(containing human Y chromosome) and GM06318B (containing human X chromosome) gave the expected product size (approximately 230 bp). sDF-2 was amplified under the same conditions used for sDF-1. The sequences of the sDF-2 primers are 5′-ACTAGGCGACTAATACAGTGGTGC-3′ and 5′-GTGAATTCATCATATGTGATTTCC-3′. The product size ranges between 163 and 177 bp. For in situ hybridization, approximately 500 ng of DNA from cosmids c364-10g and c250-9k was digested with Rsa I or Hinf I, labeled with biotinylated deoxyuridine 5′-triphosphate by random priming, and hybridized independently to metaphase spreads of normal human lymphocytes (46, XY). Signal detection was done with the use of an in situ hybridization kit from Oncor (Gaithersburg, MD) according to the manufacturer's recommendation.

19. Cycle sequencing was performed as described [B. R. Krishnan, R. W. Blakesley, D. E. Berg, *Nucleic Acids Res.* **19**, 1153 (1991)]. Fifteen cycles of 95°C for 1 min, ramp to 65°C for 5 min, and 75°C for 30 s were performed in a Perkin-Elmer Cetus thermocycler with 50 ng of cosmid DNA and 5 pmol of each end-labeled primer mix [$(AC)_{11}N$ (A, T, or G) or $(CA)_{11}N$(T, G, or C)] in a separate reaction. Gel separation and autoradiography were done according to standard protocols.
20. P. Green, CRI-MAP version 2.4, unpublished material.
21. In pedigree K1362, individual 4, a female offspring, inherited the 228-bp allele from her father's Y chromosome. In pedigree K1332, individual 6, a male offspring, inherited a 222-bp allele from his father's X chromosome.
22. J. A. Luty *et al.*, *Am. J. Hum. Genet.* **46**, 776 (1990); T. J. Hudson *et al.*, *Genomics* **13**, 622 (1992); A. Edwards, A. Civitello, H. A. Hammond, C. Caskey, *Am. J. Hum. Genet.* **49**, 746 (1991).
23. The physical distance between sDF-1 and sDF-2 was estimated from PCR amplification of these markers from overlapping YACs reported in (*17*). The smallest region of overlap among the YACs that contained both sDF-1 and sDF-2 is approximately 100 kb. For example, both markers were present in YAC yWXD630 but absent in YACs yWXD501 and yWXD37. sDF-2 (but not sDF-1) was present in yWXD250, and sDF-1 (but not sDF-2) was present in yWXD364 and yWXD427.
24. B. O. Bengtsson and P. N. Goodfellow, *Ann. Hum. Genet.* **51**, 57 (1987).
25. R. M. Speed, *Hum. Genet.* **78**, 260 (1988).
26. R. K. Saiki *et al.*, *Science* **230**, 1350 (1985).
27. NIH/CEPH Collaborative Mapping Group, *ibid.* **258**, 67 (1992).
28. Supported in part by NIH grants HG00100 and HG00201 (H.D.-K.). We thank R. Veile for assistance with in situ hybridization studies, T. Steinbrueck and R. Weaver for assistance with linkage map construction, and H. F. Willard for helpful comments on the manuscript.

18 August 1992; accepted 26 October 1992

DNA Sequencing by Primer Walking with Strings of Contiguous Hexamers

Jan Kieleczawa, John J. Dunn, F. William Studier

When template DNA is saturated with a single-stranded DNA binding protein (SSB), strings of three or four contiguous hexanucleotides (hexamers) can cooperate through base-stacking interactions to prime DNA synthesis specifically from the 3′ end of the string. Under the same conditions, priming by individual hexamers is suppressed. Strings of three or four hexamers representing more than 200 of the 4096 possible hexamers primed easily readable sequence ladders at more than 75 different sites in single-stranded or denatured double-stranded templates 6.4 kilobases to 40 kilobase pairs long, with a success rate of 60 to 90 percent. A synthesis of 1 micromole of hexamer supplies enough material for thousands of primings, so multiple libraries of all 4096 hexamers could be distributed at a reasonable cost. Such libraries would allow rapid and economical sequencing. Automating this strategy could increase the speed and efficiency of large-scale DNA sequencing by at least an order of magnitude.

