. The 3' ends of the DNA were labeled by incorporation of a single [α-^{32}P]dATP residue with Klenow fragment (18). This DNA was purified as described above and digested with 10 U of *Bam*HI under conditions suggested by the supplier, combined with 5 μl of nondenaturing loading buffer (50% glycerol, 100 mM EDTA, 0.02% bromophenol blue, 0.02% xylene cyanol FF), and then fractionated by electrophoresis in a 6% polyacrylamide gel (13). The 240-bp [3'-^{32}P]DNA fragment was recovered by electroelution (12), precipitated with ethanol as above, suspended in 25 μl of

water, and sequenced by the chemical degradation method of Maxam and Gilbert (17). One [^{32}P]dATP added to the 3′ end of the *Hin*dIII site was compensated for by the one nucleotide lost at the site of cleavage.

(ii) Circular mSV01 DNA as template. Capped products synthesized on single-stranded, circular mSV01 DNA were chromatographed on Biogel P-60, recovered in the void volume, and precipitated with ethanol as described above. The sample was suspended in 20 μl of 0.3 M NaCl–10 mM Tris-hydrochloride (pH 7.6)–1 mM EDTA, combined with 5 μg of mSV01 circular, viral DNA, and treated as described above for mSV01 insert DNA. In this case, DNA sequencing was carried out with the 418-bp *Hin*dIII-*Msp*I fragment from SV40(I) DNA. After digestion with *Hin*dIII, the five DNA fragments were separated by electrophoresis in 1% low-melting-point agarose. The gel was stained with ethidium bromide, and the section containing 1,169- and 1,118-bp fragments was incubated in 50 mM Tris-hydrochloride (pH 8.0)–0.5 mM EDTA at 65°C for 5 min to melt the agarose. After phenol extraction, DNA fragments were purified by DEAE-Sephacel column chromatography (18) and precipitated with ethanol. 3′ Ends were labeled by incorporation of one [α-^{32}P]dATP residue (18), and the DNA was purified as described above and digested with 8 U of *Msp*I under conditions suggested by the supplier. [3′-^{32}P]DNA fragments were fractionated by electrophoresis in 6% polyacrylamide, and the appropriate one was recovered by electroelution (12) and sequenced (17).

(iii) Linear mSV02 insert as template. DNA was synthesized on mSV02 insert DNA template. After phenol extraction and ethanol precipitation, DNA products were extended by 3 U of Klenow fragment in 100 μl of a solution containing 50 mM Tris-hydrochloride (pH 7.4), 10 mM $MgCl_2$, 1 mM dithiothreitol, and 0.25 mM each of dATP, dGTP, dTTP, and dCTP for 2 h at room temperature. The sample was extracted with phenol and chromatographed on Biogel P-60, and the DNA in the void volume was precipitated with ethanol and redissolved in 28 μl of H_2O. One aliquot was precipitated with ethanol; the other was cleaved with 8 U of *Sph*I under the conditions specified by the supplier and extracted with phenol, and the DNA was precipitated with ethanol. The two samples were suspended in 5 mM Tris-hydrochloride (pH 7.6)–0.5 mM EDTA, denatured for 2 min at 100°C, and then radiolabeled in the capping reaction. Both samples were then extracted with phenol, chromatographed on Biogel P-60, precipitated with ethanol, suspended in denaturing loading buffer, denatured at 90°C for 3 min, and then fractionated by electrophoresis in an 8% polyacrylamide–8 M urea gel (25) in parallel with the products from a DNA sequence analysis.

Two [3′-^{32}P]DNA fragments were prepared for sequencing. One was generated from double-stranded mSV02(RF) DNA (4.5 μg) cleaved with *Sph*I and labeled at its 3′ termini with one molecule of α-[3′-^{32}P]dATP (cordycepin triphosphate, 5,000 Ci/mmol) with calf thymus terminal transferase (29). Reactions were terminated by extraction with phenol, and the DNA was precipitated with ethanol. DNA was suspended in 20 μl of buffer and digested with 10 U of *Bam*HI under conditions suggested by the supplier. After reaction, the DNA was precipitated with ethanol, dissolved in denaturing loading buffer, heated at 90°C for 3 min, and fractionated by electrophoresis in a 6% polyacrylamide–8 M urea gel. The 280-bp [3′-^{32}P]DNA fragment was recovered by electroelution (12) and sequenced (17). The second [3′-^{32}P]DNA fragment was generated from the 311-base mSV01 insert DNA which was isolated as described above. The fragment was radiolabeled at its 3′ terminus by terminal transferase and α-[3′-^{32}P]dATP (29), purified by electrophoresis in a 6% polyacrylamide–8 M urea gel, and sequenced (17).

(iv) Linear mSV07 and mSV09 insert DNAs as templates. DNA synthesized on either mSV07 or mSV09 insert DNA was extended by Klenow fragment, extracted with phenol, chromatographed on Biogel P-60, and then radiolabeled in the capping reaction as described for mSV02. The capped products were chromatographed on Biogel P-60, precipitated with ethanol, redissolved, denatured, and then fractionated by gel electrophoresis in an 8% polyacrylamide–8 M urea gel as described above. Either mSV08 or mSV10 insert DNA, labeled at the 3′ termini with α-[3′-^{32}P]dATP (29), was used to provide sequence standards (17).

Polyacrylamide gel electrophoresis. Polyacrylamide gels (0.25 mm by 33 cm by 42 cm) containing 8% acrylamide with a 20:1 ratio of acrylamide to *N,N′*-methylenebisacrylamide were prepared in 8 M urea–100 mM Tris-borate (pH 8.3)–1 mM EDTA (25). Electrophoresis was carried out for 2 to 3 h at 2,000 V. Gels were transferred to either Whatman 3MM paper or used X-ray film and exposed to Kodak X-Omat AR film at −70°C with a Cronex intensifying screen. Electrophoresis in 22% polyacrylamide gels (0.5 mm thick) was carried out for 5 to 8 h at 1,500 V (12). Electrophoresis in preparative gels (1.5 mm by 15 cm by 30 cm) used to isolate labeled DNA fragments was carried out for 5 to 8 h at 500 V (13).

Recovery of DNA primase-DNA polymerase α products from polyacrylamide gels. At least 90% of the radiolabeled DNA was recovered from individual slices of an 8% polyacrylamide–8 M urea gel by extraction in 0.5 M ammonium acetate–10 mM magnesium acetate–0.1% sodium dodecyl sulfate at 60°C for 16 h. After centrifugation for 5 min at 12,000 × *g*, the supernatant was supplemented with 5 μg of tRNA, and nucleic acids were precipitated with ethanol, dissolved in 100 μl of water, precipitated twice with ethanol-sodium acetate to remove sodium dodecyl sulfate, and finally dissolved in 7 μl of water for subsequent treatment with nucleases.

Identification of 5′-terminal GpppN on nascent RNA-DNA chains. DNA was labeled in the capping reaction, purified, and digested with 2 μg of nuclease P1 in 10 μl of 30 mM sodium acetate (pH 5.3) containing 2 μg of tRNA for 2 h at 37°C. The products were chromatographed on PEI-cellulose in 1.6 M LiCl together with unlabeled cap standards. Radioactive spots identified by autoradiography were excised, and their radioactivities were quantitated by scintillation counting. Standards were visualized by UV light.

RESULTS

Template preference. The importance of template sequence on the ability of DNA primase-DNA polymerase α to initiate RNA-primed DNA synthesis was evaluated by using a variety of homopolymer and natural DNA templates. Initiation of RNA primer synthesis was measured by radiolabeling RNA-DNA chains produced by DNA primase-DNA polymerase α with [α-^{32}P]GTP by using vaccinia virus guanylyltransferase (i.e., the capping reaction). This reaction adds GMP specifically to the 5′ end of ppRNA or pppRNA. The advantage of the capping reaction is that only RNA-DNA chains which retained their original 5′ nucleotide are radiolabeled, and the amount of radioactivity per chain is independent of chain length or composition (27, 30, 35; Yamaguchi et al., in press). Removal of 5′-terminal phosphates with bacterial alkaline phosphatase reduced capping by 94%.

TABLE 1. Template preferences

Template	fmol (%) of $\overset{*}{G}$pppN synthesized[a]
poly(dT)	184 (100)
poly(dC)	9 (5)
poly(dA)	0 (0)
poly(dG)	0 (0)
poly(U)	0 (0)
poly(C)	0 (0)
M13mp7	45 (24)
mSV01	27 (15)
mSV02	40 (22)
mSV07	21 (11)
mSV08	15 (8)
mSV09	22 (12)
mSV10	9 (5)

[a] Synthesis of $\overset{*}{G}$pppN was calculated from the radioactivity of RNA primers in polyacrylamide gels (see Fig. 2). mSV templates were single-stranded linear insert DNA templates prepared from restriction fragments of SV40 DNA that had been cloned into M13. mSV01 and mSV02 each contain one of the two complementary strands of the 311-bp *Bst*NI G fragment from SV40 DNA that includes *ori* (12); mSV01 represents the early mRNA template and mSV02 represents the late mRNA template. Since bidirectional replication begins in the *ori* region, part of mSV01 and part of mSV02 represent a retrograde template and part represent a forward template (Fig. 4). mSV09 and mSV10 each contain one of the two complementary strands from a 263-bp fragment located about 300 bp from *ori* (13). mSV07 and mSV08 contain a 360-bp fragment located about 1,630 bp from *ori* (13). mSV08 and mSV10 contain the retrograde templates of replication forks in these regions of the genome, and mSV07 and mSV09 contain the corresponding forward templates.

Initiation of RNA-primed DNA synthesis was 20-fold better on poly(dT) than on poly(dC) (Table 1), although activity on these templates depended upon reaction conditions. Preference for poly(dC) increased 200-fold when 4 mM GTP and 0.1 mM dGTP alone were present, conditions that were optimal for this template (Yamaguchi et al., in press), and ATP has been shown to inhibit activity on poly(dC) (11). Synthesis was not detected on poly(dA), poly(dG), poly(U), or poly(C). Natural DNA (M13mp7) was about one-fourth as effective as poly(dT). Moreover, utilization of a particular natural DNA sequence (mSV templates) varied as much as fivefold.

