Resolution of DNA molecules greater than 5 megabases by contour-clamped homogeneous electric fields

Douglas Vollrath and Ronald W. Davis

Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA

Received June 8, 1987; Revised and Accepted August 31, 1987

ABSTRACT

Excellent resolution of chromosomal DNA molecules from Saccharomyces cerevisiae, Candida albicans and Schizosaccharomyces pombe has been obtained using alternating contour-clamped homogeneous electric field (CHEF) gel electrophoresis. The largest of these molecules is greater than 5 Mb in size and is resolved after 130 hours in a 0.6% agarose gel at a field strength of 1.3 V/cm and a switching interval of 1 hour. Separation of concatamers of phage lambda DNA reveals four regions of resolution in alternating CHEF gel electrophoresis. There are two regions of good resolution in which mobility approximates a linear function of molecular weight. These are separated by a region of lower resolution and bounded at high molecular weights by a region of little or no resolution. The four regions are of practical and possibly theoretical importance.

INTRODUCTION

Alternating or pulsed field gel electrophoresis has been successfully applied to resolve whole chromosomal DNA molecules up to 2 mega-basepairs (Mb) in size from numerous single-celled eukaryotes [1-7]. Although current theory does not predict an upper limit to the size of molecules which can be separated by this technique [8], in practice, chromosomal DNA larger than 2 Mb has not been well resolved [1-4]. Extension of the technique to molecules of greater than 2 Mb would significantly expand the number of organisms to which it could be applied.

This paper addresses the problem of resolving molecules larger than 2 Mb. Two types of yeast have been used as sources of these large molecules: Candida albicans, a pathogenic yeast which contains a range of molecules starting at less than 2 Mb and extending to larger sizes [7]; and Schizosaccharomyces pombe, a fission yeast which has only three chromosomes, each of which should be greater than 2 Mb based on genetic mapping and genome size estimates [9,10].

These large DNA molecules have been resolved by employing a particular configuration of the alternating field technique known as contour-clamped homogeneous electric field (CHEF) gel electrophoresis [11]. It is shown here that a combination of low field strengths, long switching intervals and low agarose gel
concentrations can result in excellent resolution of molecules greater than 5 Mb in size.

**MATERIALS AND METHODS**

**Preparation of Chromosomal DNA**

*Saccharomyces cerevisiae* DNA samples were prepared as described [11]. *S. pombe* samples were prepared by growing strain 975 (h+) to stationary phase in YPD media (2% glucose, 2% bactopeptone, 1% yeast extract). The cells were washed twice with 50 mM sodium EDTA (pH 8.0) and resuspended in 0.5 ml LET buffer (0.5 M sodium EDTA, pH 8.0, 10 mM Tris, pH 7.5) [1] to which 50-100 µl of 100 mg/ml Novozyme (Novo Laboratories) was added. The cells were warmed to 37°C and added to 0.835 ml of 1% low melting point agarose (FMC) in 125 mM sodium EDTA (pH 8.0) which had been cooled to 37°C. The mixture was allowed to solidify on ice. The agarose plug was then immersed in 1.4 ml LET buffer plus 1% beta-mercaptoethanol and incubated at 37°C overnight. The LET buffer was replaced with 1.4 ml of NDS buffer (0.5 M sodium EDTA, pH 8.0, 10 mM tris, pH 7.5, 1% sarkosyl) [1] containing 2 mg/ml proteinase K and the sample was incubated overnight at 50°C. The sample was then washed multiple times at 50°C with 50 mM sodium EDTA (pH 8.0) before storing at 40°C.

*C. albicans* (ATCC strain 14053) was grown to saturation in 5 ml of YPD media. The cells were washed twice with 50 mM sodium EDTA (pH 8.0), resuspended in 1 ml of 50 mM sodium EDTA plus 1% beta-mercaptoethanol and incubated at 30°C for 45 minutes. The cells were then washed twice with 1 M sorbitol and resuspended in 1 ml of 1 M sorbitol, 50 mM sodium phosphate (pH 7.4) and 50 mM sodium EDTA (pH 8.0). A 20 µl aliquot of 20 mg/ml Zymolyase 5000 (Kirin Breweries) was added and the cells were incubated overnight at 37°C. The resulting spheroplasts were carefully washed twice with 1 M sorbitol, 250 mM sodium EDTA (pH 8.0), and resuspended in 1.8 ml of this buffer. The cell suspension was warmed to 37°C and 2.5 ml of 1% low melting point agarose in 125 mM sodium EDTA was added. The mixture was allowed to solidify on ice, immersed in 4.3 ml of NDS buffer containing 2 mg/ml proteinase K, and incubated overnight at 50°C. The sample was then washed multiple times at 50°C with 50 mM sodium EDTA (pH 8.0). All three types of yeast chromosomal DNA preparations were stable for over 12 months when stored in 50 mM sodium EDTA at 40°C.

Human genomic DNA was prepared from peripheral blood leukocytes. Two volumes of fresh, EDTA-treated peripheral blood was gently overlaid on one volume of Ficoll-Paque (Pharmacia) and centrifuged at 900 x g for 12 minutes at room temperature. The "buffy coat," containing mononuclear leukocytes, was drawn off.
with a pasteur pipet. The cells were washed twice with ice cold RSB (10 mM NaCl, 10 mM Tris, pH 7.5, 25 mM sodium EDTA, pH 8.0) [12], first with two volumes and then with one volume of buffer. The cells were counted in a hemocytometer and brought to a concentration of 5-7 x 10^7 cells/ml with RSB. The cell suspension was warmed to 37°C and immediately mixed with an equal volume of 1% low melting point agarose (also at 37°C) in RSB with 60 mg/ml proteinase K. The mixture was transferred into a teflon trough (14.5 cm x 0.6 cm x 0.5 cm) and allowed to solidify on ice. The agarose plug was then placed in a tube and completely immersed in an equal volume of 1% sodium dodecyl sulfate (SDS) in RSB and incubated at 50°C for 18-24 hours. The sample was washed extensively with 50 mM sodium EDTA (pH 8.0) at 50°C until no SDS precipitate was visible upon cooling to 4°C.