Improvement in the efficiency of nucleotide sequencing is needed if the entire human genome is to be sequenced in the next 15 years at reasonable cost (*1*). A potentially efficient method is enzymatic sequencing by primer walking. A primer within a segment of known sequence (such as a vector sequence) is used to extend the sequence into an unknown region. The sequence thus determined is in turn used to select a primer to extend the sequence further, and this process is repeated until the sequence of the entire molecule has been determined. Advantages of primer walking are that the entire sequence can be determined on a single preparation of template DNA without subcloning and that the sequence can be determined with a minimum number of sequencing reactions.

A disadvantage of primer walking has been the inconvenience and expense of having to synthesize a primer for each sequencing reaction. A proposed solution is to use primers that are short enough that a manageable library of primers would allow any DNA molecule to be sequenced entirely with primers selected from the library (*2*). A potential way to obtain specific priming with a small library of primers is to generate longer, more specific primers from combinations of shorter, less specific ones. One proposal for doing this is by template-directed ligation (*3*), which would be compatible with current sequencing procedures. Efficient ligation requires that oligonucleotides pair at adjacent sites in the template DNA (*4*). We refer to sets of contiguous oligonucleotides as strings.

In trying to optimize conditions for specific priming by short oligonucleotides and to develop conditions for template-directed ligation, we discovered that saturating the template DNA with SSB stimulated strings of three or more unligated hexamers to prime specifically at the position of the string and at the same time suppressed priming by individual hexamers or by many pairs of contiguous hexamers. This stimulation of priming required that the hexamers be contiguous: priming by sets of three or four hexamers whose complements were separated by one or two nucleotides was suppressed by added SSB.

Conditions for priming by hexamer strings. The DNA polymerase used for sequencing was modified T7 DNA polymerase (*5*). The SSB of *Escherichia coli*, the gene *32* protein of bacteriophage T4, and the gene *2.5* protein of bacteriophage T7 (*6*) each stimulated good sequence ladders, but the *E. coli* SSB gave the best initial results and was used for the experiments described here. We used single-stranded M13 DNA (6407 nucleotides) or M13mp18 DNA (7250 nucleotides) (*7*) and one string of four hexamers (Table 1, A4–A1) to establish optimal conditions for priming by hexamer strings (*8*).

Standard reaction conditions were derived from the two-step Sequenase labeling and termination protocols for sequencing with ^{35}S label (U.S. Biochemical). An equilibration reaction (10 μl) typically contained 0.7 μg (33 nM) of M13 DNA, 3 μg of SSB (16 μM SSB monomer), and 50 pmol (5 μM) of each hexamer in 40 mM tris-HCl (pH 7.5), 50 mM NaCl, and 10 mM $MgCl_2$, plus 5% glycerol contributed by the stock solution of SSB. After equilibration for at least 5 min at 0°C, labeling was for 5 min at 0°C, followed by termination for 5 min at 37°C (*9*). We added SDS (0.1%) to the stop solution to prevent SSB from interfering with electrophoresis of the DNA on sequencing gels.

Labeling intensity was reduced at 5 or 20 mM $MgCl_2$ relative to 10 mM but was insensitive to NaCl concentration between 40 and 100 mM. The priming complex appears to reach equilibrium in 2.5 min at 0°C, whether SSB is added to an equilibrated mixture of DNA and hexamers or hexamers are added to an equilibrated mixture of DNA and SSB, as judged by intensity of labeling at different times after the last addition.

SSB concentration. *Escherichia coli* SSB

Biology Department, Brookhaven National Laboratory, Upton, NY 11973.

was obtained from two commercial sources (U.S. Biochemical and Promega) or purified by us after expression from the cloned gene. Titration showed that a mass ratio of SSB to DNA of slightly greater tha 2.5 maximally stimulated priming at the position of the string and almost completely suppressed priming elsewhere (Fig. 1). This saturating ratio corresponds to about 22 nucleotides of DNA per SSB monomer, 88 per tetramer, or 176 per octamer, consistent with the value of 175 nucleotides per octamer estimated for the beaded form of the SSB-DNA complex (*10*).

Concentrations of SSB greater than about 3.5 μg per 10-μl reaction were increasingly inhibitory for labeling of the sequence ladders (Fig. 1). Inhibition is due to the concentration of SSB and not the SSB to DNA ratio. Most early experiments used 5 μg of SSB, a slightly inhibitory amount that may, however, be slightly more effective in suppressing priming at secondary sites.

