Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 4641–4645, May 1995
Biochemistry

Rolling replication of short DNA circles

ANDREW FIRE* AND SI-QUN XU

Department of Embryology, Carnegie Institution of Washington, 115 West University Parkway, Baltimore, MD 21210

Communicated by Donald D. Brown, Carnegie Institution of Washington, Baltimore, MD, February 10, 1995

ABSTRACT **Natural genes and proteins often contain tandemly repeated sequence motifs that dramatically increase physiological specificity and activity. Given the selective value of such repeats, it is likely that several different mechanisms have been responsible for their generation. One mechanism that has been shown to generate relatively long tandem repeats (in the kilobase range) is rolling circle replication. In this communication, we demonstrate that rolling circle synthesis in a simple enzymatic system can produce tandem repeats of monomers as short as 34 bp. In addition to suggesting possible origins for natural tandem repeats, these observations provide a facile means for constructing libraries of repeated motifs for use in "*in vitro* evolution" experiments designed to select molecules with defined biological or chemical properties.**

Analysis of naturally occurring macromolecular sequence has revealed repetitive structure at a variety of levels (1, 2). Particularly relevant to gene expression and replication are sets of short sequence motifs that often occur in multiple copies around promoter/enhancer regions and replication origins (3). The repetition of motifs within a control region has been shown in many cases to allow individual trans-acting factors to exert additive and/or cooperative effects; this design can improve the specificity of a control mechanism by increasing the signal of appropriate activity while decreasing the possibility of fortuitous inappropriate activity (4). Requirements for repeated sequence motifs have also been found in characterizing the activities of specific RNA (5–8) and protein (9–11) functions.

In investigating structure–function relationships *in vitro* and *in vivo*, several researchers have used strategies that involve the production of a large library of random sequences followed by selection for sequences with a given property (12). These schemes can produce experimentally useful reagents and provide a wealth of information about sequence requirements for the selected activity (e.g., refs. 13–15). Application of such a selection strategy depends on the ability to produce large libraries of random sequences, efficient selection procedures, and appropriate means for recovering and characterizing the selected molecules. Frequently, the techniques for selection or screening of molecules are insufficient to find active sequences. In particular, if several tandem copies of a functional segment are required for activity, then the problem of recovering an active sequence from a random pool becomes increasingly more difficult.

To circumvent the insufficiency of available selection techniques for many interesting biological and biochemical activities, we sought to produce libraries of random repeated sequences: pools of molecules in which each member contains tandem repeats of a different sequence element. The potential usefulness of such concatamer libraries can be illustrated by calculating the probability that a given 8-bp element will occur independently in three positions in a single random 60-mer sequence (≈1 in 250 million). If we replace the random 60-mers with a library of trimerized random 20-mers, then this probability improves by ≈5 orders of magnitude.

We considered several different methods for constructing concatamer libraries. Chemical synthesis can be used to generate random pools of DNA sequence oligomers (12), but straightforward ligation to concatenate elements from these pools would not produce the desired result, since there is no means to ensure that ligation joins molecules of the same sequence. A straightforward method for generation of a small library would be to separately synthesize and concatenate a number of separate oligonucleotides (16); unfortunately, this would be cumbersome and expensive for large libraries.

As a more general procedure we chose a scheme based on the rolling circle replication used by many plasmids and viruses (17, 18). Rolling circle replication involves two simultaneous processes: (*i*) DNA polymerase must synthesize sequences complementary to a circular template. (*ii*) As this replication proceeds, some mechanism must unwind the parental duplex to allow the polymerase to advance. Models for physiological rolling circle replication generally involve a template that is predominantly double stranded, with a helicase or single-strand DNA binding activity preceding the polymerase to allow replication to continue (17, 18). Characterized rolling circle replication mechanisms have been found to operate on templates on the order of kilobases and larger (18). Rolling circle replication of templates smaller than 100 bp by previously described mechanisms would be considered unlikely, since formation of very short double-stranded circles would be topologically obstructed (19). Although there was no precedent, we chose to examine the ability of predominantly single-stranded circles to act as templates for rolling circle synthesis.