**Digestion of Human Chromosomal DNA**

A slice, with dimensions about equal to those of a gel well and containing about 5 μg of DNA, was cut from the agarose plug. The slice was placed in a 1.5 ml polypropylene tube with 1 ml of the appropriate restriction enzyme buffer and incubated at 37°C for 10-15 minutes to equilibrate the gel slice with the buffer. The buffer was then replaced. A total of three such incubations was performed, after which the slice was removed and blotted with a tissue. The slice was then placed in a 0.5 ml polypropylene tube containing 50 μl of restriction enzyme buffer. A two fold excess of restriction enzyme was added and the sample was incubated at the temperature recommended by the supplier (New England Biolabs) for 3 hours. This was repeated for a total incubation of 6 hours with a four-fold excess of enzyme. The reaction was terminated with 0.3 ml of 50 mM sodium EDTA. The sample could be stored for several days at 4°C with no deleterious effects.

**Lambda Ladders**

A total of 10 μg of bacteriophage lambda DNA (CI857S7) was added to 0.250 ml of 2% polyethylene glycol (M_r 8000), 2 mM ATP and 2 mM dithiothreitol. Three Weiss units of T4 DNA ligase (New England Biolabs) were added along with an equal volume of 1% low melting point agarose in 20 mM MgCl_2, 100 mM Tris (pH 7.5) which had been cooled to 37°C. The mixture was allowed to solidify on ice. The resulting plug was immersed in an equal volume of ligase buffer (1% PEG 8000, 1 mM ATP, 1mM DTT, 10mM MgCl_2, 50 mM Tris, pH 7.5) and incubated at room temperature for 18 to 24 hours. The sample was stored in 50 mM sodium EDTA (pH 8.0) at 4°C and remained stable for at least 3 months.

**Gel Electrophoresis**

CHEF gel electrophoresis was performed using a hexagonal array of electrodes as described [11], except that, for some gels, vertical rather than horizontal electrodes were employed. For all gels, the buffer was cooled to 90°C during electrophoresis and
the reorientation angle was 120°. Samples were either cut into blocks the size of a
gel well and inserted, or they were melted by heating to 65-70°C for 2-5 minutes and
loaded as a molten mixture using a 100 μl glass micropipet. In either case, the wells
were then sealed with 1% low melting point agarose. Melting the samples can, at
times, be more convenient. However, a larger proportion of DNA greater than 1 Mb
in size seems to enter the gel when the samples are not melted.

Transfer and Hybridization

After staining with ethidium bromide (0.5 μg/ml), the gel was irradiated for 1
minute with 254 nm ultraviolet light (UV Products) to nick the DNA. The gel was
then placed in 250 ml of denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 30
minutes at room temperature with gentle shaking followed by a 30 minute incubation
in 250 ml of neutralizing solution (0.5 M Tris, pH 7.5, 1.5 M NaCl) also at room
temperature with shaking. The DNA was then transferred to a Genatran 45 filter
(Plasco) using a reservoir of 500 ml of 10X SSPE [13] for 15-20 hours. The filter
was baked for 1-2 hours at 80°C under vacuum and then prehybridized for 6 hours at
42°C in 15 ml of 3X SSPE, 50% formamide, 100 mg/ml salmon sperm DNA, 0.1%
bovine serum albumin, 0.1% Ficoll (MW 400k), 0.1% polyvinyl pyrrolodine (MW
300k) and 1% SDS. Hybridization was carried out in 15 ml of the same solution with
5% dextran sulfate added for 24-36 hours at 42°C using 1-2 X 10^6 cpm/ml of random
hexamer primed probe [14]. The filter was washed once in 500 ml of 50%
formamide, 3X SSPE, 0.2% SDS at 42°C for 15 minutes and twice in 500 ml of
0.5X SSPE, 0.1% SDS at 60°C for 30 minutes each.

RESULTS
Resolution Characteristics of CHEF

To achieve separation of molecules greater than 2 Mb in size, an understanding
of the resolution characteristics of the CHEF system is essential. Excellent resolution
of S. cerevisiae chromosomal DNA can be obtained using CHEF (Fig. 1A).
However, these molecules are not uniformly distributed throughout the range of
resolution and their sizes are not known with precision. A more accurate picture of
the resolution characteristics can be obtained by electrophoresis of concatamers of
bacteriophage lambda. Figure 1B shows the result obtained when a "ladder" of
lambda concatamers is separated under conditions suitable for resolving most of the
S. cerevisiae molecules. Four regions of separation are apparent. From bottom to
top, they are region 1, an area of good resolution in which mobility appears to be a
linear function of molecular weight; region 2, a small area of reduced resolution in
the middle of the gel; region 3, an area of maximal resolution in which mobility is
again a linear function of molecular weight; and region 4, an area at the top of the gel
of little or no resolution.
Figure 1
Four regions of resolution in alternating CHEF gel electrophoresis. (A) Chromosomal DNA from S. cerevisiae strain YNN295 was electrophoresed for 24 hours at 160 volts using an 80 second switching interval. The distance between driving electrodes was 26.5 cm. (B) and (C) Concatamers of phage lambda DNA molecules were subjected to two separate electrophoresis runs in 1% agarose gels using a field strength of 210 volts and a switching interval of either 60 or 80 seconds. The distance between driving electrodes was 30 cm. The 60 second gel is shown in (B) and arrows to the right of the gel indicate the approximate extent of the four regions described in the text. Data from both runs were used to plot the mobility versus molecular size curves shown in (C). Mobility was calculated relative to the lambda monomer (48.5 kb) to facilitate comparison between the two gels. Arrows to the right of the plot indicate the approximate extent of the four regions for the 80 second curve.

Figure 1C shows the effect of increasing the switching interval from 60 seconds (used in Fig. 1B) to 80 seconds on this resolution profile. Increasing the switching interval displaces region 3 to higher molecular weights. However, the slope of the curve and the extent of the region remain largely constant. This relationship holds
**Figure 2**

*Four regions of resolution with different electrophoresis conditions and a variety of samples.*

(A) *S. cerevisiae* and lambda samples were electrophoresed for 53.5 hours in 1% agarose at 120 volts using a switching interval of 4 minutes and an electrode spacing of 30 cm. Band 11, containing chromosomes VII and XV, is about 1.2 Mb in size and is shown as a point of reference [19]. (B) Human genomic DNA restricted with the infrequently cutting enzymes BssH II, Mlu I, and Not I was electrophoresed for 22 hours in 1% agarose at 160 volts using a switching interval of 60 seconds and an electrode spacing of 26.5 cm. Arrows to the right of both gels indicate the approximate extent of the four regions described in the text.