Hexamer concentration. Priming by hexamer strings depends on having a high enough concentration of hexamers to displace SSB and pair contiguously to the template at the position of the string. About 5 μM of each hexamer appeared to be sufficient to promote maximum intensity and uniformity of labeling of the sequence ladders under our reaction conditions; higher concentrations (10 to 50 μM) showed only slight improvement. At lower hexamer concentrations, labeling of shorter DNAs in the sequence ladder decreased, reflecting a lower frequency of priming during the labeling reaction. Labeling was almost undetectable at 0.5 μM, even though hexamer was in 15-fold molar excess over template.

DNA concentration. Labeling intensity decreased only below 150 ng (7 nM) of M13 DNA, and sequence ladders remained easily detectable at 15 ng (0.7 nM). The patterns of labeling were similar whether the SSB concentration remained constant or was diluted in parallel with the DNA. Labeling was also similar when the M13 DNA was diluted in the presence of denatured T7 DNA, keeping a total of 0.6 μg of DNA in each reaction mixture. Although T7 DNA is 12-fold more complex than M13 DNA, the density of sites complementary to individual hexamers was largely unchanged because M13 DNA contains 1.5 times as many hexamers as the 4096 that are possible.

Reaction temperature. Specific priming at the hexamer string decreased markedly as reaction temperature increased above 0°C; with 5 μg of SSB per reaction, considerable priming remained at 5°C, much less at 10°C, and little at 15°C. The average length of the DNA chains in the sequence ladder also increased with temperature, consistent with a reduced frequency of priming. This decrease in priming apparently reflects competition between the hexamer string and SSB for binding the template DNA because priming decreased little if at all over this temperature range in the absence of SSB.

Priming by hexamer strings in M13 DNA. To test the generality of priming by hexamer strings in the presence of SSB, we made 20 different strings containing from 4 to 18 contiguous hexamers complementary to 15 different regions of M13 or M13mp18 DNA. The hexamers in the first two strings (A and B) are listed in Table 1. Most subsequent strings were built around an available hexamer whose complement was found at more than one site in M13 DNA. Thus, hexamer B1 is the same as A4. Altogether, these 20 strings contained 119 different hexamers of widely different composition.

In the first three strings, A, B, and C, we tested each single hexamer and every contiguous string of two to six hexamers for the ability to prime sequencing reactions in the presence and absence of SSB. In the remaining strings, most contiguous strings of two, three, and four hexamers were tested in the presence of SSB, for a total testing of 63 strings of two hexamers, 70 strings of three, and 55 strings of four in the 19 strings that excluded string D. The sequence ladders in Fig. 2 illustrate the principal findings.

In the absence of SSB, sequence ladders were generally weak and ambiguous, whether primed by individual hexamers or by any of the strings of hexamers. Exceptions include hexamer A6, which primed moderately well at its single complementary site in M13 DNA, and the contiguous pair B3–B2 (Fig. 2B), which primed selectively as a pair even though neither hexamer by itself primed significantly at this site (and each is complementary to three additional sites in M13 DNA).

In the presence of SSB, priming by individual hexamers was almost always strongly suppressed. Priming by most contiguous pairs of hexamers was also suppressed, but about 40% of those tested primed specifically as a pair in the presence of SSB. Examples include the weakly priming A6–A5 pair and the strongly priming B3–B2 pair in Fig. 2. Most strings of three or four hexamers were stimulated by SSB to prime intensely and specifically at the position of the string (Fig. 2). Strings of more than four hexamers did not seem to offer any advantage.

The sequence ladders obtained in the presence of SSB were usually primed exclusively by the hexamer at the 3′ end of the string, as shown by a shift of the sequence ladder by six nucleotides with the addition or subtraction of a hexamer

Table 1. Sequences of hexamers used in strings A and B in M13 DNA.

Number*	Sequence 5′ to 3′	Position in M13 DNA†	Other sites‡
A6	ACCCCC	1194	0
A5	AGCGAT	1188	1
A4	TATACC	1182	1
A3	AAGCGC	1176	2
A2	GAAACA	1170	6
A1	AAGTAC	1164	3
B6	TACCTT	830	4
B5	ATGCGA	824	0
B4	TTTTAA	818	9
B3	GAACTG	812	3
B2	GCTCAT	806	3
B1	TATACC	800	1

*Numbers decrease in the 5′ to 3′ direction, so that the hexamer with the lowest number is at the 3′ end of a string. †The nucleotide in M13 DNA that is complementary to the 3′ nucleotide of the hexamer at the position of the string. Priming proceeds toward lower numbers. ‡Number of sites in M13 DNA complementary to the hexamer at positions outside of the string.