MATERIALS AND METHODS

Reagents. Oligonucleotides were made on an Applied Biosystems DNA synthesizer (model 380B), desalted, precipitated with ethanol, and used without further purification. Bacteriophage T4 DNA ligase, polynucleotide kinase, and DNA polymerase were from New England Biolabs. *Escherichia coli* DNA polymerase was from Bethesda Research Laboratories; DNA polymerase large fragment (Klenow) and *Sty* I were from Boehringer Mannheim. Modified phage T7 DNA polymerase (Sequenase) was from United States Biochemical.

Generation of Putative Concatamer Products. *Step 1.* Primer/template was produced by dilute annealing followed by ligation. Template oligonucleotide (zf42 or zf43) (0.8 nmol) was phosphorylated with T4 polynucleotide kinase (60 Richardson units) in 250 μl of LK buffer (50 mM Tris·HCl, pH 7.8/10 mM $MgCl_2$/10 mM dithiothreitol/1 mM ATP/25 μg of bovine serum albumin per ml). After heating to 70°C for 10 min, the material was immediately diluted at 37°C into 12 ml of LK buffer containing 0.7 nmol of primer oligonucleotide (zf39). After 30 min at 23°C, the sample was transferred to 16°C and incubated for 4 hr with T4 DNA ligase (120 Weiss units). Ligase reactions were stopped by addition of EDTA (to 10 mM), SDS (to 0.1%), NH_4OAc (to 1 M), and 50 μg of glycogen, extracted once each with 8 ml of phenol/chloroform

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*To whom reprint requests should be addressed.

(1:1) and chloroform and precipitated twice with ethanol. Final samples were resuspended in 25 μl of TEN (10 mM Tris·HCl, pH 7.5/1 mM EDTA/25 mM NaCl) and stored at −70°C. In handling the annealed primer/template, care was taken to avoid transient low-salt or high-temperature conditions that could lead to denaturation.

Step 2. Primer extension reactions were carried out in 20 μl with 0.8 μl of primer/template. Extension reaction mixtures contained 10 mM Bistris propane chloride; 10 mM NaCl; 1 mM dithiothreitol; 1 mM each dATP, dTTP, dGTP; 0.25 mM dCTP (including [^{32}P]dCTP to a specific activity of 0.5 Ci/mmol; 1 Ci = 37 GBq); 0.1 mg of bovine serum albumin per ml; and the indicated polymerase (in some experiments 50 mM NaCl was added; this did not change the product electrophoresis pattern). Incubations were at 23°C for T4 and Klenow [large fragment of *E. coli* DNA polymerase I (18)] and 37°C for *E. coli* DNA polymerase I and coliphage T7 DNA polymerase (Sequenase).

Although *E. coli* DNA polymerase was most efficient under the reaction conditions used, longer exposures revealed similar patterns for Klenow and T7; we have not extensively varied reaction conditions to optimize output with these enzymes. Similar reactions (not shown) with thermophilic DNA polymerases (Vent, *Pfu*, and *Taq*; reactions at 58°C) produced no detectable rolling circle products but instead produced a series of products indicative of multiround rolling hairpin replication as described by Cavalier-Smith (20) (although we cannot rule out the possibility that some rolling hairpin products were also produced in reactions with the other polymerases, neither the partial digestion products nor the structures of cloned products were indicative of this as a major component of the *E. coli* polymerase I reaction material).

Step 3. For further characterization by restriction digestion and cloning, we used material from 4-hr extension products produced with *E. coli* DNA polymerase I. This analysis required that the product be converted to a double-stranded form. In initial experiments, an extra round of DNA synthesis was performed to produce the second strand [this was done by isolating the reaction product and allowing self-priming in a standard hairpin reaction (20)]. Subsequently, we found that much of the product was double stranded even without this second round of synthesis [a plausible explanation would be that self-priming hairpin structures (20) form at some frequency during the initial 4-hr synthesis reaction, perhaps after unraveling of the link between the elongating DNA terminus and the rolling circle template].