Even when region 3 is shifted to higher molecular weights (Fig. 2A). Molecules differing by a single lambda monomer (48.5 kb) can be resolved in the 1.2-1.5 Mb range. In contrast, decreased resolution in region 1 is observed under these conditions. Thus, by properly positioning region 3, resolution of multi-megabase molecules which differ in length by only a few percent should be possible using CHEF.

Another noteworthy characteristic of region 3 is that DNA bands within it are
Figure 3
Effect of field strength and agarose gel concentration on the resolution characteristics of CHEF.

(A) and (B) S. cerevisiae and C. albicans samples were electrophoresed at 200 volts in a 0.7% agarose gel for 24 hours using a switching interval of (A) 90 seconds and (B) 120 seconds. (C) and (D) S. cerevisiae and C. albicans samples were electrophoresed at 100 volts in a 0.7% agarose gel for 44 hours using a switching interval of (C) 3.5 minutes and (D) 5 minutes. An S. cerevisiae rDNA [20] probe hybridizes to the lowest mobility C. albicans band in (D). (E) and (F) S. cerevisiae and C. albicans samples were electrophoresed at 200 volts using an 80 second switching interval and (E) a 1% agarose gel for 28 hours or (F) a 0.6% agarose gel for 24 hours. In all panels, the distance between driving electrodes was 26.5 cm.

more diffuse and thus appear fainter relative to bands in regions 1 and 2. This effect is well illustrated in genomic digests of human DNA. Figure 2B shows that all four regions are visible in such digests. Region 2 appears as an increase in staining intensity near the middle of the gel. Region 3 appears as a more faintly staining region between region 4 and region 2. These differences in intensity are not likely to be due to discontinuities in the distribution of fragment sizes in the samples since their presence is independent of both the restriction enzyme used and the switching interval.

Resolution of Multi-megabase Yeast Chromosomal DNA

The previous section establishes that CHEF gel electrophoresis displays a region of high resolution which, as with other alternating field techniques, can be shifted to higher molecular weights by increasing the switching interval. This knowledge was applied to the resolution of yeast chromosomal DNA of greater than 2 Mb in size.
Figure 4
Resolution of S. pombe chromosomal DNA molecules.
(A) S. pombe and S. cerevisiae samples were electrophoresed in a 0.6% agarose gel for 120 hours using a switching interval of 30 minutes and a field strength of 50 volts. (B) S. pombe, C. albicans and S. cerevisiae samples were electrophoresed in a 0.6% agarose gel for 130 hours using a switching interval of 60 minutes and a field strength of 40 volts. The distance between driving electrodes in both cases was 30 cm. Bands were identified by Southern blot using specific sequence probes for S. pombe chromosomes I (mei2.8) and II (mei3.2) [21,22]. S. cerevisiae rDNA [20] cross-hybridizes to S. pombe rDNA and is thus a probe for both S. pombe chromosome III and S. cerevisiae chromosome XII. CEN4 [23] was used to identify S. cerevisiae chromosome IV.

C. albicans is thought to possess larger chromosomal DNA than S. cerevisiae [7]. Figure 3A shows a CHEF gel run under conditions which optimize the resolution of the largest S. cerevisiae chromosomes and the smallest chromosomes of C. albicans. The intense ethidium bromide staining in the upper portion of the C. albicans lanes indicates that these large molecules are in region 4 of the gel. When the switching interval is increased from 80 seconds to 120 seconds to shift region 3 to a higher molecular weight, the larger chromosomal bands begin to smear in both samples (Fig. 3B).

The smearing effect observed for the largest DNA molecules could be a result of
requiring them to move under too strong an electric field. This hypothesis can be tested by resolving these same molecules at a lower voltage. However, in CHEF, lowering the voltage displaces region 3 to a lower molecular weight range. So, in order to resolve the same sized molecules at a lower voltage, the switching interval must be increased. In fact, a given range of molecular weights can be resolved using virtually any combination of field strength and switching interval as long as the product of these two values remains approximately constant (data not shown).

In Figure 3C, the field strength has been decreased to 100 volts and the switching interval increased to 3.5 minutes. The smearing effect is largely eliminated. When the switching interval is increased further to 5 minutes and the voltage held constant, good resolution of the larger \textit{C. albicans} chromosomal DNA is obtained (Fig. 3D). Thus, large molecules which smear at one field strength are better behaved when the field strength is decreased. With this lesson in mind, resolution of the \textit{S. pombe} chromosomal DNA was attempted.

The field strength was further decreased to 50 volts and the switching interval increased by a factor of six to 30 minutes. The agarose concentration was also decreased from 0.7\% to 0.6\%. Lowering the agarose concentration causes region 3 to shift to a higher molecular weight when the voltage and switching interval are held constant (Fig. 3E and F). However, the chromosomal bands tend to be more diffuse at lower agarose concentrations, limiting the usefulness of very dilute gels.

Figure 4A shows the results of this initial attempt to separate the \textit{S. pombe} chromosomes. Good resolution of \textit{S. cerevisiae} chromosomes IV and XII, which differ in size by about 1 Mb, is obtained. These chromosomes are also well resolved from the bulk of DNA in the \textit{S. pombe} sample. However, the sample itself is still in region 4 of the gel.

To better resolve the \textit{S. pombe} chromosomes, the voltage was decreased to 40 volts and the switching interval doubled to one hour. All other parameters were held constant. Figure 4B shows that good resolution is now obtained. Three DNA bands are evident. Probes specific for each chromosome identify these bands as the three individual chromosomal DNA molecules of \textit{S. pombe} (data not shown). These assignments are consistent with those of other workers [15].