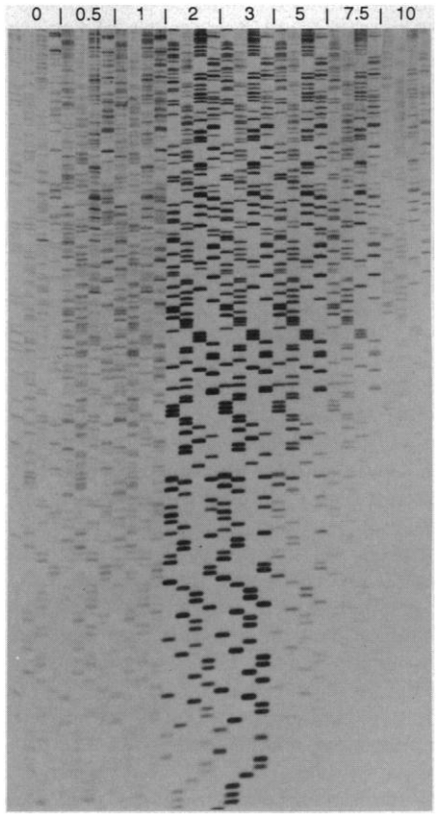

Fig. 1. Titration with SSB. Sequencing reactions were primed by hexamers A4–A1 on 0.7 μg of M13 DNA in the presence of different amounts of *E. coli* SSB (micrograms used per reaction are given above each set). The order of lanes in the sequence ladders is GATC.

at the 3′ but not the 5′ end of the string (Fig. 2). However, in many cases priming by one or two internal hexamers was observed, producing superimposed sequence ladders six nucleotides apart. Substantial amounts of such double priming are evident in the patterns generated by B3–B1 and B4–B1 in Fig. 2B. In a few cases, the 3′ hexamer of the string primed weakly if at all, and priming was predominantly or almost exclusively at the next hexamer. Thirteen of the 70 strings of three (19%) and 16 of the 55 strings of four (29%) had enough double priming that reading the sequence ladder was difficult. Increasing the concentration of the 3′ hexamer was only marginally effective in reducing double priming in the few cases tested.