Partial digestion was carried out with restriction endonuclease *Sty* I, which should cut just once in each tandem repeat. To check the efficiency of the partial digest, samples were resolved after denaturation on a 6% acrylamide sequencing gel. Preparative samples were resolved in parallel without denaturation; these produced a shifted ladder of bands consistent with the double-stranded character of the material. The nondenatured samples from the *Sty* I partial digest were excised, eluted, ligated into a suitable *Sty* I-cut bacterial plasmid vector, and transformed into a Rec$^-$ bacterial host. Production of cloned libraries from the partial digest material was relatively efficient; by using standard protocols for bacterial transformation (10^6 colonies per μg of control plasmid DNA), several hundred clones were obtained with material derived from 1 pmol of input oligonucleotides. Much larger libraries could readily be obtained by optimizing and scaling up the bacterial transformation.

RESULTS

The experimental design (Fig. 1) was based on a consideration of possible behavior for an oligonucleotide primer that is being extended by DNA polymerase after annealing to a single-strand circular template. As the primer is extended, a double-

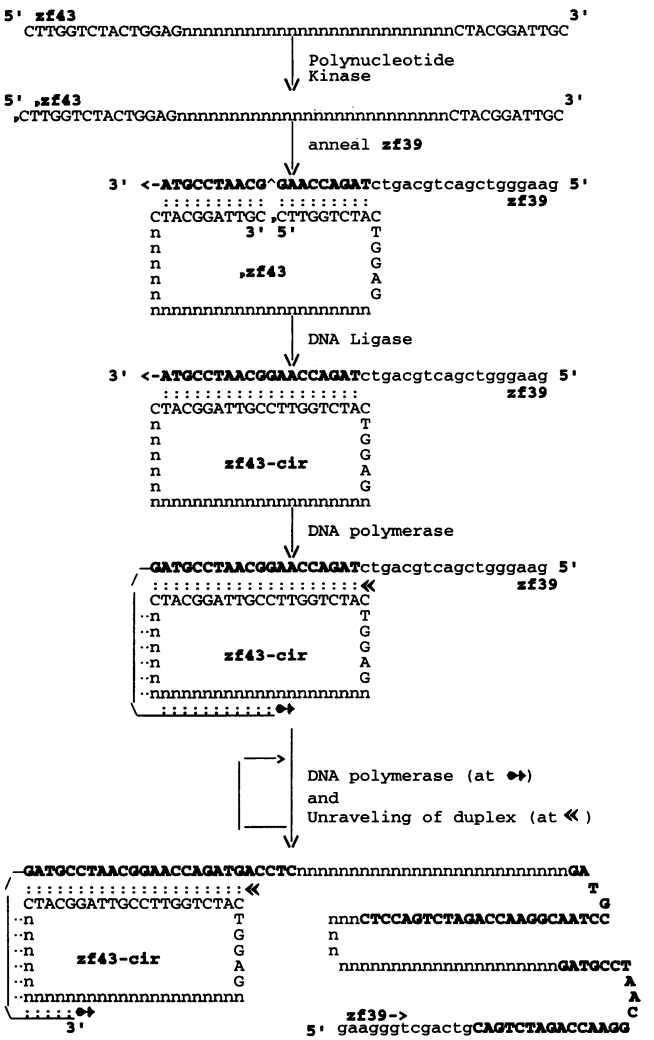


FIG. 1. Scheme for generation of concatamer library using rolling circle replication. The template oligonucleotide (zf43) is circularized using a partially complementary oligonucleotide (zf39) as a guide. This primer template is then reacted with DNA polymerase in the presence of deoxynucleoside triphosphate precursors. As synthesis proceeds, the polymerase creates duplex DNA. At some point, the circle becomes constrained so tightly that polymerization cannot proceed without some relief of the restraint. One possible outcome is shown: the constraint on the small circle might drive unwinding at the lagging end of the duplex. If this process is combined with continued polymerization, then a long repeating polymer would be produced. For this figure, the 52-base oligonucleotide zf43 is shown as template; oligonucleotide zf42, used in other experiments, is identical in the constrained regions but has 10 fewer nonspecified residues (total length, 42 bases).