**DISCUSSION**

The resolution profile of CHEF gel electrophoresis displays four distinct regions: two regions of linear resolution differing in slope, each bounded above by a region of decreased resolution. This biphasic resolution profile is suggestive of the superposition of two different resolving mechanisms, one acting primarily in region 1 and another predominating in region 3. The mechanism acting in region 3 most likely results from the exquisite dependence on molecular weight of the time required
for a molecule to reorient when the direction of the electric field is changed, as proposed by Schwartz and Cantor [1]. The sensitivity of the position of region 3 to the switching interval supports this notion. Molecules in this region probably spend a large amount of time reorienting and only a small amount traveling through the gel.

In region 1, by contrast, the molecules probably reorient rapidly relative to the switching interval, so little resolution is obtained on the basis of reorientation time. Instead, alternating the fields may prevent the molecules from assuming a conformation in which mobility is independent of molecular weight, as when large DNA is subjected to conventional agarose gel electrophoresis [16]. This would allow sieving through the gel matrix to occur. In this model, the time required for a molecule to assume the non-sieving conformation would have to be long relative to the switching interval. The fact that the slope of region 1 approaches zero (i.e. no resolution) as the switching interval increases (see, for example, Fig. 2A) supports this idea.

The specific range of molecular weights present in region 3 of any given gel is a complex function of field strength, switching interval, agarose concentration and temperature. The magnitude of the force applied to the molecule by the electric field, the duration of that force, the resistance to reorientation caused by the agarose matrix and the thermal energy of the DNA all contribute to the reorientation time. In addition, there is an upper limit to the strength of the electric field that can be used to resolve a given large molecule when the agarose matrix and temperature are held constant. Above this limit, the molecules are not well behaved. For example, at field strengths normally used to resolve molecules less than 2 Mb, the largest S. pombe molecules do not even enter the gel. Presumably, these large molecules become trapped when required to move too rapidly through the gel matrix.

The low field strengths necessary to prevent trapping of multi-megabase DNA molecules require electrophoresis times of several days to a week. Thus, extension of CHEF gel electrophoresis to even larger molecules is limited by the extremely long run times involved. Perhaps this problem could be circumvented by the development of new gel matrices which reduce trapping. Alternatively, condensation of the molecules in a way that preserves the molecular weight dependence of reorientation time may prove fruitful.

By recognizing the importance of field strength in CHEF gel electrophoresis, it has been possible to resolve the three chromosomal DNA molecules of S. pombe which constitute the organism's 15 Mb haploid genome [10]. This has also recently been achieved using the pulsed field gradient system [15]. The ratio of the sizes of the two largest S. pombe molecules (chromosomes II and I) is estimated to be 1:1.4 from the genetic map and measurements of fluorescently stained condensed chromosomes [9,17]. Chromosome III, which contains over 1 Mb of the rDNA
cluster, is not adequately assessed by either of the above techniques because the rDNA exhibits unique condensation and recombination characteristics compared to the rest of the genome [17,18]. An independent estimate of the size of chromosome III is required.

Given the resolution profile in CHEF gel electrophoresis, extrapolations of molecular weight beyond the range of size markers must be done with caution. It is apparent, however, that chromosome III has a lower mobility in CHEF (Fig. 4B) than S. cerevisiae chromosome XII which is about 2.5 Mb in size (G. Carle and M. Olson, personal communication). Assuming chromosome III is about 3 Mb in size, chromosomes II and I should be 5 and 7 Mb respectively. If these estimates are reasonably accurate, then resolution of the chromosomes of many other organisms should be possible using this technique.

ACKNOWLEDGEMENTS
This work was supported by the Medical Scientist Training Program at Stanford (D.V.), Public Health Service grants GM21891 and AG02908, and the MacArthur Foundation (R.W.D.). We thank Adam Kuspa and Gilbert Chu for reading the manuscript, Gilbert Chu and Marjorie Thomas for helpful discussions and David Beach for providing the S. pombe DNA probes.

REFERENCES
Initiation of simian virus 40 DNA replication in vitro: identification of RNA-primed nascent DNA chains

Janos Taljanidisz1,2, R. Scott Decker1,3, Zong-Sheng Guo2,4, Melvin L. DePamphilis2,4 and Nilima Sarkar1,2*

1Department of Metabolic Regulation, Boston Biomedical Research Institute, 20 Staniford Street, Boston, MA 02114 and 2Department of Biological Chemistry, Harvard Medical School, Boston, MA 02115, USA

Received May 26, 1987; Revised and Accepted September 3, 1987

ABSTRACT

Cell-free extracts of simian virus 40 (SV40)-infected CV-1 cells can initiate large tumor antigen dependent bidirectional replication in circular DNA molecules containing a functional SV40 origin of replication (ori). To determine whether or not DNA replication under these conditions involves RNA-primed DNA synthesis, replication was carried out in the presence of 5-mercuri-deoxyctydine triphosphate to label nascent DNA chains. Newly synthesized mercurated DNA was isolated by its affinity for thiol-agarose, and the 5'-ends of the isolated chains were radiolabeled to allow identification of RNA primers. At least 50% of the isolated chains contained 4 to 7 ribonucleotides covalently linked to their 5'-end; 80% of the oligoribonucleotides began with adenosine and 19% began with guanosine. About 60% of the nascent DNA chains annealed to the SV40 ori region, and about 80% of these chains were synthesized in the same direction as early mRNA. These results are consistent with the properties of SV40 DNA replication in vivo and support a model for initiation of SV40 DNA replication in which DNA primase initiates DNA synthesis on that strand of ori that encodes early mRNA.

INTRODUCTION

Simian virus 40 (SV40) provides a simple model of a single eukaryotic chromosomal replicon (see refs. 1 and 2 for review). The viral genome is a 5.2 kilobase circular double-stranded DNA molecule that replicates in the nucleus of its host as a chromosome with a histone composition and nucleosome structure similar to cellular chromatin. When the virus-encoded protein large tumor antigen (T-ag) and permissive cell factors interact with the cis-acting sequence genetically defined as the SV40 origin of replication (ori), replication is initiated bidirectionally at a unique site (the origin of bidirectional replication) within ori. Except for T-ag, all of the components required for the replication and assembly of SV40 chromosomes are provided by the host cell, and the events at replication forks in cellular chromosomes appear identical to those in SV40 and its close relative, polyoma virus (3, 4).