To determine the 5′-terminal ribonucleoside di- or triphosphate composition of nascent chains, the products from each reaction were digested with nuclease P1 to release $\overset{*}{G}$pppN cap structures which were then chromatographed on PEI-cellulose. In each case, all of the radiolabel was released as either GpppA or GpppG (Fig. 1). With natural DNA templates, DNA primase-DNA polymerase α initiated synthesis predominantly (79 to 87%) with ATP, suggesting that initiation sites containing dT are preferred over ones containing dC. However, on the mSV02 template, initiation events began predominantly (78%) with GTP. Thus, the choice of ATP or GTP for initiating RNA primer synthesis on natural DNA templates depended strongly on the template sequence.

The effect of a particular template on the size of RNA primers was also determined. DNA primase-DNA polymerase α products were radiolabeled in the capping reaction and then digested with phage T4 DNA polymerase, 3′ to 5′ exonuclease (T4 exo) under conditions that remove all but a single deoxynucleoside monophosphate dNMP from the 3′ ends of RNA chains (16; Yamaguchi et al., in press). Before digestion with T4 exo, the radiolabeled polynucleotides synthesized on mSV templates ranged in length from 25 bases to the size of the template, and products synthesized on M13 DNA were 50 to about 700 bases (Fig. 2, T4 exo − lanes). Accurate sizes were determined by electrophoresis in an 8% polyacrylamide–8 M urea gel (data not shown). The major products produced on poly(dT) were about 30 bases long, whereas most of those synthesized on poly(dC) were shorter than 10 bases. Treatment of radiolabeled nascent DNA with NaOH converted all radiolabel to GpppNp (identified by PEI-cellulose chromatography [Yamaguchi et al., in press]), demonstrating that all of the radiolabel was in RNA (Fig. 2, + OH^-). GpppNp comigrated with $p(Ap)_2$ (nucleotide position 2) during gel electrophoresis.

After digestion with T4 exo, 5′-end-labeled oligonucleotides appeared in the region identified by $p(Ap)_3$ to $p(Ap)_{12}$ (Fig. 2, T4 exo + lanes). Their apparent size, judged by the

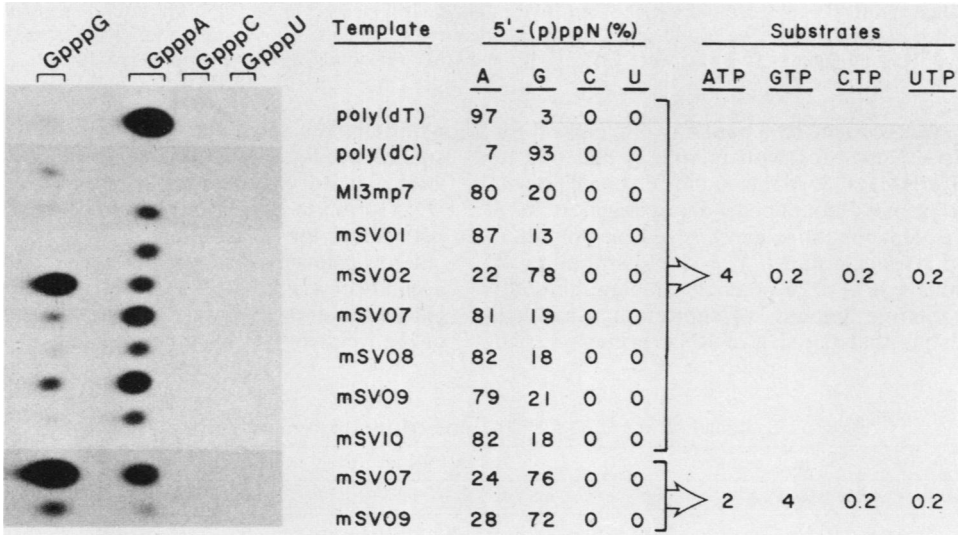

FIG. 1. 5′-Terminal nucleotide composition of RNA primers synthesized on various templates. Nascent DNA was synthesized by DNA primase-DNA polymerase α on various templates under either standard conditions (4 mM ATP, 0.2 mM GTP, 0.2 mM CTP, 0.2 mM UTP) or under high-GTP condition (2 mM ATP, 4 mM GTP, 0.2 mM CTP, 0.2 mM UTP). DNA was labeled in the capping reaction, purified, and digested with nuclease P1, and the products were chromatographed on PEI-cellulose. The positions of cap standards are indicated. mSV templates were single-stranded linear insert DNA.

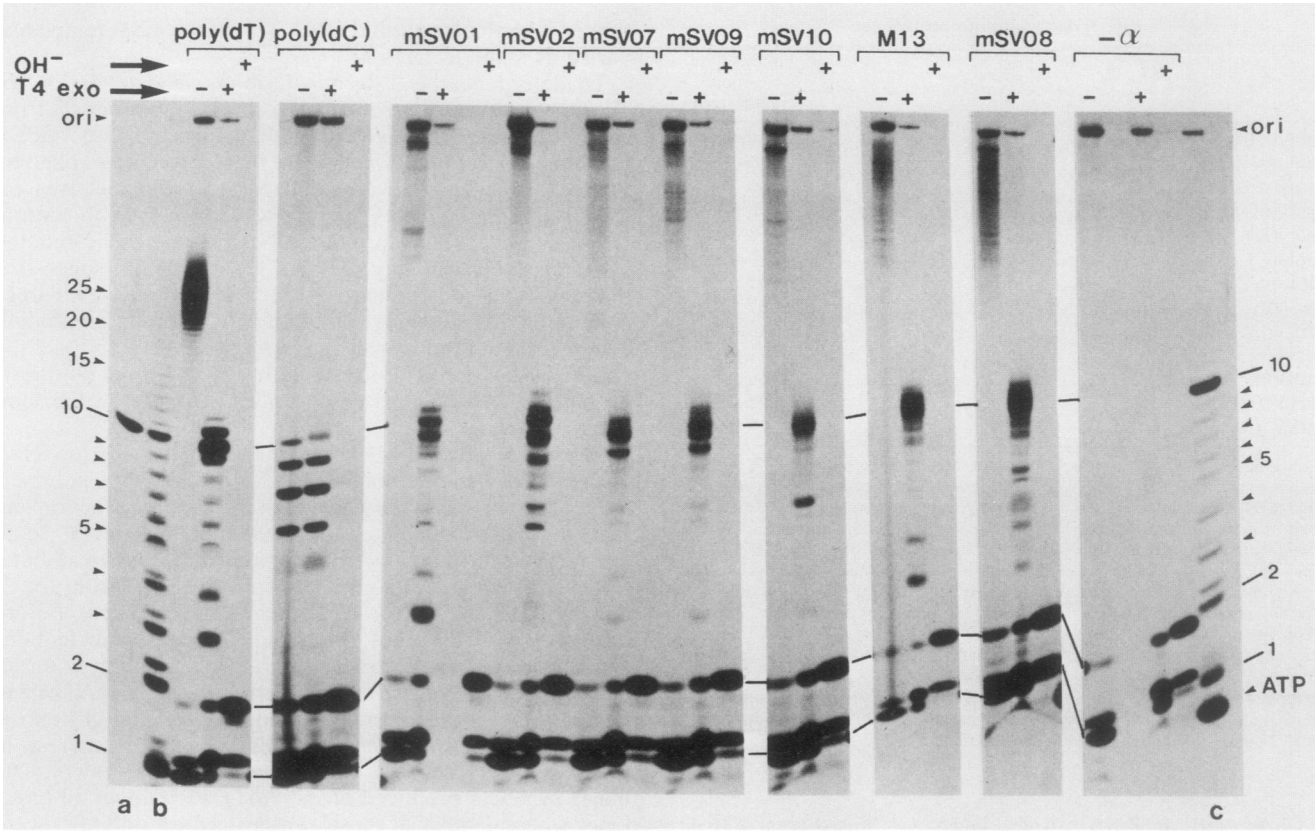

FIG. 2. Sizes of RNA primers synthesized on homopolymer and natural DNA templates. (p)ppRNA-DNA chains were synthesized under standard assay conditions with different DNA templates. Their 5′ ends were then labeled with [α-^{32}P]GTP by using vaccinia guanylyltransferase. The GpppN-labeled products were either left untreated (T4 exo − lanes) or treated with either T4 exonuclease (T4 exo + lanes) or NaOH (OH^- + lanes) and then fractionated by electrophoresis in a 22% polyacrylamide–8 M urea gel. The nine templates used are indicated at the top and described in the text. Lanes −α, Control reactions with mSV08 insert DNA but without DNA primase-DNA polymerase α; lane a, 5′-^{32}P-labeled $(pA)_{10}$ size standard; lane b, 5′-^{32}P-labeled $p(Ap)_n$ size standards generated by partial alkaline hydrolysis of poly(rA) (8); lane c, both 5′-^{32}P-labeled $p(Ap)_n$ and mSV02 5′-^{32}P-labeled insert DNA, which remains at the origin (ori). Numbers on the vertical axis indicate the size in nucleotides of $p(Ap)_n$ containing a 2′ or 3′ phosphate. The slower migrating satellite bands in lanes b and c appear to be the 2′,3′-cyclic-phosphate form (4). Unreacted [γ-^{32}P]ATP from the polynucleotide kinase reaction appears in lane c. To obtain autoradiograms with bands of comparable intensities, poly(dC), mSV10, and mSV08 products were exposed for 48 h; the other products were exposed for 24 h. Radiolabeled material formed in the absence of DNA primase-DNA polymerase α (lanes −α) that remained at the origin of the gel was not digested by T4 exo, although it was sensitive to alkali. It appeared to be [^{32}P]GMP bound to guanylyltransferase in the absence of an acceptor molecule (26). The bands appearing in the vicinity of nucleotide positions 1 and 2 [corresponding to standards pAp and $p(Ap)_2$] in the undigested samples (−lanes) were GTP, GDP, and GMP remaining from the capping reaction.