stranded region of the circle will be formed. At a certain point, this double-stranded region will become sufficiently long to "strain" the circular topology. At this point, one of three events might occur: (*i*) the polymerase might stall or stop; (*ii*) the template might be forced to unravel behind the polymerase; or (*iii*) unwinding might occur at the site of polymerization—this could lead to self-priming of the product strand (20).

If unraveling behind the polymerase occurs, then a rolling circle mechanism could be set up; the production of new double-stranded regions as the polymerase progresses 5′ to 3′ would provide a steric force unwinding the DNA duplex behind the active polymerase. Under these circumstances, the energy for both polymerization and helicase activity would be derived from utilization by DNA polymerase of nucleoside triphosphates.

The substrates we used to test this rolling circle replication scheme are shown in Fig. 1. A primer template pair is set up

by using two oligonucleotides with a permuted region of complementarity. Annealing and ligation of these two oligonucleotides should produce a primed, partially single-stranded circle. To provide diversity in the template population, most of the unpaired residues are nonspecified (i.e., an equimolar mixture of A, G, C, and T). Rolling circle synthesis should then generate a population of long single-stranded products, each containing repeats complementary to an individual template molecule.

In initial experiments with several available DNA polymerase enzymes, we found that *E. coli* DNA polymerase I was particularly efficient in synthesizing high molecular weight products from the primer template; after 3 hr, a large amount of high molecular weight product had formed (Fig. 2*A*).

After conversion to a double-stranded product, the material was further characterized by partial restriction digestion, with an enzyme cutting once in each repeat. The resulting ladder of bands (Fig. 2*B*) was consistent with a tandem array structure derived from rolling circle synthesis. As expected, more complete digestion with the restriction enzyme (Fig. 2*C*) produced material that was predominantly monomer length. An estimate of the yield from a 4-hr extension with *E. coli* polymerase I was obtained from these experiments: ≈15 ng of *Sty* I-sensitive high molecular weight material was synthesized per pmol of input oligonucleotides.

As a critical test of the rolling circle model, it was essential to show that the reaction indeed produced repeats derived from individual input templates. This was demonstrated by