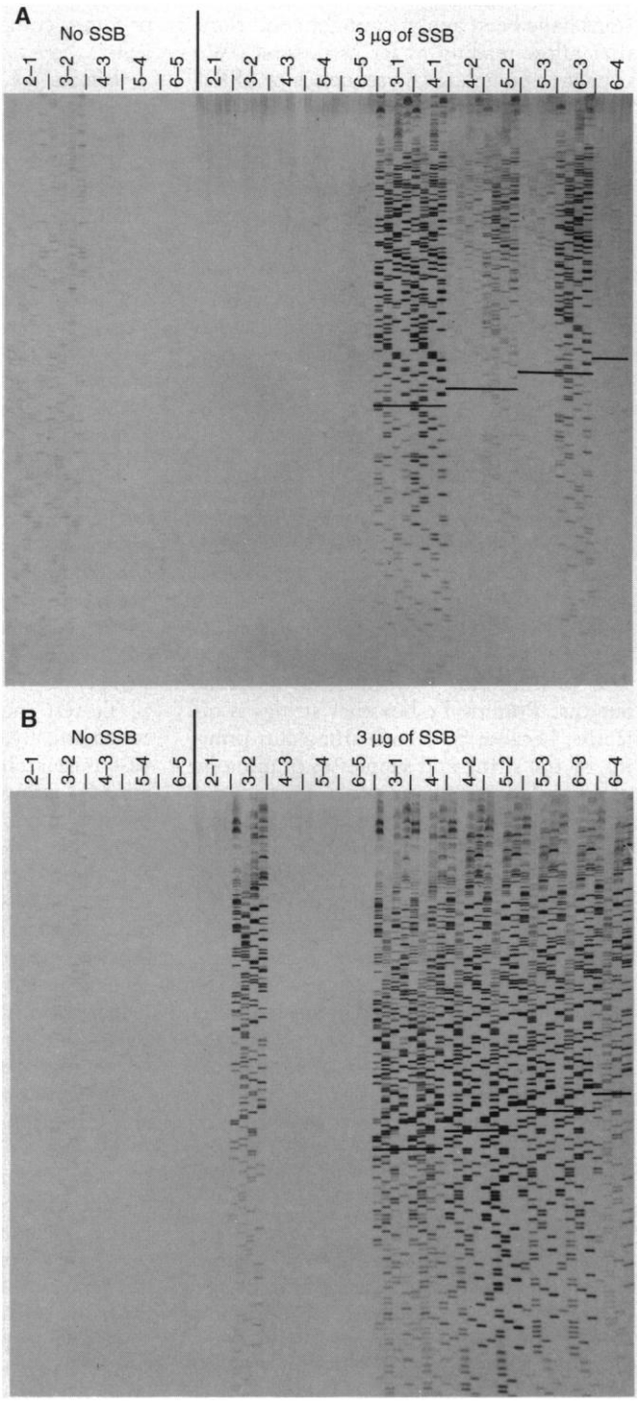

Fig. 2. Priming by contiguous hexamers (**A**) in string A and (**B**) in string B. Sequencing reactions were primed on 0.7 μg of M13 DNA in the absence or presence of 3 μg of SSB, as indicated. All sets of two, three, or four contiguous hexamers are shown, with the hexamer numbers (Table 1) given above each sequence ladder. Equivalent positions in the sequence ladders primed by strings of three or four hexamers are indicated by the horizontal lines in the figure. The order of lanes in the sequence ladders is GATC.

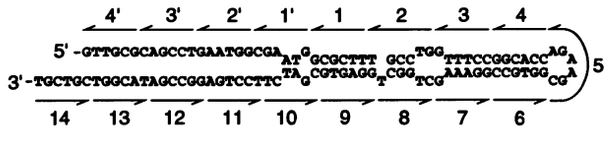

Fig. 3. Potential base-paired structure in M13 DNA in relation to the hexamers of string D (numbered above and below the sequence). Nucleotides 6422 to 6529 of M13 DNA are shown; the 11-base palindrome is centered on nucleotides 6472 to 6473.

Two other problems interfered with the determination of sequence primed by hexamer strings: weak priming and priming at secondary sites outside of the string. A few strings primed so weakly that we required a 2- to 5-day exposure of the autoradiogram to read the sequence. Relatively weak priming in several strings of three hexamers, including A4–A2, A5–A3, and A6–A4 in Fig. 2A, increased substantially when a fourth hexamer was added. Significant interference by priming at secondary sites was observed in three cases affecting 11 strings of three or four hexamers.

Ladders with sequences that could be read unambiguously and without difficulty were obtained from 49 of the 70 strings of three hexamers (70%) and 33 of the 55 strings of four hexamers (60%). At least some sequence information could be obtained from many of the other ladders as well. Overlapping ladders primed by two or more hexamers in a string, the most frequent problem, might be resolvable by computer analysis to generate reliable sequence information.

Interference by base pairing in template DNA. Hexamer string D is a special case not included in the above analysis. This string was built from a site complementary to hexamer A3 at nucleotides 6446 to 6451 in M13mp18 DNA and initially contained six hexamers, D6–D1. Unlike the other strings, no combinations of these six hexamers primed a sequence ladder at the position of the string. Examination of the template sequence in this region revealed a perfect 11-base palindrome, plus the potential for additional base pairing (Fig. 3). Extending the string, we found that D10–D7 also did not prime, but D14–D11 on the 3′ side of the palindrome and D1′–D4′ on the 5′ side both primed specifically. Thus, competition from base pairing in the template DNA seems to prevent priming by hexamer strings.

SSB is thought to remove most base-paired structures from single-stranded DNA (6), but the structure at string D may be too stable to be removed by SSB under priming conditions at 0°C. Higher temperatures should favor stabilization of the unfolded structure by SSB, but heating the mixture of primers, DNA, and SSB to temperatures as high as 90°C before attempting sequencing reactions at 0°C did not promote specific priming by hexamer string D4–D1

(SSB is thermostable). Perhaps the structure reforms rapidly after cooling in the presence of SSB, or perhaps the displacement of SSB to form the hexamer string allows the structure to form and displace the hexamers. Increasing the hexamer concentration from 5 to 50 μM was not sufficient to promote specific priming by string D4–D1.

Although strong local base pairing in the template DNA seems to prevent priming by hexamer strings, inspection of the sequence of template DNA in the region where priming is desired should allow most such problem areas to be identified and avoided.