$p(Ap)_n$ standards, was reduced by 3 bases: two to correct for the 5′-cap [$Gpp(pA)_n$ migrates with $p(Ap)_{n+2}$] and one to correct for the 3′-dNMP (35; Yamaguchi et al., in press). Therefore, RNA primers contained, on average, 6 to 8 ribonucleotides, except for those synthesized on poly(dC), which averaged 5 ribonucleotides. The size distribution of oligoribonucleotides synthesized on each template differed, however, in the relative amounts of shorter than average molecules, suggesting that the size of RNA primers varied with template sequence (Fig. 2; Table 2). The similarity in size of products synthesized on poly(dC) template before and after T4 exo digestion demonstrated that coupling of DNA primase with DNA polymerase α activity was very poor under these conditions.

DNA primase-DNA polymerase α initiation sites in and around the origin of SV40 DNA replication. (i) mSV01 template. The above results suggested that the number of sites or the frequency at which each site was used or both varied

TABLE 2. Length distribution for RNA primers synthesized on various templates

Length (bases)	Distribution (%)[a]								
	poly(dT)	poly(dC)	m13mp7	mSV01	mSV02	mSV07	mSV08	mSV09	mSV10
1–2	10	12	22	38	6	9	19	10	9
3–4	4	61	12	11	17	11	16	9	21
5–9	86	27	66	51	77	80	65	81	70

[a] The proportion of RNA primers of each length was calculated by excising the appropriate parts of each gel lane (Fig. 2) and measuring the radioactivity by scintillation counting. Lengths of RNA primers were corrected for the presence of a cap structure and 3′-dNMP as described in the text. mSV templates were single-stranded linear insert DNA.

with the sequence of the template. Therefore, the nucleotide locations for initiation of RNA primer synthesis were mapped on several templates representing different regions of the SV40 genome. mSV01 insert DNA is a 311-base, single-stranded, linear molecule containing the SV40 *ori* region. The 5′ termini of polynucleotide chains synthesized on this template by DNA primase-DNA polymerase α were labeled in the capping reaction. 5′-G*pppN chains were then hybridized with single-stranded, circular mSV01 DNA and extended with *Escherichia coli* DNA polymerase I (Klenow fragment) to insure that all nascent DNA chains were elongated past the single *Hin*dIII site located 239 bases downstream from the 3′ end of the template (position 5184). One aliquot was digested with *Hin*dIII to release nascent chains with identical 3′ termini, and one aliquot was left untreated as a control. Both aliquots were heat denatured and fractionated by electrophoresis in a polyacrylamide-urea gel. The nucleotide locations of 5′ ends of RNA primers were determined by comparing the migration of these bands with those released by sequencing DNA chains of the same polarity as nascent DNA. A similar approach was used previously to map nascent SV40 DNA chains synthesized in vivo (12, 13). The length of each [3′-^{32}P]DNA fragment produced in the sequencing reaction marked the location of its 5′ end with respect to the *Hin*dIII site.

Digestion of DNA hybrids with *Hin*dIII released six unique RNA-DNA chains that mapped around position 5234 (Fig. 3, lanes b and e). No other bands were detected even in longer autoradiographic exposures. In this and subsequent experiments, the relative amount of each 5′-end-labeled RNA-DNA chain reflected the relative frequency of initiation at that site. The precise template location for the 5′ ribonucleotide of each RNA-DNA chain was obtained by correcting for the presence of a cap and for the fact that nascent DNA chains migrated approximately one-half nucleotide out of register with the DNA sequencing tracks. However, to avoid any ambiguities in the interpretation of these data, each of the six bands (Fig. 3, lane e) was excised from the gel, and the 5′ end nucleotide of each was identified (Fig. 3, lanes p through u). Thus, the six RNA-DNA chains in lane e (top to bottom) contained 5′ terminal (p)pp-A, -A, -A, -G, -A, and -A. The template positions consistent with both analyses (Fig. 4) revealed that these 5′-capped RNA-DNA chains migrated as chains 2.5 bases longer than a DNA chain of the correct length; 2 bases can be accounted for by the cap moiety, and 0.5 base can be accounted for by the RNA primer (13). Only initiation events from SV40 nucleotide 5210 to 161 were mapped in this experiment (Fig. 4).

Since the initiation sites were restricted to a small region in the center of the linear template, concern was given to the possibility that free 5′ or 3′ ends of the linear template might have influenced site selection by the DNA primase-DNA polymerase complex. Therefore, single-stranded, circular mSV01 DNA was used as a template. Since this template was 23-fold longer than mSV01 insert DNA, the relative intensity of each start site was reduced, and the background was increased in both digested and undigested lanes. Nevertheless, *Hin*dIII released at least three distinct bands (Fig. 3, lane k), and these were identical to the major bands observed with mSV01 insert DNA (lane b). These bands were absent from undigested samples (Fig. 3, lane j). Therefore, the initiation sites mapped in these experiments were specified by DNA sequences rather than by DNA ends.

(ii) mSV02 template. DNA primase initiation sites on mSV02 insert DNA, the complementary sequence to mSV01 insert DNA, were mapped by using a similar technique (28). The products synthesized by DNA primase-DNA polymerase α were extended to the end of the template by incubation with Klenow fragment. The 5′ ends of (p)ppRNA-DNA chains were labeled in the capping reaction, and the products were fractionated by gel electrophoresis under denaturing conditions. In the 295-base region that was mapped in this experiment (nucleotides 5106 to 145; Fig. 4), 21 unique DNA chains were observed (Fig. 5, lanes b, i, and j). To locate the 5′ ends of these RNA-primed DNA chains, two corrections in length were necessary. First, both mSV01 and mSV02 insert DNA were released from the vector by cleavage with *Bam*HI, which leaves a 4-base 5′ overhang. DNA sequence analysis was performed on 3′-^{32}P-labeled mSV01 insert DNA, and in vitro DNA synthesis was carried out with mSV02 insert DNA. Alignment of these two complementary sequences reveals that nascent DNA chains synthesized to the 5′ end of mSV02 templates will be 4 bases longer than the 3′-terminal nucleotide on mSV01 sequence standards (28). Second, RNA-DNA chains that have been capped with GTP migrate as chains whose lengths have been increased by 2.5 nucleotides (see above). Therefore, the lengths of capped nascent RNA-DNA chains (Fig. 5) must be reduced by 6.5 bases each to identify the initiation sites shown in Fig. 4.

This conclusion was confirmed by excising each of the six major bands shown in lane i of Fig. 5, digesting them with nuclease P1, and then identifying the 5′-G*pppN products by PEI chromatography (Fig. 5, lanes q through v). These initiation sites (top to bottom) began with (p)pp-A, -A, -G, -G, -G, and -G.

To confirm that nascent chains had been elongated to the end of the template, the extended products were cut at their single *Sph*I site, which lies very near the 5′ end of this template (nucleotide 133), and fractionated by gel electrophoresis. The RNA-DNA chains present before *Sph*I digestion (Fig. 5, lanes b, i, and j) were all shortened by approximately 30 bases (Fig. 5, lanes c, k, and l). The initiation sites identified on mSV02 by *Sph*I cleavage were identical to those of the undigested sample (Fig. 4).

DNA primase-DNA polymerase α initiation sites on templates outside the origin of SV40 DNA replication. The 5′ termini of RNA-DNA chains produced by DNA primase-DNA polymerase α on single-stranded linear mSV09 insert DNA (263 bases) or mSV07 insert DNA (360 bases) templates were mapped as described for mSV02. mSV07 and mSV09 were chosen rather than mSV08 and mSV10 because they were better templates (Table 1). Of the 247 bases examined in mSV09, 19 distinct nucleotide initiation sites were detected (Fig. 6, lane g). Of the 345 bases examined in mSV07, 21 distinct initiation sites were observed (Fig. 6, lane b). To confirm the relationship between the electrophoretic migration of RNA-DNA chains and DNA sequence products, each of the major bands was excised from the gel, and the 5′-terminal nucleotide of each band was identified by nuclease P1 digestion and PEI-cellulose chromatography as described above (data not shown). As explained for mSV02, the combined data revealed that the lengths of nascent chains derived by a direct comparison with the lengths of sequenced DNA chains had to be shortened by 6.5 bases in the case of mSV09 and 7 bases in the case of mSV07 to locate the actual initiation sites. Positions of 15 major sites on mSV09 DNA and 14 major sites on mSV07 DNA are provided in Fig. 4. The lengths of nascent RNA-DNA chains before and after incubation with Klenow fragment were identical, indicating that DNA polymerase α had elongated