Short nascent DNA chains are repeatedly initiated on average once every 145 base pairs (bp). This process of discontinuous DNA synthesis occurs predominantly, if not exclusively, on the retrograde arm of replication forks where the direction of synthesis must be opposite to the direction of fork movement; synthesis on the forward arm is a relatively continuous process. A short oligoribonucleotide is synthesized first, serving as a primer for DNA synthesis. It is
then excised at the same rate at which completed Okazaki fragments are joined to the 5'-end of the continuously growing daughter strand. All steps in the synthesis of these RNA primed-DNA chains appear to be carried out by a multiprotein complex that includes DNA primase, DNA polymerase-α [or DNA polymerase-δ (5, 6)], C1C2 primer recognition proteins and other proteins relevant to DNA replication such as topoisomerase and exonuclease activities. Initiation sites for RNA-primed DNA synthesis are also found within the SV40 [and polyoma virus (3)] ori sequence on the strand that encodes early mRNA. These sites are indistinguishable from those found outside ori in terms of their frequency, sequence composition, confinement to retrograde arms of forks and average lengths of RNA primer, suggesting that initiation of DNA synthesis at ori occurs by the same mechanism used to initiate Okazaki fragments throughout the genome. Similar results have recently been obtained with polyoma virus-replicating DNA (4).

Soluble extracts from various permissive mammalian cells are capable of initiating replication in circular DNA molecules containing ori 6-12). Replication is bidirectional from ori and requires a functional T-ag, ori-core sequence, and DNA primase-DNA polymerase-α (11, 13) [or -i (5, 6)]. Although these results imply that DNA replication in these in vitro systems occurs in the same way that SV40 replicates in vivo, there has been no direct evidence of RNA-primed DNA synthesis. In fact, some data indicate that significant differences may exist. First, although DNA synthesis in these in vitro systems depends upon addition of deoxyribonucleoside triphosphates, with the exception of ATP, they do not depend on addition of ribonucleoside triphosphates, indicating that RNA synthesis is not a prerequisite of DNA synthesis. Secondly, although SV40-directed DNA replication in vitro is resistant to α-amanitin (6), it is inhibited by ribonuclease A (7, 14), suggesting a role for preformed RNA in the replication process. Finally, DNA primase-DNA polymerase-α purified from permissive monkey cells initiates synthesis at distinctly different sites and makes RNA primers significantly shorter when using SV40 DNA templates in vitro than observed on the same template sequences during SV40 DNA replication in vivo (15-17), indicating that additional proteins or even different enzymes are involved in DNA replication in vivo. Therefore, we employed a novel approach involving mercurated nucleotides in an effort to identify RNA-primed nascent DNA chains synthesized during SV40-directed DNA replication in vitro.

5-mercuri-deoxycytidine triphosphate (Hg-dCTP) has been used as a substrate for in vitro DNA replication in both bacterial (18-22) and mammalian cell systems (23) to facilitate the isolation and characterization of RNA-linked DNA under conditions where RNA primers are rapidly excised and low in abundance. Nascent Hg-labeled DNA chains are isolated by affinity chromatography on thiol-agarose and their 5'-ends analyzed for the presence of oligoribonucleotides. Application of this approach to a system that initiates SV40 DNA replication in vitro demonstrated the presence of RNA-primed nascent DNA chains in and around the ori region and revealed that the initial direction of DNA synthesis was the same as.

7878
that of early mRNA synthesis. The results are consistent with the model for initiation of SV40 DNA replication proposed by Hay and DePamphilis (24).

MATERIALS AND METHODS

In Vitro DNA Replication

African Green monkey kidney cell line CV-1 was infected with SV40 wild-type 800 (25). From 36 to 38 hrs post infection, cytosol and nuclear extract was prepared as previously described (6, 11, 12). SV40-infected cells were washed with hypotonic buffer [20 mM HEPES (pH 7.8), 5 mM potassium acetate, 0.5 mM MgCl2, and 0.5 mM dithiothreitol] and lysed in a Dounce homogenizer. Nuclei were sedimented by centrifugation at 1,200 x g for 5 min and then resuspended in hypotonic buffer supplemented with 500 mM potassium acetate to prepare a high-salt nuclear extract. SV40 chromosomes were removed from the high-salt nuclear extract by centrifugation at 300,000 x g for 1 hr. Cytoplasm (supernatant from nuclei sedimentation) was centrifuged at 100,000 x g for 1 hr to yield a cytosol fraction. The DNA substrate in these experiments was pSVori, a 2,643 bp plasmid constructed by replacing nucleotide 29 to 562 of plasmid vector pML-1 with a 206 bp segment of SV40 DNA containing ori (12).

DNA replication was carried out in a final volume of 125 µL containing 30 mM HEPES (pH 7.8), 7 mM magnesium acetate, 1 mM EGTA, 4 mM ATP, 0.2 mM each of CTP, GTP, and UTP, 0.1 mM each of dATP and dGTP, 0.05 mM Hg-dCTP (Sigma), 100 µCi [3H]dTTP (26 Ci/mmol gives 0.03 mM dTTP), 0.5 mM dithiothreitol, 10 mM phosphoenol pyruvate, 3.75 µg pyruvate kinase, 3% polyethylene glycol (14,000 Mr, Aldrich Chemical Co.), 0.25 µg pSVori DNA, 50 µL cytosol (7-12 mg protein/mL) and 25 µL high salt nuclear extract (5-9 mg protein/mL). This mixture was incubated for 15 min at 30°C, unless otherwise indicated, before stopping the reaction by addition of 0.5 % sodium dodecyl sulfate and 15 mM EDTA. DNA was purified by treatment with proteinase K, extraction with organic solvents and precipitation with ethanol, as previously described (6, 11, 12).

Affinity Chromatography of Hg-DNA

DNA purified from the in vitro DNA replication reaction was dissolved in 50 µL of water and chromatographed on Bio-Gel P4 (0.8 x 25 cm column) in 0.1 M triethyl ammonium bicarbonate (pH 7.5) to remove residual dithiothreitol. Use of siliconized columns and tubes was essential to prevent loss of Hg-DNA. DNA was recovered in the void volume, evaporated to dryness under vacuum, and then dissolved in 100 µL 0.05 M Tris-Cl (pH 7.5), 0.1 M NaCl, and 1 mM EDTA. The DNA was denatured by heating for 5 min at 100°C, diluted to 1 mL with 10 mM Tris-Cl (pH 7.5), 1 mM EDTA, and 10 mM NaCl and then chromatographed on a 1 mL column of thiol-agarose (Affi-Gel 401, Bio-Rad), as previously described (20-22).