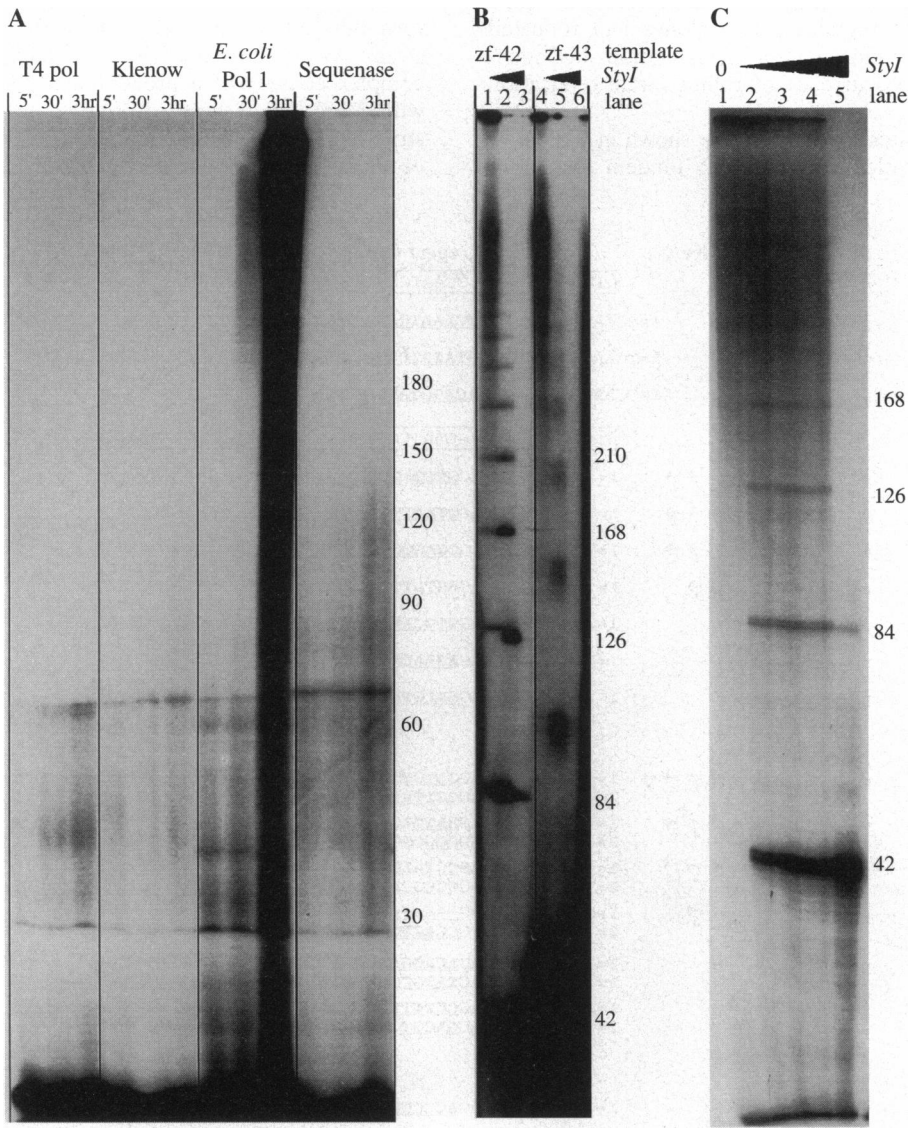


FIG. 2. Electrophoretic separation of reaction products. (*A*) Time course of extension reactions. The annealed primer/template diagrammed in Fig. 1 was extended with the indicated polymerase in the presence of [^{32}P]dCTP. At the indicated time points, aliquots were removed, denatured, and resolved on a sequencing-type gel. Base numbers on the right are sizes deduced from standard DNA sequencing reactions run on the same gel (not shown). A delayed time course in the synthesis reaction with *E. coli* polymerase I (Pol I) (compare 30-min and 3-hr samples) has been seen with a variety of primer/template combinations and polymerase preparations. The nature of this apparent lag has not been investigated. (*B*) Partial digestion of reaction products with restriction enzyme *Sty* I. As above, samples were denatured and resolved on sequencing-type gels. Lanes: 1–3, digestion of zf42 extension product; 4–6, identical except that the template has been derived from oligonucleotide zf43, which is longer than zf42 by 10 bases. The broadened bands observed in lanes 4–6 apparently result from the presence of shorter (internally deleted) oligonucleotide molecules in the zf43 preparation. (*C*) Full digestion of reaction products with *Sty* I. Lanes: 1, undigested products from a 4-hr extension using zf42 as template; 2–5, digestion with concentrations of *Sty* I in a range expected to yield full digestion. The major band at 42 nt presumably represents completely digested unit-length material (the doublet band may represent the two different product strands). A small amount of material migrating at 84 bp with the highest *Sty* I concentration is likely to result from a residual level of incomplete digestion. Other (minor) bands have not been characterized.

cloning and sequencing of reaction products. Since the original oligonucleotides contained 16 (or 26) nonspecified bases, repetition of the identical sequence in several consecutive members of a concatamer would be conclusive evidence that the cloned multimer is indeed derived by a replication event from a single template molecule.