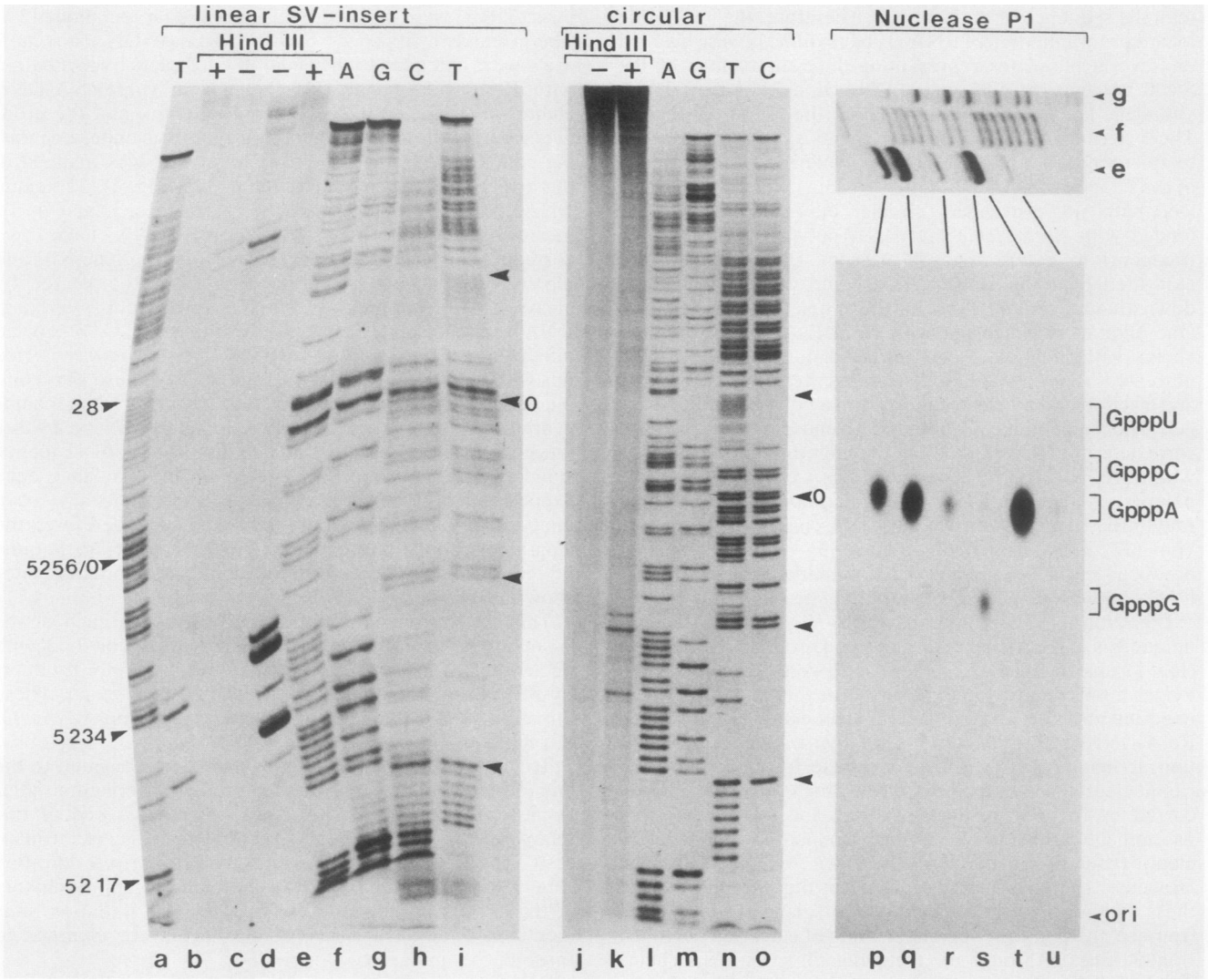

FIG. 3. Mapping the locations of 5′ ends of RNA primers synthesized on mSV01 (the early mRNA template strand containing the SV40 *ori* region). (p)ppRNA-DNA chains synthesized on either single-stranded linear mSV01 insert DNA (linear SV insert) or single-stranded circular mSV01 DNA (circular) were radiolabeled in the capping reaction, extended beyond the *Hin*dIII restriction site, and then digested with *Hin*dIII to release G*pppN-RNA-DNA chains, as described in the text. The sizes of these chains were determined by electrophoresis in an 8% polyacrylamide–8 M urea gel. G*pppN-RNA-DNA chains before *Hin*dIII digestion (*Hin*dIII − lanes) are shown in lanes c, d, and j. The same material after *Hin*dIII digestion (*Hin*dIII + lanes) is shown in lanes b, e, and k. Lanes b and c contain one-fourth as much sample as do lanes d and e. The nucleotide sequence of DNA chains with the same polarity as nascent DNA was determined concurrently by using the sequencing procedure of Maxam and Gilbert on [3′-^{32}P]DNA labeled at the *Hin*dIII site (A, cleavage at A and G; G, cleavage at G only; T, cleavage at T and C; C, cleavage at C only). Numbers on the vertical axis are nucleotide locations on the SV40 wt800 genome (13). The identity of 5′ nucleotides in G*pppN-RNA-DNA chains was confirmed by eluting each of the six major bands in lane e from the gel, digesting them individually with nuclease P1, and then fractionating the products by PEI-cellulose thin-layer chromatography (Nuclease P1, lanes p through u). The bands in lane e, along with the sequencing tracks in lanes f and g, are shown at the top of the PEI-cellulose strip. GpppU, GpppC, GpppA, and GpppG indicate the positions of cap standards. Samples were spotted on PEI-cellulose at the position marked ori.

the nascent DNA chains to the end of the template (Fig. 6, compare *Pol*I − lanes with *Pol*I + lanes).

Effect of ATP/GTP ratio on initiation site selection. The effect of the ATP/GTP ratio on site selection was determined by mapping the locations of nascent RNA-DNA chains synthesized on mSV07 and mSV09 insert DNA templates under different reaction conditions. RNA-DNA chains synthesized under conditions in which the ATP/GTP ratio was 20 began about four times more frequently with ATP than with GTP, whereas when the ratio was 0.5, DNA was initiated about three times more frequently with GTP than with ATP (Fig. 1). Based on the amount of [α-^{32}P]GTP incorporated in the capping reaction, the extent of RNA primer synthesis was the same under both conditions for mSV07 but was reduced 36% under high-GTP conditions for mSV09. Analysis of the initiation sites selected confirmed that the primary effect of changing the ATP/GTP ratio was to change the frequency at which each initiation site was used (Fig. 6). In most cases, a high ATP/GTP ratio promoted initiation at dT residues in the template, whereas a low ATP/GTP ratio promoted initiation at dC residues (Fig. 4). Some sites were only observed at either high or low GTP concentrations, reflecting a change in preference for template sequences. Surprisingly, however, three GTP initiation

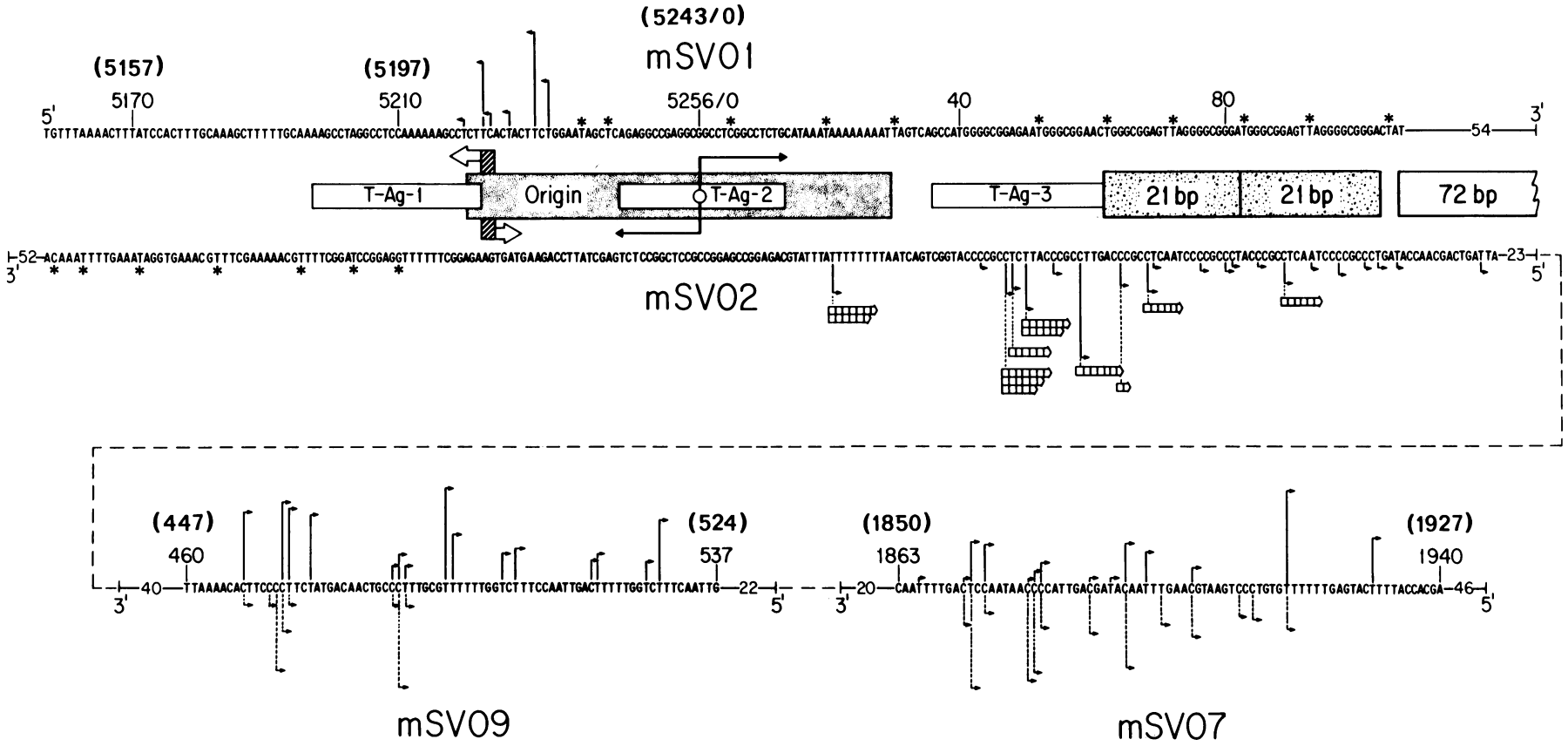

FIG. 4. Nucleotide locations of the 5′ ends of RNA primers synthesized on SV40 DNA templates in vitro and in vivo. In vitro results were as follows. The nucleotide sequence for the region analyzed in each of the four templates (mSV01, mSV02, mSV09, and mSV07) is shown, except in the case of mSV01 where position 5210 is the 5′ end of the region analyzed. In some cases, sequences at the 3′ and 5′ ends of the regions analyzed that did not contain initiation events were deleted for simplicity, and the number of missing bases is indicated. The mSV02, mSV09, and mSV07 templates are connected with a broken line to indicate that they are different regions of the same DNA strand. Vertical bars with an arrowhead designate the template nucleotide complementary to the 5′-terminal ribonucleotide in the RNA primer and indicate the direction of synthesis (Fig. 3, 5, and 6). Solid vertical bars represent initiation events under standard assay conditions (ATP/GTP ratio, 20), and broken vertical bars represent initiation events under high GTP conditions (ATP/GTP ratio, 0.5) (Fig. 6). The relative height of each bar represents the relative frequency of initiation events at each site on that template. The sizes of RNA primers initiated at specific sites on mSV02 (Fig. 7) are represented by open, blocked arrows; each block represents one nucleotide and the arrow indicates direction of synthesis. In vivo results were as follows. An asterisk indicates the 5′-terminal ribonucleotides of RNA primers observed during SV40 DNA replication in vivo (12). T-antigen (T-Ag)-binding sites 1, 2, and 3 (6), the minimum sequence required for the origin of replication (Origin [6]), a 27-bp palindrome ( ), 21- and 72-bp repeated sequences which contain gene promoter and enhancer elements, and the transition point from discontinuous to continuous DNA synthesis in vivo (i.e., the origin of bidirectional replication [12]) are indicated. Nucleotides are numbered with respect to the SV40 wt800 sequence (13) which contains an adenosine in place of a guanosine at position 5222 and also a change in the enhancer sequence resulting in 13 additional bases (13). Numbers for SV40 strain 776 are in parentheses.