DNA Hybridization Assays

Single-stranded mSV01, mSV02, and M13mp8 DNA were treated with 0.3 N NaOH for 1
hr at 65°C, diluted with an equal volume of 2 M ammonium acetate, and then applied to MS1 Magna nylon filters in serial dilutions using a Schleicher and Schuell filtration manifold, as previously described (26). Pre-hybridization and hybridization were performed with [5'-32P]Hg-DNA in a buffer containing 0.75 NaCl, 0.05 M NaH2PO4, 5 mM EDTA, 0.1% (v/v) Ficoll, 0.1% (v/v) polyvinylpyrrolidone, 0.1% (w/v) bovine serum albumin, 1% sodium dodecyl sulfate, 200 μg/mL denatured salmon sperm DNA and 50% (v/v) formamide. After 24 hrs of hybridization at 37°C, the filters were washed as described (26). These conditions were optimal for preserving RNA in RNA-p-DNA chains and for maintaining Hg-DNA:DNA hybrids which have a lower melting point than DNA:DNA hybrids (27). Following preliminary autoradiography to locate samples, each spot was cut out of the filter and the amount of [5'-32P]Hg-DNA measured by scintillation counting.

RESULTS

Labeling Nascent DNA with Hg-dCTP

The role of RNA primers during the initiation and continuation of SV40-directed DNA replication in vitro was explored using a cell-free system supplemented with 5-mercurideoxyctydine triphosphate (Hg-dCTP) and programmed with a circular plasmid (pSVori) containing the SV40 ori sequence. The original low-salt extract of SV40-infected CV-1 cells described by Su and DePamphilis (28, 29) that allowed SV40 replicating chromosomes to continue replication in vitro was modified to include a high-salt nuclear extract and polyethylene glycol (6). These modifications allowed initiation of multiple rounds of bidirectional replication from ori in SV40 chromosomes as well as in plasmid DNA. Replication depended on a functional SV40 T-ag and ori-sequence (6, 11, 12). Furthermore, promoter sequences that facilitate activation of ori-core in vivo [the 'GGGCGG'-repeats (1, 2)] were utilized in binding of initiation factor(s) specifically to SV40 ori-core (12). The products of replication included covalently-closed, superhelical DNA . Stillman (9) has shown that a high-salt nuclear extract is required for supercoiling of newly replicated DNA, which results from nucleosome assembly, although such extracts also reduce the amount of DNA synthesis observed. Therefore, several important characteristics of initiation of viral chromosome replication in vivo were reproduced under these in vitro conditions.

Substitution of Hg-dCTP for dCTP had no discernible effect on the characteristics of DNA replication that we have previously described for this system (6, 11). An average of 4.5 pmol [α-32P]dCTP (18 pmols of all four dNTPs) was incorporated during the first hour of incubation using either dCTP or Hg-dCTP, and synthesis with either nucleotide was inhibited 50% in the presence of 3 μg/mL aphidicolin. Incorporation of [α-32P]dNTPs was reduced at least 100-fold when plasmids in which SV40 ori was either absent or contained a 6 bp deletion at the BglII site were used, at least 100-fold when T-ag specific monoclonal antibody PAb419 was added, and at least 20-fold when linear DNA containing a functional ori was used. These results were
consistent with observations of other in vitro systems that initiate SV40 DNA replication (7-10). The products of replication with either dCTP or Hg-dCTP were essentially the same when fractionated by electrophoresis in agarose gels (data not shown). DNA synthesis with either dCTP or Hg-dCTP exhibited a delay of about 15 min; DNA synthesis during the first 15 min of incubation was only 6% of that observed after 2 hr. About 87% of the Hg-DNA synthesized during the initial 15 min of incubation could be recovered by affinity chromatography on thiolagarose. Therefore, this material was further analyzed for the presence of 5'-terminal oligoribonucleotides.

RNA Primers on Nascent DNA Chains

The mercurated [3H]DNA isolated above (Hg-DNA) was treated with bacterial alkaline phosphatase and then incubated with [γ-32P]ATP in the presence of phage T4 polynucleotide kinase (24, 25) so that each polynucleotide chain carries a single phosphate and is radiolabeled at its 5'-end. Nonutilized [γ-32P]ATP was removed by chromatography on Bio-Gel P4. Rechromatography of the [5'-32P]DNA through Bio-Gel P4 confirmed that it eluted in the void

![Figure 1](image_url)

**Figure 1.** Fraction of [5'-32P]HgDNA covalently attached to RNA primers. Panel A: [5'-32P]HgDNA purified from the in vitro DNA replication reaction (see Materials and Methods) was subjected to gel filtration through Bio-Gel P4 (0.8 x 50 cm column) in 0.1 M triethylammonium bicarbonate (pH 7.5). Fractions (0.15 mL) were collected and analyzed for 32P by scintillation counting. Panel B: An aliquot of the [5'-32P]HgDNA analyzed in panel A was precipitated with ethanol, resuspended in 0.15 M NaOH, and incubated for 20 hrs at 37°C to hydrolyze RNA primers (24, 25). This material was subjected to gel filtration as in panel A.
volume (Fig. 1A). Since oligo(dT)18 also eluted in the void volume of this column, [5'-32P]DNA chains were at least 18 residues in length. This material was incubated with alkali to hydrolyze RNA and then chromatographed on Bio-Gel P4 (Fig. 1B). At least 50% of the 32P-label behaved as 3', 5'-ribonucleoside bisphosphates (20), revealing that at least 50% of the [3H]Hg-DNA carried one or more ribonucleotides at their 5'-ends.

The 32P-labeled ribonucleosides generated by alkaline hydrolysis of isolated nascent DNA were chromatographed on poly(ethylenimine) cellulose (20). About 80% of the radiolabel comigrated with adenosine 3', 5'-bisphosphate, 19% with guanosine 3', 5'-bisphosphate, and 1% with cytidine and uridine 3', 5'-bisphosphates (Fig. 2). Therefore, the putative 5'-oligoribonucleotides covalently linked to Hg-DNA began predominantly with adenosine and secondarily with guanosine.