From one set of synthesis reactions, we sequenced 18 different cloned products (Fig. 3). All gave different sequences. Ten of these sequences had the predicted structures, with each having 3–5 tandem copies of a unique oligonucleotide from the original pool. Six of the characterized clones had an alternating structure in which two of the initial oligonucleotides alternate in a mixed tandem array. These products could result from dimer circles (bridged by two primer molecules) formed during annealing and ligation of the initial primer template. The remaining two clones had repeated structures that require more complex explanations but nonetheless evidence the occurrence of rolling circle replication events.

Using other templates similar to those shown in Fig. 1, we have observed replication products with tandem repeats of monomers as short as 34 bp. In these and previous experiments, there has been no indication of specific sequence requirements for rolling circle synthesis. In particular, examination of the recovered sequences from Fig. 3 reveals no evidence for bias in either base composition or secondary structure.

DISCUSSION

We have shown that *E. coli* DNA polymerase I is capable of carrying out a rolling circle type synthesis reaction on a very short circular template. Although the mechanism of this reaction has not been studied in detail, several theoretical considerations of topology and scale are relevant. First, the short templates used in these experiments would be unlikely to form fully double-stranded circles without extreme topological strain. As an alternative, it seems likely that the replicating complex consists of a predominantly single-stranded circle, with DNA polymerase working to extend a short double-stranded region. At a certain point, extending the total length of the double-stranded region would be energetically unfavor-