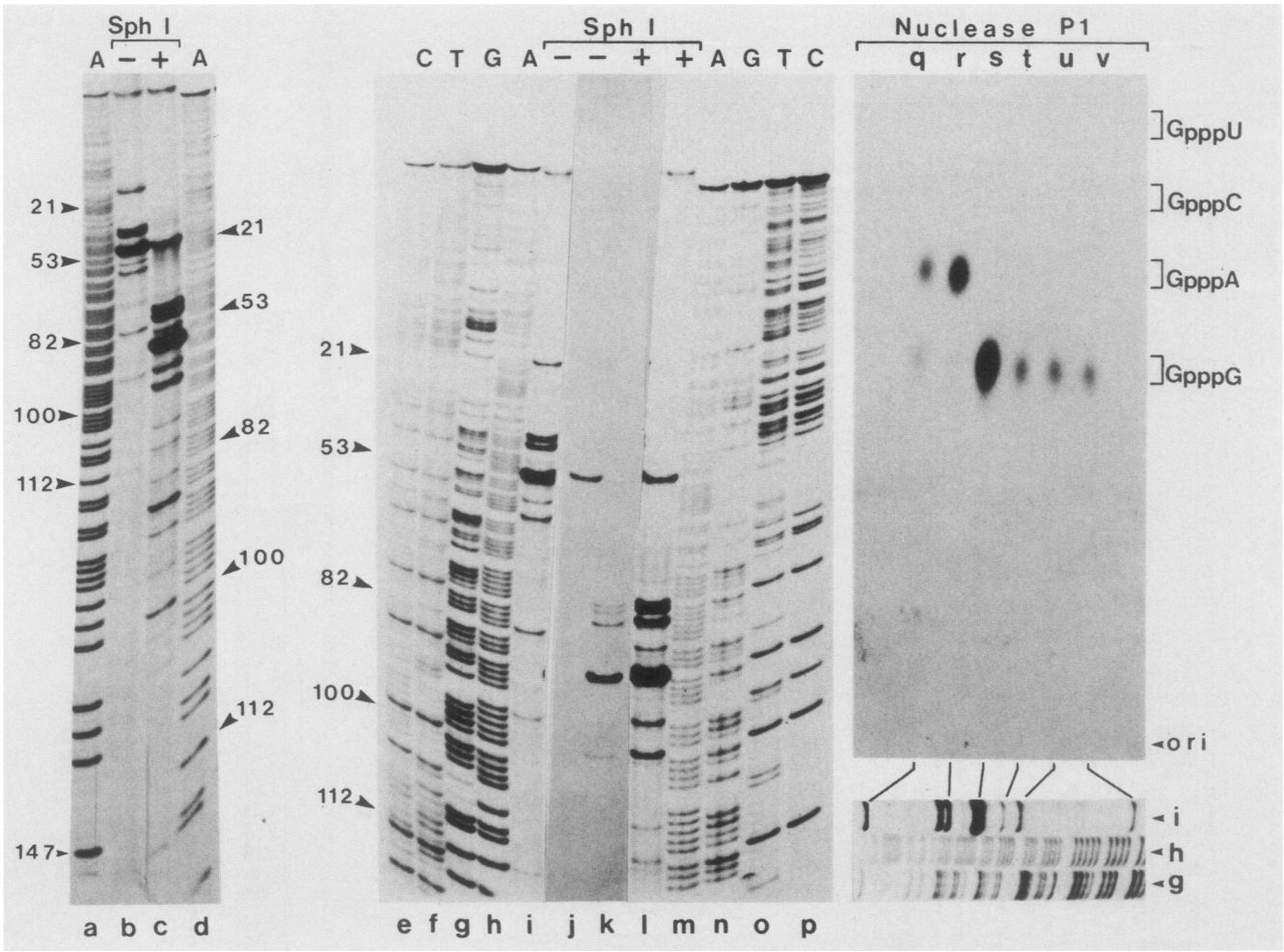

FIG. 5. Mapping the locations of 5′ ends of RNA primers synthesized on mSV02 (the late mRNA template strand containing the SV40 *ori* region). (p)ppRNA-DNA chains were synthesized on single-stranded linear mSV02 insert DNA and extended to the end of the template. The products were labeled in the capping reaction and analyzed by electrophoresis in an 8% polyacrylamide–8 M urea gel, either before (*Sph*I −, lanes b, i, and j) or after (*Sph*I +, lanes c, k, and l) digestion with *Sph*I restriction endonuclease. To reveal all of the DNA products, electrophoresis was carried out for 2 h (left panel) and 4 h (middle panel), and both light (lanes j and k) and dark (lanes i and l) autoradiographic exposures are presented. mSV01 3′-^{32}P insert DNA provided the appropriate sequence standards for DNA primase products not digested with *Sph*I (lanes a, e, f, g, and h). A 3′-^{32}P-labeled DNA fragment with the same polarity as DNA primase products but terminally labeled at the *Sph*I site provided the appropriate sequence standards for DNA primase products digested with *Sph*I (lanes d, m, n, o, and p). Sequence tracks (A, G, T, and C) and genomic nucleotide positions (numbers on vertical axis) are described in the legend to Fig. 3. Identification of the 5′-terminal nucleotide in G$\overset{*}{p}$ppN-RNA-DNA chains was carried out with the six major bands in lane i as described for Fig. 3 (Nuclease P1, lanes q through v). GpppU, GpppC, GpppA, and GpppG indicate the positions of cap core standards.

events occurred more frequently at low GTP concentrations than at high GTP concentrations, and two ATP initiation events occurred more frequently at high GTP concentrations. These 5′ ends were identified both by mapping their locations on the genome and by 5′-terminal nucleotide analysis of individual DNA bands. Thus, changing the ATP/GTP ratio in the reaction mixture dramatically altered the selection and utilization of initiation sites. Since similar increases in the concentration of either CTP or UTP did not alter the 5′ ribonucleotide composition of RNA primers (Yamaguchi et al., in press), initiation sites were not mapped under these conditions.

Sequence of an individual RNA primer. To confirm the accuracy of DNA primase-DNA polymerase α initiation sites determined in the preceding experiments, the sequence of the most prominent RNA primer synthesized on mSV02 insert DNA (Fig. 5, lane i, position 58) was partially determined (Fig. 7, right panel). This band was excised and treated with either RNase T1, which cuts RNA preferentially on the 3′ side of guanosine residues, or RNase PhyM, which cuts preferentially on the 3′ side of adenosine or uridine residues (8, 9). The randomly cleaved products from partial alkaline (Fig. 7, lane A) hydrolysis revealed the position of cytidine residues and the total number of nucleotides present (8). The sequence for this RNA primer was 5′-G(A/U)(A/U)C(A/U)GG, which was in excellent agreement with the sequence predicted from its DNA template initiation site, 5′-GAACUGG. Thus, the nucleotide locations shown in Fig. 4 represent true in vitro RNA primer initiation sites.

Sizes of individual RNA primers. Analysis of the total population of RNA primers synthesized on different DNA templates showed that although the average size was typically 6 to 8 bases, the size distribution depended on the template sequence (Fig. 2; Table 2). These data suggested that the size of individual RNA primers may vary with the