To determine whether or not the oligoribonucleotides attached to the 5'-ends of nascent DNA chains were the size expected of RNA primers, [5'-32P]Hg-DNA was digested exhaustively with the 3' to 5' exonuclease of bacteriophage T4 DNA polymerase (20, 24, 25). This enzyme specifically digests single-stranded DNA in RNA-primed DNA chains leaving a single deoxyribonucleotide attached to the 3'-end of the RNA moiety (15). The products were then fractionated by polyacrylamide gel electrophoresis (15, 20, 24, 25). Prior to digestion, [5'-32P]DNA chains remained at the top of the gel (Fig. 3, lane A), whereas after digestion,

![Figure 2](image)

Figure 2. Identification of the 5'-ribonucleotide associated with RNA-p-[Hg]DNA. The [5'-32P], 2',3'-ribonucleoside diphosphates released from [5'-32P]HgDNA by alkaline hydrolysis and isolated by gel filtration (see Figure 1B) were fractionated by thin layer chromatography on PEI cellulose using 0.55 M ammonium sulfate as the solvent and a mixture of the four ribonucleoside 3', 5'-bisphosphates as standards. 32P-nucleotides were detected by autoradiography and nucleotide standards by adsorption of short wave UV light. The autoradiogram was scanned using a soft laser scanning densitometer (Zeinek).
most of the 32P-labeled material migrated with oligoribonucleotides 5 to 8 residues in length (Fig. 3, lane B). Digestion of this material with RNase T2 generated 5'-labeled mono- and dinucleotides (Fig. 3, lane C). Therefore, nascent DNA chains synthesized in vitro had 4 to 7 ribonucleotides covalently attached to their 5'-ends.

**RNA-Primed Nascent DNA Chains Initiated In the Ori-Region**

To determine whether or not any of these RNA-primed nascent DNA chains originated in the SV40 ori-region, Hg-DNA isolated after 15 minutes of incubation in vitro was hybridized with SV40 specific probes consisting of single-stranded M13 phage DNA containing either the SV40 ori-region DNA strand that encodes early mRNA (mSVO1) or the strand that encodes late mRNA (mSV02). All 5'-ends of one aliquot of the isolated Hg-DNA were radiolabeled as

![Image](figure3.png)

**Figure 3. Size of RNA primers covalently attached to HgDNA.** [5'-32P]HgDNA was subjected to electrophoresis in a 20% polyacrylamide gel containing 7 M urea (20) before (lane A) and after (lane B) digestion with the 3' to 5' exonuclease of phage T4 DNA polymerase or with both T4 DNA polymerase exonuclease and RNase T2 (lane C). The gel was calibrated by concurrent electrophoresis of a partial alkaline digest of [5'-32P]poly(A). Numbers indicate chain length in nucleotides.
TABLE 1. Hybridization of [5'-32P]Hg-DNA to the SV40 Ori-Region

<table>
<thead>
<tr>
<th>Nascent DNA Structure</th>
<th>Amount* of [5'-32P]Hg-DNA Hybridized to:</th>
<th>Ratio#</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mSV01</td>
<td>mSV02</td>
</tr>
<tr>
<td>RNA-p-DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(15,000 cpm)</td>
<td>3460</td>
<td>844</td>
</tr>
<tr>
<td>(30,000 cpm)</td>
<td>4890</td>
<td>1522</td>
</tr>
<tr>
<td>RNA-p-DNA + DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(15,000 cpm)</td>
<td>2254</td>
<td>1356</td>
</tr>
<tr>
<td>(30,000 cpm)</td>
<td>5070</td>
<td>2096</td>
</tr>
</tbody>
</table>

* All numbers are counts per minute.

# Nonspecific DNA binding (M13mp8) was subtracted from the values obtained with mSV01 and mSV02 before calculating ratio. mSV01 and mSV02 are single-stranded circular DNA molecules obtained by inserting the 311 bp BstN1 SV40 DNA fragment containing ori into recombinant vector M13mp7 (24). mSV01 carries the early mRNA template strand of ori, and mSV02 carries the late mRNA template strand.

described above; about half the sample was [5'-32P]RNA-p-[Hg]DNA and half was [5'-32P]Hg-DNA. A second aliquot was first incubated in the presence of ATP and T4 polynucleotide kinase to insure that the 5'-ends of all polynucleotide chains were terminated by a phosphate. RI:A primers were then removed by alkaline hydrolysis, and the resulting 5'-hydroxyl terminated DNA chains were radiolabeled using [Y-32P]ATP and T4 polynucleotide kinase (24, 25). Thus, all [5'-32P]DNA chains in the second aliquot represented those DNA chains whose 5'-ends used to be covalently attached to RNA (i.e. RNA-p-DNA chains). Two different amounts of 32P-labeled material was hybridized to each M13 probe to insure that the probe's hybridization capacity was not saturated. About 24% of all nascent DNA chains originated from the SV40 ori-region, (Table 1, RNA-p-DNA + DNA). About 20% of RNA-primed nascent DNA chains hybridized specifically to mSV01 and about 5% hybridized to mSV02 (Table 1, RNA-p-DNA).

Hybridization of Hg-DNA is comparatively inefficient relative to unsubstituted DNA due to the low melting point of Hg-DNA hybrids (27). Hybridization efficiency in these experiments was 40% as measured with a 5'-end labeled, Hg-substituted restriction fragment that was complementary to mSV01 (26). Therefore, about 60% of all 5'-end labeled nascent DNA contained a sequence complementary to the 209 bp SV40 ori-containing sequence present in pSVori, the template for in vitro DNA synthesis. About 50% of RNA-primed nascent DNA chains hybridized to mSV01 and about 13% hybridized to mSV02, revealing that about 80% were synthesized in the same direction as early mRNA and that as many as 63% of these chains
may have originated in the SV40 ori-region. The remaining 30-40% of RNA-primed nascent DNA chains presumably represented Okazaki fragments that were initiated at replication forks outside this ori-region as a result of bidirectional replication (6-11). If the in vitro metabolism of RNA-primed nascent DNA chains is representative of their in vivo metabolism, then most of the nascent DNA chains that did not contain RNA at their 5'-ends resulted from normal excision of RNA primers as Okazaki fragments are ligated to the ends of growing daughter strands (1, 2, 25). Alternatively, some nascent DNA chains may have been broken during their purification.