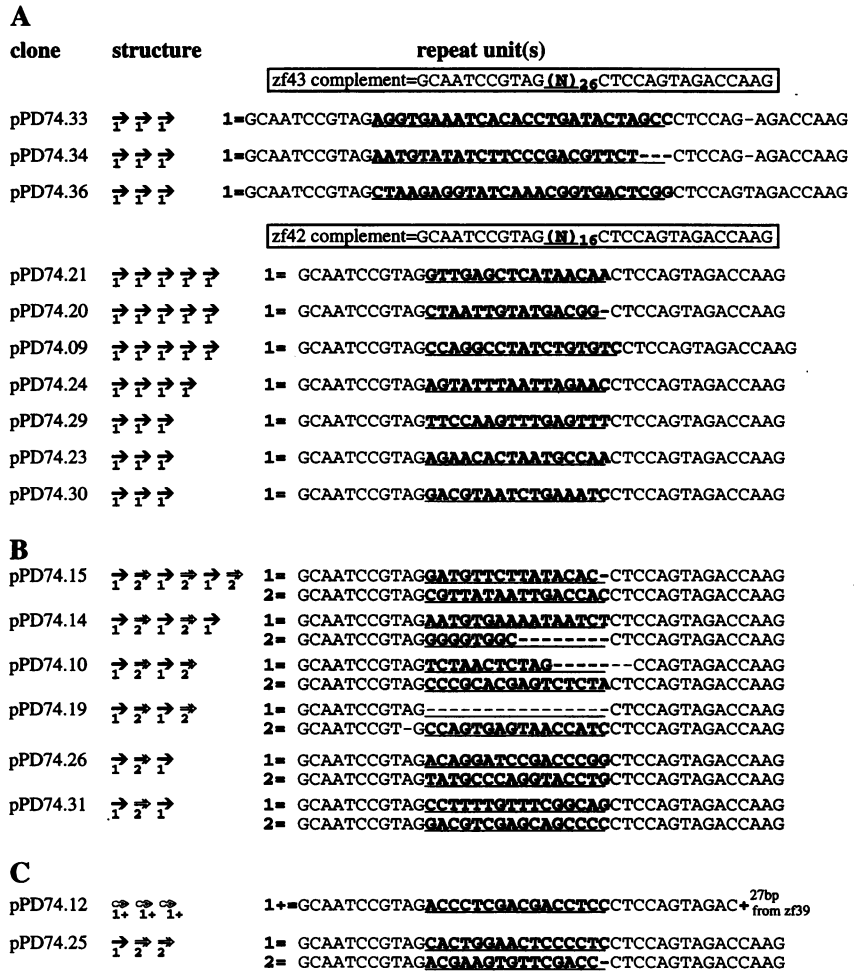


FIG. 3. Structures of 18 clones from *E. coli* DNA polymerase I extension of primer/templates. Sequences of DNA from 18 arbitrarily chosen clones from putative rolling circle synthesis are shown. Plasmids pPD74.33, -74.34, and -74.36 were derived from zf43; all others were derived from zf42. In three cases (pPD74.19, -74.21, and -74.23), the plasmids analyzed apparently contained two different inserts that had been joined during ligation of insert to plasmid vector; in these cases, only the first insert is described. (*A*) The ten "simple" products each had the structure predicted for rolling circle synthesis from the input primer/template. In several cases, the periodic repeat in these products differs in size by 1–4 bp from the design of the input oligonucleotide; this probably reflects a population of internally deleted oligonucleotides in the original zf42 and zf43 preparations. (*B*) Six products had an alternating pattern derived from two different input molecules. These are assumed to result from formation of dimeric primer/template circles in the initial ligation (two input molecules joined end to end to end). Consistent with this hypothesis, when the initial ligation to form zf42 circles was performed at a 10-fold higher concentration of template and primer, a majority of clones (70%) exhibited an alternating structure. (*C*) Two products had structures that were not simply explained, although each had a repeated structure consistent with a rolling circle event.

able. At this point, continued polymerase activity (at the leading end) could drive unwinding behind the polymerase.

A second constraint comes from a requirement for rotational positioning of DNA polymerase on the template. The twist in the local DNA helix at the site of synthesis necessitates a rotation of polymerase relative to template. This must occur without the polymerase physically passing through the cyclized DNA template [DNA polymerase I is sufficiently large (from the Klenow crystal structure; see ref. 21) that even with maximum contortion, a nucleotide chain of <60 residues could not encircle it]. As an alternative, polymerase could act essentially as a fixed surface while the template continually twists inward on itself. The combined forward and twisting motion of the template could provide the necessary constant interface between polymerase and template.

The ability of isolated polymerase to carry out rolling circle synthesis on a small DNA circle *in vitro* suggests that this process could have played an evolutionary role [along with mechanisms such as unequal recombination, transposition, and replication slippage (2, 18)] in generating the plethora of natural tandem-repeat structures that appear in coding and noncoding sequences. The enzymatic activities used in our scheme (kinase, ligase, and DNA polymerase) are readily available *in vivo*. Template oligonucleotides might come from a pool of replication intermediates or nucleic acid breakdown products present in cells. The final requirement (permuted complementary between primer and template oligonucleotides) would be a relatively rare coincidence; nevertheless, even a small genome would be expected statistically to contain many suitable combinations of sequences.

Given the efficiency of the *in vitro* reaction, it is conceivable that short circles might replicate *in vivo* as part of a concerted physiological or pathological process. The smallest known circular replicons are small viroid and satellite virus RNAs of several hundred base pairs (22). It is conceivable that there are (as yet undiscovered) shorter plasmids or viroid-like parasites that replicate by using the single-stranded rolling circle mechanism described here.

From an experimental point of view, the ability to produce large libraries of random or semirandom concatamers should have numerous applications. We have begun using these concatamer libraries to characterize cis-acting control sequences with defined activation patterns, with the goal of identifying the corresponding developmentally regulated transcription factors. Three other potential uses for concatamer libraries are notable.

(*i*) Vergnaud *et al.* (16) have used individually synthesized and concatamerized random oligonucleotides as hybridization probes to identify human DNA polymorphisms. The ability to make large libraries of random concatamers should greatly facilitate this approach.