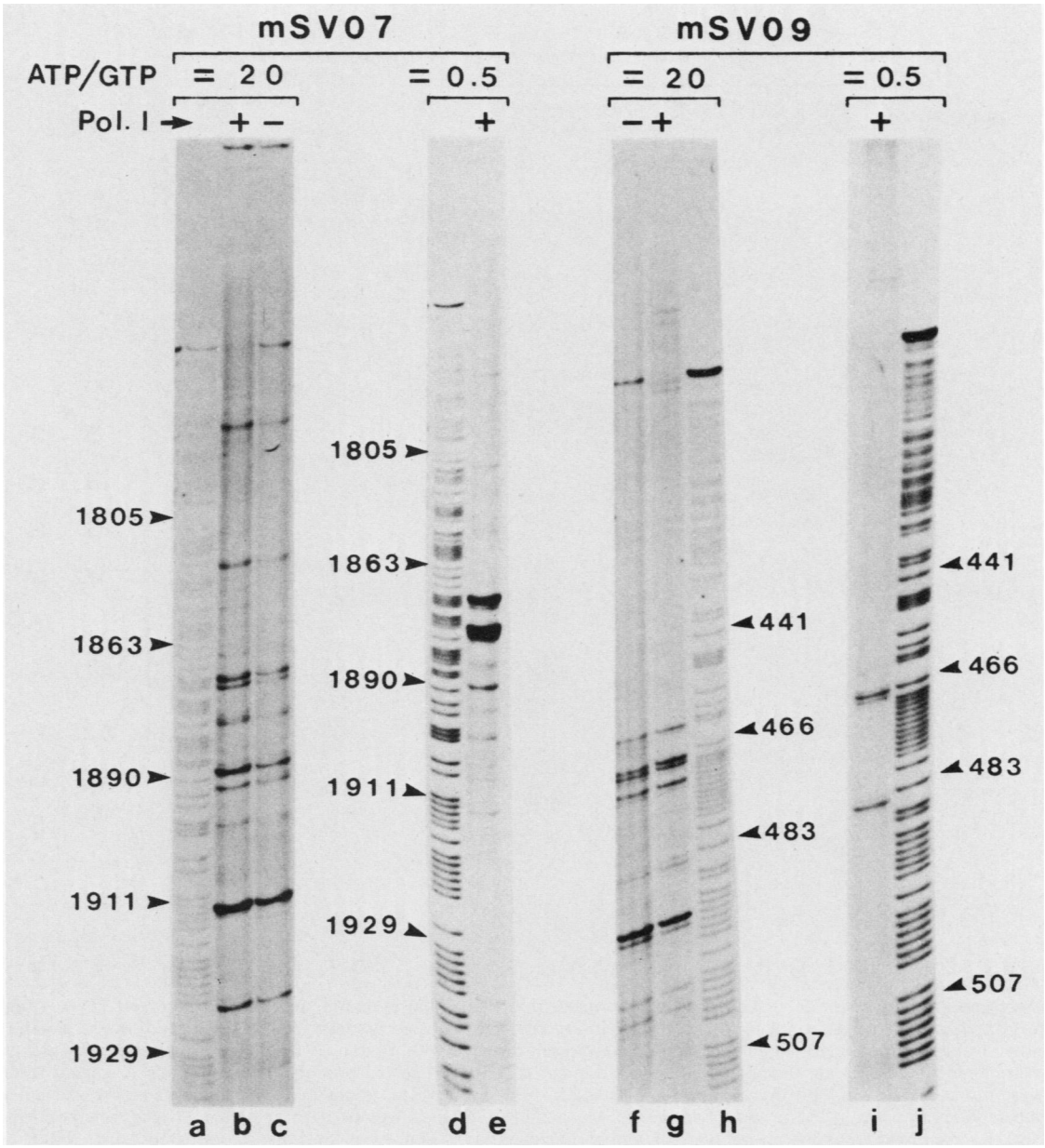

FIG. 6. Mapping the locations of 5′ ends of RNA primers synthesized at high and low ATP/GTP ratios on mSV07 or mSV09. (p)ppRNA-DNA chains were synthesized on single-stranded, linear mSV07 (left panel) or mSV09 (right panel) insert DNA either under standard assay conditions (4 mM ATP, 0.2 mM each of GTP, CTP, and UTP; ATP/GTP = 20 lanes) or in the presence of a high concentration of GTP (2 mM ATP, 4 mM GTP, 0.2 mM CTP, 0.2 mM UTP; ATP/GTP = 0.5 lanes). One portion of the sample was then treated with *E. coli* DNA polymerase I in the presence of all four dNTPs (*PolI* + lanes) to ensure that nascent chains were extended to the end of the template. Both the *PolI* treated (lanes b, e, g, and i) and untreated (lanes c and f) samples were then labeled in the capping reaction and fractionated by electrophoresis in 8% polyacrylamide–8 M urea gels. mSV08 (left panel) or mSV10 (right panel) 3′-^{32}P-labeled insert DNA was sequenced concurrently (lanes a, d, h, and j). Only the A track is shown (cleavage at both purines). Numbers on the vertical axis are nucleotide locations in SV40 wt800 (13).

template site used for initiation. Therefore, eight individual nascent capped-RNA-DNA chains that had been synthesized on mSV02 insert DNA were separated by gel electrophoresis, excised, and eluted from the gel. The size of their RNA primers was determined by digestion with T4 exo followed by gel electrophoresis (Fig. 7, T4 exonuclease + lanes). T4 exo conditions were such that about 80% of the primers contained a single 3′-dNMP, 7% were intact RNA primers alone, and 13% were RNA primers whose 3′ ends had been partially degraded (Yamaguchi et al., in press). Therefore, the structure of the major RNA chains released by T4 exo was $\overset{*}{G}pppN(pN)_npdN$. At least three examples of RNA primers of unique length were observed (Fig. 7, lanes g through i); however, other sites appeared to produce from two to three major products (lanes j through n). After correcting for the cap structure and 3′-dNMP (3.5 bases), most of the primers contained 5 to 8 ribonucleotides. However, one contained only 2 ribonucleotides (Fig. 7, lane i). Since the 5′ end of each RNA-DNA chain originated at a single, unique base in the template, variation in primer

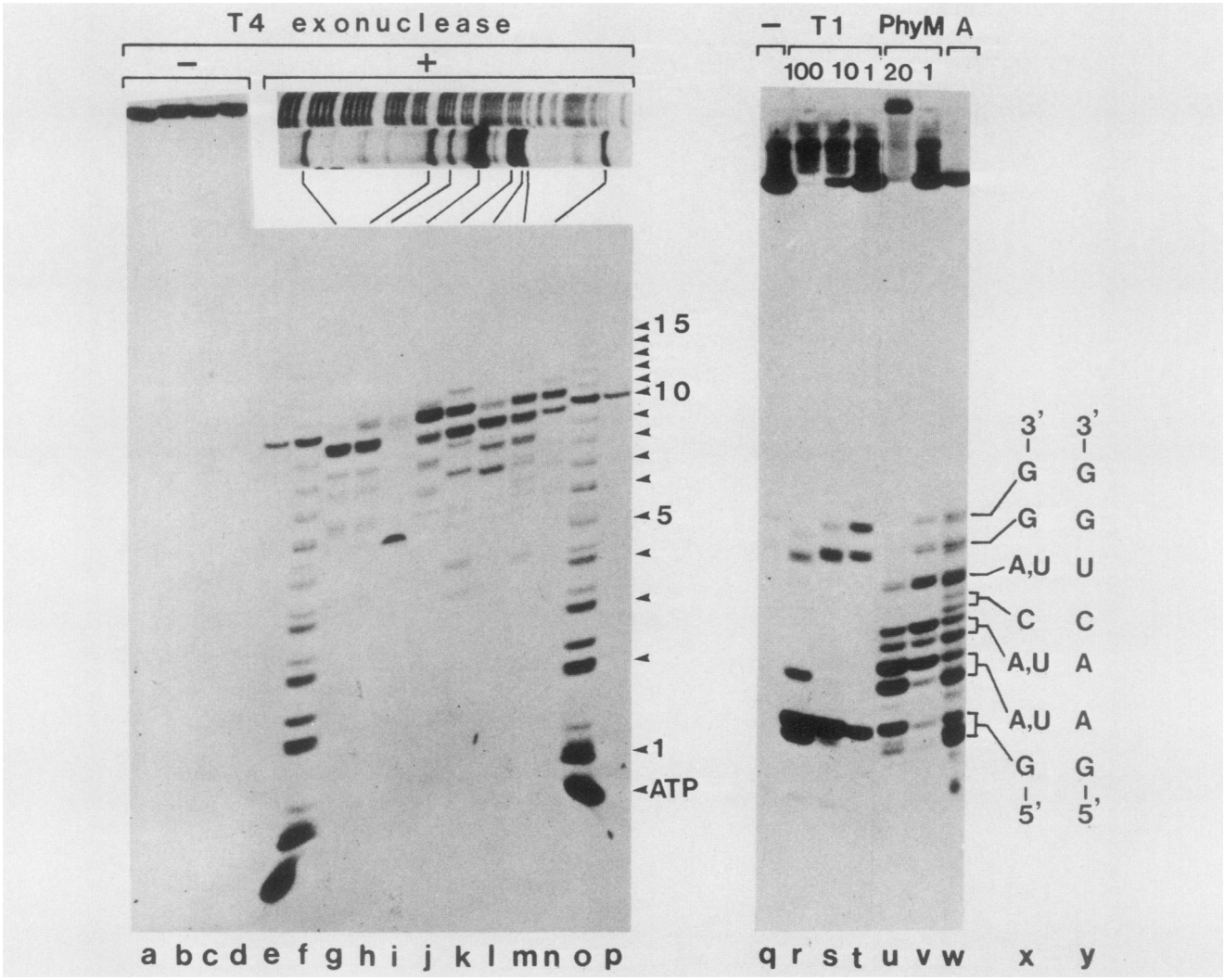

FIG. 7. Size and partial nucleotide sequence of RNA primers from individual initiation sites on mSV02 insert DNA. (p)ppRNA-DNA chains were synthesized on single-stranded linear mSV02 insert DNA and labeled as described in the legend to Fig. 5. The eight major products shown in Fig. 5 lane i were eluted separately from the gel, digested with T4 exo (+ lanes), and fractionated by electrophoresis in a 22% polyacrylamide–8 M urea gel (left panel). Lanes a through d show undigested products corresponding to lanes j through m. Size standards were run in parallel: 5′-^{32}P-$(pA)_{10}$ (lanes e and p) and 5′-^{32}P-labeled $p(Ap)_n$ (lanes f and o). Numbers on the vertical axis indicate the size in nucleotides of $p(Ap)_n$. The product whose size was analyzed in lane j was partially sequenced (right panel) with RNase T1 (G preference [8]), RNase PhyM (A and U preference [9]), and partial alkaline (A) hydrolysis (8). Lane q, untreated; lane r, 100 U of RNase T1, 55°C, 15 min; lane s, 10 U of RNase T1; lane t, 1 U of RNase T1; lane u, 20 U of RNase PhyM, 55°C, 15 min; lane v, 1 U of RNase PhyM; lane w, 50 mM sodium citrate (pH 9.0), 90°C, 15 min; lane x, nucleotide sequence deduced from this experiment; lane y, nucleotide sequence expected from the mapping experiment in Fig. 5.

length must depend on the sequence of the initiation site. The sequences of these initiation sites (Fig. 4) and the lengths of the most prominent RNA primers synthesized at each site are indicated.