DISCUSSION

Our earlier attempts to detect RNA primers during replication of plasmid DNA containing the SV40 ori-region by incorporation of [α-32P]dNTPs into newly replicated DNA were unsuccessful (unpublished data). Since pool sizes for rNTPs in mammalian cells exceed those of dNTPs by 500 to 1000-fold (30), dilution of radiolabeled rNTPs in the cell extracts may have prevented detection of RNA primer synthesis. In fact, a dependency of RNA-primed DNA synthesis on ribonucleotides has been difficult to demonstrate in subcellular systems where RNA-primed nascent DNA chains have been identified in a variety of ways (31-33). Therefore, we isolated newly synthesized DNA from the in vitro reaction by incorporation of an affinity label (Hg-dCTP) that allowed nascent DNA chains to be isolated from other DNA and RNA in the reaction by its affinity for thiol-agarose and then examined their 5'-ends for the presence of an oligoribonucleotide. This approach has been successfully used to detect RNA-primed DNA synthesis in other systems (18-23). The data revealed that at least 50% of the DNA chains synthesized during the initial "lag" period of pSVori DNA replication in an extract of SV40-infected CV-1 cells had 4 to 7 ribonucleotides covalently attached to their 5'-ends and that about 80% of these oligoribonucleotides began with adenosine and 19% with guanosine.

The results obtained with pSVori replication in vitro are remarkably similar to SV40 DNA synthesis in vivo. About 50% of the nascent DNA chains isolated from SV40 DNA replicating in vivo contain an average of 9 to 11 ribonucleotides attached to their 5'-ends; 61% of all oligoribonucleotides began with adenosine and 23% began with guanosine (25). Rapid degradation of RNA primers accounts for the fraction of oligoribonucleotides containing cytidine or uridine at their 5'-ends (25). When only (p)ppRNA-p-DNA chains were radiolabeled using vaccinia virus guanyltransferase, 80% of the RNA primers began with adenosine and 20% with guanosine; no initiation events were detected that began with either CTP or UTP (4).

The data on RNA-primed DNA synthesis during SV40 DNA replication in vivo and pSVori replication in vitro are consistent with the mechanism of action of purified DNA primase-DNA polymerase-α from permissive cells. The same regions of SV40 DNA analyzed in vivo have been used as single-stranded templates for purified DNA primase-DNA polymerase-α from CV-1 cells (15,16). This enzyme complex initiated RNA-primed DNA
synthesis 80% of the time with ATP and 20% of the time with GTP; initiation with either CTP or UTP was not detected. The typical length of RNA primers was 6 to 8 residues, although the length and base composition of individual RNA primers depended on the specific template initiation site chosen. A multiprotein complex containing DNA primase, DNA polymerase-α, C1C2 and other replication-related proteins initiated RNA-primed DNA synthesis on the same templates with similar characteristics except that RNA primers were typically 1 to 5 nucleotides long (17).

Part of the difference between the lengths of RNA primers synthesized under various conditions appears to result from the use of DNA size standards in the in vivo experiments and homo-oligomers of ribonucleotides as size standards in the in vitro experiments. Intact SV40 RNA primers synthesized in vivo typically contained 6 to 9 ribonucleotides when measured against oligo(rA) standards (4). Therefore, the typical RNA primer length synthesized during pSVori replication in vitro (4-7 bases) was characteristic of SV40 DNA replication in vivo (6-9 bases) and of CV-1 cell DNA primase-DNA polymerase-α DNA synthesis in vitro (6-8 bases). One important comparison remains. The initiation sites selected by purified DNA primase-DNA polymerase-α complexes in vitro were distinctly different from those selected by the cellular replication complex in vivo (15-17), suggesting that either additional proteins or different enzymes are required. Unfortunately, we were not able to map the initiation sites for RNA-primed nascent DNA chains synthesized on pSVori in vitro because restriction endonucleases failed to cut recognition sites containing Hg-dCMP (G. Banfalvi, S. Bhattacharya and N. Sarkar, manuscript in preparation). New methods for mapping these initiation sites are currently under development.

The data presented in this paper are consistent with the model proposed by Hay and DePamphilis (24, 25) for the initiation of SV40 DNA replication. In this model, a T-ag initiation complex separates the two strands of ori to allow DNA primase-DNA polymerase-α [or δ (5, 6)] to select one of several potential sites on the strand that encodes early mRNA to initiate RNA-primed DNA synthesis. From 60-70% of the RNA-primed nascent DNA chains synthesized during the first 15 minutes of incubation in vitro contained sequences complementary to the 209 bp ori region in pSVori. Ori itself consists of a 64 bp core sequence flanked on each side by an auxiliary sequence of 43-45 bp (1, 2). Since this comprises about 73% of the SV40 sequence present in pSVori, some RNA-primed DNA chains likely originated within ori. Moreover, 80% of these initiation events occurred on the template that also encodes early mRNA, suggesting that synthesis of the first DNA chain was initiated on that side of ori which encodes early mRNA. In fact, we have shown previously that the initial direction of DNA synthesis in SV40 chromosomes initiating replication under these in vitro conditions is the same as early mRNA synthesis (11) and that RNA-primed initiation events occur exclusively on the early mRNA template strand of ori during both SV40 (24) and polyoma virus (3) DNA replication in vivo. Further application of the techniques used in this study may allow precise
identification of the initiation site(s) for the first DNA chain synthesized during SV40 DNA replication.

ACKNOWLEDGEMENTS

We are grateful to Dr. H. Paulus for a critical reading of the manuscript. This investigation was supported by grants GM 25022 (N.S.) and CA 15579 (M.L.D.) from the National Institutes of Health.

*To whom correspondence should be addressed

Present addresses: 2Gene-Trak Systems, 31 New York Ave., Framingham, MA 01701, USA and
4Department of Cell and Developmental Biology, Roche Institute of Molecular Biology, Nutley, NJ 07110, USA

REFERENCES