(*ii*) In a variety of systems, sequences controlling translation have been localized to short repeated sequence motifs in the 3′ nontranslated leader of the mRNA (5–8). By constructing a library of concatamers inserted into the 3′ leader sequence for an easily assayed reporter gene/expression vector, it should be possible to identify control sequences producing different patterns of translational regulation.

(*iii*) Repeated motifs in proteins have been used in evolution to produce tight and very specific protein–protein interactions (9–11). Construction of a concatamer library in a protein coding context should allow *in vitro* or *in vivo* selection of repeated peptide sequence motifs that allow specific binding to a target protein.

Many of these selective schemes could be envisioned as a two-step process. Once a molecule with desired characteristics has been identified in any of the above screens, a second round of selection could be carried out to isolate sequences with optimal activity. This would be done by starting with material in which each base is only partially randomized relative to the initially recovered active sequence (e.g., an A residue in the initial selected sequence might be replaced by a mixture of 90% A + 5% G + 3% T + 2% C).

We are grateful to D. Brown, J. Corden, J. Gall, N. Fedoroff, V. Jantsch-Plunger, D. Kaiser, D. Koshland, M. Krause, P. Okkema, A. Pinder, H. Rienhoff, R. Schleif, Z. Wang, an anonymous reviewer, and members of the Department of Embryology for their help and suggestions. This work was supported by the National Institutes of Health (Grant R01-GM37706 to A.F.) and the Carnegie Institution. A.F. is a Rita Allen Scholar.

1. Britten, R. J. & Kohne, D. E. (1968) *Science* **161,** 529–540.
2. Beridze, T. (1986) *Satellite DNA* (Springer, Berlin).
3. Trifonov, E. N. & Brendel, V. (1986) *Gnomic: A dictionary of Genetic Codes* (Balaban, Rehovot, Israel).
4. Ondek, B., Shepard, A. & Herr, W. (1987) *EMBO J.* **6,** 1017–1025.
5. Theil, E. C. (1990) *J. Biol. Chem.* **265,** 4771–4774.
6. Wharton, R. P. & Struhl, G. (1991) *Cell* **67,** 955–967.
7. Goodwin, E. B., Okkema, P. G., Evans, T. C. & Kimble, J. (1993) *Cell* **75,** 329–339.
8. Wightman, B., Ha, I. & Ruvkun, G. (1993) *Cell* **75,** 855–862.
9. Corden, J. L. (1990) *Trends Biochem. Sci.* **15,** 383–387.
10. Bork, P. (1992) *FEBS Lett.* **307,** 49–54.
11. Engel, J. (1992) *Biochemistry* **31,** 10643–10651.
12. Szostak, J. (1992) *Trends Biochem. Sci.* **17,** 89–93.
13. Bartel, D. P. & Szostak, J. W. (1993) *Science* **261,** 1411–1418.
14. Beutel, B. A. & Gold, L. (1992) *J. Mol. Biol.* **228,** 803–812.
15. Barbas, C. F., Amberg, W., Simoncsits, A., Jones, T. M. & Lerner, R. A. (1993) *Gene* **137,** 57–62.
16. Vergnaud, G., Mariat, D., Apiou, F., Aurias, A., Lathrop, M. & Lauthier, V. (1991) *Genomics* **11,** 135–144.
17. Gilbert, W. & Dressler, D. (1968) *Cold Spring Harbor Symp. Quant. Biol.* **33,** 473–484.
18. Baker, T. A. & Kornberg, A. (1992) *DNA Replication* (Freeman, New York).
19. Koo, H.-S., Drak, J., Rice, J. A. & Crothers, D. M. (1990) *Biochemistry* **29,** 4227–4234.
20. Cavalier-Smith, T. (1974) *Nature (London)* **250,** 467–470.
21. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. (1985) *Nature (London)* **313,** 762–766.
22. Diener, T. O. (1991) *FASEB J.* **5,** 2808–2813.