DNA sequence encoding initiation sites for RNA-primed DNA. The average frequency of initiation events (5′ ends of RNA primers) under standard reaction conditions was one per 16 nucleotides of template, although this value varied two- to threefold among different sequences (mSV01, 1/35; mSV02, 1/14; mSV07, 1/13; mSV09, 1/16). The 5′ termini of each of the 56 RNA primers (Fig. 4) were aligned so that the DNA sequence in and around their respective templates could be compared. The stronger half of these sites was considered as one group, and the weaker half was considered as a second group. The frequency of pyrimidines at each nucleotide position in the template relative to the average pyrimidine composition in the entire 1,102 nucleotides examined revealed that DNA primase-DNA polymerase α had a distinct preference for initiating synthesis in the center of pyrimidine-rich sequences (Fig. 8A). In our standard reaction conditions (ATP/GTP ratio, 20), the stronger DNA template sites were 3′-$(Py)_8$CTTT$(Py)_4$ (Fig. 8B) whereas the weaker sites were 3′-$(Py)_7$PuCC(Py)$_5$ (Fig. 8C). The underlined base was complementary to the first ribonucleotide in the primer (Fig. 8). Thus, at least two classes of initiation sites existed, one that promoted ATP starts and one that promoted GTP starts. Both groups showed a preference for 3′-CPy at the actual site for RNA synthesis.

When the ratio of ATP to GTP was high, initiation sites

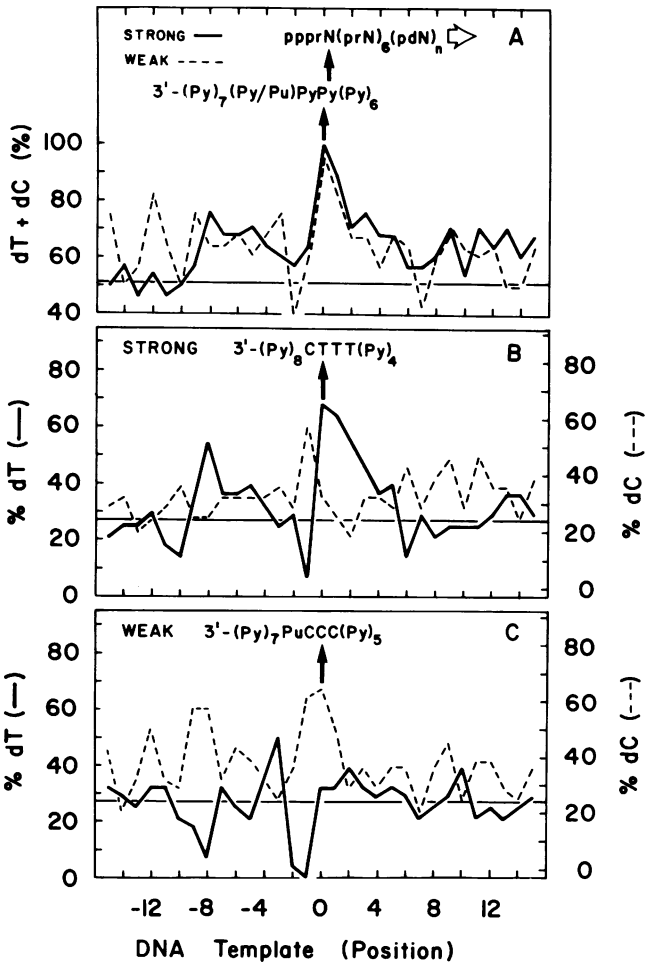

FIG. 8. Consensus DNA template sequences that initiate synthesis of RNA primers by DNA primase-DNA polymerase α. The 56 initiation sites identified for RNA primer synthesis under standard assay conditions (ATP/GTP ratio, 20) were divided into two groups, the stronger half and the weaker half. In each group, 5′ ends of RNA primers (DNA template position 0) were aligned to allow comparison of the DNA template sequences both upstream and downstream from this nucleotide position. The fraction (%) of pyrimidines at each template position for the strong sites (solid line) and weak sites (broken line) was calculated, as well as the average pyrimidine composition (51%; horizontal line) for the 1,102 bases analyzed (A). The frequency of occurrence (percentage) for dT (solid line) and dC (broken line) at each template position was calculated for the strong sites (B) and weak sites (C) along with the average fraction (horizontal line) of dT (27%) and dC (24%) in the 1,102 bases of template.

containing 3′-CTTT were favored. Nevertheless, the 21 sites on mSV02 (Fig. 4), which shared a consensus of 3′-PyCCCPy (data not shown), were utilized efficiently under these conditions (Table 1). Therefore, selection of initiation sites was influenced strongly by template sequence. Conversely, the ATP/GTP ratio also affected site selection. When the ATP/GTP ratio was 0.5, sites containing 3′-CCC were favored. The 21 sites used on mSV07 and mSV09 under these conditions (Fig. 4) were analyzed as described for Fig. 8 (data not shown) and shown to contain the consensus sequence, 3′-CCCCPy. Thus, both template sequence and ribonucleotide concentration determined which initiation sites were used by DNA primase-DNA polymerase α.

DISCUSSION

Parameters that affect selection of DNA primase initiation sites. (i) Template sequence. The experiments presented in this paper revealed precise nucleotide locations of initiation sites for CV-1 cell DNA primase-DNA polymerase α on several SV40 DNA sequences in vitro. Initiation events were found, on average, once every 16 bases, although the actual frequency varied from 1/13 to 1/35 among different DNA template sequences and some sites were highly preferred over others. Since the same sites observed on linear mSV01 insert DNA template were also observed on circular mSV01 DNA, site recognition was a consequence of the sequence and not the linear structure of the template. The sites shared a consensus sequence of either 3′-CTTT or 3′-CCC centered within a pyrimidine-rich sequence. The actual size of these regions was perhaps as long as 25 residues (Fig. 8A). Since several pyrimidine-rich sequences of less than 7 nucleotides did not function as initiation sites, the minimum recognition sequence for DNA primase-DNA polymerase α may be 7 bases. Initiation events tended to be clustered in certain regions of the template, suggesting that the predominant selection was for pyrimidine-rich areas in which RNA primer initiation could occur at one of several possible nucleotides. Thus, only deoxypyrimidine homopolymers were effective templates; deoxypurine and ribohomopolymers were inactive (Table 1). Consistent with this observation, natural DNA sequences rich in purines or with purines interspersed equally among pyrimidines were free of initiation sites (e.g., nucleotides 5246 to 161 in mSV01 and nucleotides 5106 to 10 in mSV02; Fig. 4). Conversely, long stretches of pyrimidine CCCGCC repeats in the mSV02 insert DNA template strongly promoted initiation of primer synthesis with GTP, even at low concentrations of GTP. Therefore, template sequence was a key parameter in selecting initiation sites.

Similar results have been reported for mSV01 and mSV02 insert DNA sequences in experiments with purified mouse DNA primase separated from DNA polymerase α (28). Although the sites were not mapped precisely, they appear to be the same as ones recognized by CV-1 DNA primase-DNA polymerase α, suggesting that site selection is carried out by the DNA primase component alone.

(ii) ATP/GTP ratio. Preference for specific initiation sites and the extent of their utilization were modulated by the ratio of ATP to GTP in the reaction mixture. This was apparent with homopolymer as well as with natural DNA templates. Poly(dC) was utilized more efficiently if GTP and dGTP were the only nucleotides present. When natural DNA templates were used in the presence of an ATP/GTP ratio of 20, about 80% of the RNA primers began with ATP and 20% began with GTP. When the ATP/GTP ratio was reduced to 0.5 by increasing the GTP concentration from 0.2 to 4 mM, several changes occurred. First, 40% of the original start sites were not utilized, all but one of which had initiated synthesis with ATP. Second, 24% of the start sites used at high GTP concentrations were new sites that had not been detected at low GTP concentrations. All but one of these new sites initiated synthesis with GTP. Third, the relative frequency of initiation events at sites using GTP was substantially increased, whereas the frequency of initiation events at sites using ATP was correspondingly decreased. Together, the data showed that high ATP concentrations promoted initiation of RNA primer synthesis in vitro at pyrimidine-rich sites containing 3′-CTTT whereas high GTP concentrations promoted initiation at pyrimidine-rich sites containing 3′-CCC (Fig. 4).

(iii) Template structure. The influence of secondary structures (i.e., double-stranded DNA hairpins) in the template was demonstrated by Tseng and Ahlem (28), who showed that a 6-bp deletion in the 27-bp palindromic sequence located within the SV40 *ori* (Fig. 4) reduced the frequency of initiation at sites adjacent to the palindrome in mSV01 by 80%. This palindromic sequence has been shown to form a hairpin when present as a single-stranded DNA template under the conditions employed in DNA polymerase assays (D. T. Weaver and M. L. DePamphilis, J. Mol. Biol., in press). It is also possible that hairpin formation prevents initiation within the palindromic sequence itself, since no initiation events were detected within the 27-base palindromic sequence in mSV01 or mSV02 (Fig. 4). Folding of the DNA template also appears to affect the selection of initiation sites for *E. coli* DNA primase (27).

Parameters that affect RNA primer length. RNA primers containing a 5′ di- or triphosphate consisted, on average, of six to eight ribonucleotides (Table 2). The relative number of short (p)ppRNA primers (1 to 4 ribonucleotides) varied from 14% on poly(dT) to 73% on poly(dC), with the values for natural DNA templates ranging from 19 to 49%. The size of RNA primers initiated at individual sites varied from 1 to 9 ribonucleotides (Fig. 2 and 7). Even RNA primers initiated at the same nucleotide location could vary in length from 1 to 3 bases (Fig. 7). This phenomenon was also observed with mouse DNA primase (28) and *E. coli* replication proteins (19). Thus, the size of RNA primers synthesized in vitro was not unique but was modulated by template sequence.

The size of RNA primers may be determined by the secondary structure of the RNA:DNA duplex formed. The average length of an RNA primer represents one-half to three-fourths of one helical turn in an RNA:DNA duplex. This length may be optimal for DNA polymerase α to bind to the primer-template and initiate DNA synthesis. However, since duplex polynucleotide structures are highly sensitive to base composition and environmental conditions (14, 32, 36), it is reasonable that variations in template sequence could produce variations in the secondary structure which could affect the binding.

Another factor may be template structure. The largest fraction of small RNA primers synthesized on a natural DNA template occurred with mSV01 (Table 2), indicating that the 27-bp hairpin immediately upstream may influence RNA primer size.

Finally, dC residues in the template may promote termination of primer synthesis. Sequence analysis of RNA primer initiation sites revealed a preference for dC at RNA-p-DNA junctions (Fig. 8), consistent with a previous observation that ribonucleotides at RNA-p-DNA junctions synthesized under these conditions were guanosine rich (Yamaguchi et al., in press). Furthermore, the average size of RNA primers synthesized on poly(dC) was significantly shorter than that of RNA primers from other templates.

Comparison between in vitro and in vivo initiation sites on SV40 DNA. The DNA primase-DNA polymerase α used in this study was purified from CV-1 cells, an established cell line of African green monkey kidney cells, the permissive host for SV40 replication. Therefore, it was appropriate to compare the initiation sites selected by this enzyme in vitro with the sites selected by the cellular DNA replication machinery in vivo. The locations of 5′ ends of RNA primers synthesized in vivo at the SV40 *ori* region have been previously determined (12) and are indicated by asterisks in Fig. 4. Although the 5′ ends of these RNA primers were not labeled in the capping reaction, subsequent analysis confirmed that 80% of the RNA primers bore the original ATP or GTP 5′ nucleotide (13; unpublished data). However, none of these in vivo initiation sites were used in vitro, even though the DNA templates used in vitro were similar in size to the proposed Okazaki fragment initiation zone at replication forks (6, 7). In fact, the templates representing the forward arms of replication forks, mSV07 and mSV09, were more active with DNA primase-DNA polymerase α than were the corresponding retrograde templates, mSV08 and mSV10 (Table 1). Furthermore, in vivo initiation of RNA-primed DNA synthesis occurs preferentially at 3′-purine-pyrimidine sites located within random sequences (13) instead of at the 3′-pyrimidine-pyrimidine sites within pyrimidine-rich regions observed in vitro. Initiation sites in vivo are found, on average, once every 7 bases, at least twice the frequency that was observed in vitro. Furthermore, it does not appear that the 27-base palindrome in *ori* forms a hairpin structure in vivo (Weaver and DePamphilis, in press), although this sequence is clearly important for initiation of DNA replication (6). Thus, the biological significance, if any, of the cluster of in vitro initiation events within *ori* on mSV01 is not clear. The simplest explanation for the difference in site selection between in vitro and in vivo conditions is a need for other proteins that may associate either with the DNA primase-DNA polymerase α complex or with the DNA template (2, 3, 22–24, 34).

ACKNOWLEDGMENTS

We thank Eriko Yamaguchi for her help in the analysis of DNA sequence data.

This research was supported by the National Cancer Institute and the American Cancer Society.

LITERATURE CITED

1. **Anderson, S., and M. L. DePamphilis.** 1979. Metabolism of Okazaki fragments during SV40 DNA replication. J. Biol. Chem. **254**:11495–11504.
2. **Arai, K., and A. Kornberg.** 1981. Mechanism of dnaB protein action; II. ATP hydrolysis by dnaB protein dependent on single- or double-stranded DNA. J. Biol. Chem. **256**:5253–5259.
3. **Arai, K., and A. Kornberg.** 1981. Mechanism of dnaB protein action; IV. General priming of DNA replication by dnaB protein and primase compared with polymerase. J. Biol. Chem. **256**:5267–5272.
4. **Bock, R. M.** 1967. Controlled partial hydrolysis of RNA. Methods Enzymol. **12**:218–221.
5. **DePamphilis, M. L., and P. M. Wassarman.** 1980. Replication of eukaryotic chromosomes: a close-up of the replication fork. Annu. Rev. Biochem. **49**:627–666.
6. **DePamphilis, M. L., and P. M. Wassarman.** 1982. Organization and replication of papovavirus DNA, p. 37–114. *In* A. Kaplan (ed.), Organization and replication of viral DNA. CRC Press, Boca Raton, Fla.
7. **DePamphilis, M. L., L. E. Chalifour, M. F. Charette, M. E. Cusick, R. T. Hay, E. A. Hendrickson, C. G. Pritchard, L. C. Tack, P. M. Wassarman, D. T. Weaver, and D. O. Wirak.** 1983. Papovavirus chromosomes as a model for mammalian DNA replication. UCLA Symp. Mol. Cell. Biol. New Ser. **10**:423–447.
8. **Donis-Keller, H., A. M. Maxam, and W. Gilbert.** 1977. Mapping adenines, guanines, and pyrimidines in RNA. Nucleic Acids Res. **4**:2527–2538.
9. **Donis-Keller, H.** 1980. PhyM: an RNase activity specific for U and A residues useful in RNA sequence analysis. Nucleic Acids Res. **8**:3133–3142.
10. **Fisher, P. A., and D. Korn.** 1979. Enzymological characterization of KB cell DNA polymerase-α. II. Specificity of the protein-nucleic acid interaction. J. Biol. Chem. **254**:11040–11046.
11. **Gronostajski, R. M., J. Field, and J. Hurwitz.** 1984. Purification

of a primase activity associated with DNA polymerase α from HeLa cells. J. Biol. Chem. **259**:9479–9486.
12. **Hay, R. T., and M. L. DePamphilis.** 1982. Initiation of SV40 DNA replication *in vivo*: location and structure of 5'-ends of DNA synthesized in the *ori* region. Cell **28**:767–779.
13. **Hay, R. T., E. A. Hendrickson, and M. L. DePamphilis.** 1984. Sequence specificity for the initiation of RNA-primed SV40 DNA synthesis *in vivo*. J. Mol. Biol. **175**:131–157.
14. **Hogan, M., J. LeGrange, and B. Austin.** 1983. Dependence of DNA helix flexibility on base composition. Nature (London) **304**:752–754.
15. **Kitani, T., K. Y. Yoda, and T. Okazaki.** 1984. Discontinuous DNA replication of *Drosophila melanogaster* is primed by octaribonucleotide primer. Mol. Cell. Biol. **4**:1591–1596.
16. **Kurosawa, Y., and T. Okazaki.** 1979. Structure of the RNA portion of the RNA-linked DNA pieces in bacteriophage T4-infected *Escherichia coli* cells. J. Mol. Biol. **135**:841–861.
17. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
18. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning laboratory manual. p. 113–116, 156, 166, 438. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
19. **Ogawa, T., K. Arai, and T. Okazaki.** 1983. Site selection and structure of DNA-linked RNA primers synthesized by the primosome in phage φX174 DNA replication *in vitro*. J. Biol. Chem. **258**:13353–13358.
20. **Patton, R. J., and C.-B. Chae.** 1982. A method for isolation of a large amount of a single-stranded DNA fragment. Anal. Biochem. **126**:231–234.
21. **Peacock, A. C., and C. W. Dingman.** 1967. Resolution of multiple ribonucleic acid species by polyacrylamide gel electrophoresis. Biochemistry **6**:1818–1827.
22. **Pritchard, C. G., and M. L. DePamphilis.** 1983. Preparation of DNA polymerase α-C_1C_2 by reconstituting DNA polymerase α with its specific stimulatory cofactors, C_1C_2. J. Biol. Chem. **258**:9801–9809.
23. **Pritchard, C. G., D. T. Weaver, E. F. Baril, and M. L. DePamphilis.** 1983. DNA polymerase α cofactors C_1C_2 function as primer recognition proteins. J. Biol. Chem. **258**:9810–9819.
24. **Reha-Krantz, L. J., and J. Hurwitz.** 1978. The dnaB gene product of *Escherichia coli*: II. Single-stranded DNA-dependent ribonucleoside triphosphatase activity. J. Biol. Chem. **253**: 4051–4057.
25. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**:107–110.
26. **Shuman, S., and J. Hurwitz.** 1981. Mechanism of mRNA capping by vaccinia virus guanylyltransferase: characterization of an enzyme-guanylate intermediate. Proc. Natl. Acad. Sci. U.S.A. **78**:187–191.
27. **Sims, J., and E. M. Benz, Jr.** 1980. Initiation of DNA replication by the E. *coli* dna G protein: evidence that tertiary structure is involved. Proc. Natl. Acad. Sci. U.S.A. **77**:900–904.
28. **Tseng, B. Y., and C. N. Ahlem.** 1984. Mouse primase initiation sites in the origin region of simian virus 40. Proc. Natl. Acad. Sci. U.S.A. **81**:2342–2346.
29. **Tu, C., and S. Cohen.** 1980. 3'-End labeling of DNA with [α-^{32}P]cordycepin-5'-triphosphate. Gene **10**:177–183.
30. **Venkatesan, S., A. Gershowitz, and B. Moss.** 1980. Modification of the 5' end of mRNA; association of RNA triphosphatase with the RNA guanylyltransferase - RNA(guanine - 7 -)methyltransferase complex from vaccinia virus. J. Biol. Chem. **255**:903–908.
31. **Weaver, D. T., and M. L. DePamphilis.** 1984. The role of palindromic and nonpalindromic sequences in arresting DNA synthesis *in vitro* and *in vivo*. J. Mol. Biol. **180**:961–986.
32. **Wells, R. D., R. W. Blakesley, S. C. Hardies, G. T. Horn, J. E. Larson, E. Selsing, J. F. Burd, H. W. Chan, J. B. Dodgson, F. K. Jensen, I. F. Nes, and R. M. Wartell.** 1977. The role of DNA structure in genetic regulation. Crit. Rev. Biochem. **4**:305–340.
33. **Wilson, S. H., A. Matsukage, E. W. Bohn, Y. C. Chen, and M. Sivarajan.** 1977. Polynucleotide recognition by DNA polymerase α. Nucleic Acids Res. **4**:3981–3996.
34. **Yagura, T., T. Kozu, and T. Seno.** 1983. Mechanism of stimulation by a specific protein factor of *de novo* DNA synthesis by mouse DNA replicase with fd phage single-stranded circular DNA. Nucleic Acids Res. **11**:6369–6380.
35. **Yoda, K., and T. Okazaki.** 1983. Primer RNA for DNA synthesis on single-stranded DNA template in a cell free system from *Drosophila melanogaster* embryos. Nucleic Acids Res. **11**: 3433–3450.
36. **Zimmerman, S. B.** 1982. The three-dimensional structure of DNA. Annu. Rev. Biochem. **51**:395–427.