Priming by hexamer strings in denatured double-stranded DNAs. M13 viral DNA is naturally single-stranded, but hexamer strings also prime specifically on denatured double-stranded DNAs (*11*), apparently unaffected by the presence of the complementary strand. Hexamer string A4–A1 primed on denatured linear or supercoiled forms of double-stranded M13 DNA, and good sequence ladders were obtained from a heat-denatured 2.1-kbp polymerase chain reaction (PCR) product from T7 gene 5.

To test a DNA in the size range of cosmid DNAs, we tried priming in three different regions in T7 DNA, a linear double-stranded DNA of 39,937 bp of known sequence (*12*). Because of its greater length, 1 μg of T7 DNA per reaction provided only 3.8 nM of unique priming sites, a concentration where the intensity of sequence ladders primed on M13 DNA was reduced but still substantial. In each region of T7 DNA, strings of three or four hexamers primed specific sequence ladders that were usually readable after overnight exposure of the autoradiograms. Some hexamer strings primed weakly or primed double or triple ladders, problems similar to those observed in M13 DNA. Reducing the size of the molecule that contains the priming site to 20, 14, 7, or 4 kb by cutting the T7 DNA with different restriction enzymes had no detectable effect on the sequence ladders obtained.

To test primer walking with hexamer strings on a cosmid-sized DNA of unknown sequence, we used the DNA of LPP-1, a T7-like cyanophage (*13*) whose DNA we are in the process of sequencing. We constructed 27 strings of four hexamers designed to prime within blocks of known sequence and to extend them into unknown regions. We chose most of these strings in such a way that they would contain at least some hexamers predicted to have relatively high affinities for template DNA (*14*). Twenty-four of these strings (89%) gave readable sequence ladders whose quality ranged from fair to excellent. *Escherichia coli* SSB alone was successful in about half of these cases, and addition of T7 gene *2.5* protein produced readable ladders in the others. The longest sequence read was 461 nucleotides, but the sequencing reactions were not optimized for long reads, and only some reactions have been analyzed under conditions that allow reading as far as possible. We expect to complete the sequence of LPP-1 DNA entirely by primer walking with hexamer strings.

Nucleotide sequences of hexamers in priming strings. Altogether, more than 200 hexamers with a wide range of sequence and composition were used in strings that primed successfully in 45 regions in three different template DNAs. The only consistent correlation we made between hexamer sequence and poor performance was that some hexamers containing only T and A were ineffective as the 3′ or 5′ hexamer in a string, although they functioned well internally. TAATAA, ATTATT, and TTAATT showed this behavior, but TTTTAA primed effectively as the 3′ hexamer in B6–B4 (Fig. 2 and Table 1). Further experience may uncover other limitations in priming ability, but it seems clear that a wide range of hexamer combinations is effective.

SSB inhibition of primers of different lengths. Priming by hexamer strings is effective because SSB both stimulates priming by the string and suppresses priming by individual hexamers at other sites in the DNA. To test the ability of SSB to suppress priming by oligonucleotides of different lengths, we synthesized a nested set of oligonucleotides complementary to M13 DNA at the position of string B, all having the same 3′ nucleotide as hexamer B2. Although B2 has four complementary sites in M13 DNA, each of the longer oligonucleotides has only one perfectly complementary site. Priming by 5 μM oligonucleotide on 0.6 μg of M13 DNA was tested in the presence of 0, 2, or 5 μg of SSB under standard reaction conditions, equilibrating 1 hour before labeling.

In the absence of SSB, maximum priming efficiency was reached at primer lengths of nine or greater; priming by the octamer or heptamer was weaker, and priming by the hexamer was very weak. Addition of 2 μg of SSB only slightly suppressed priming by the heptamer and seemed to enhance priming by oligonucleotides of lengths 8 to 11. Addition of 5 μg of SSB strongly suppressed priming by oligonucleotides of up to length 8, moderately suppressed priming by those of length 9 and 10, and did not suppress priming by those of length 11 and longer. An increase of the temperature to 22° or 37°C increased the length of primer needed for maximum efficiency by only one nucleotide or so and only moderately increased the suppression by SSB. These results suggest that interaction between contiguous hexamers in a string need not be great to drive the establishment of priming complexes in the presence of SSB.

Effects of mismatches. We detected about a dozen instances of secondary priming in the presence of SSB where we were able to pinpoint the site of priming. In almost every case, the site of secondary priming was a perfect complement to one of the hexamers in the string, which was flanked by one or more contiguous but mismatched pairing sites for the same hexamer or another hexamer in the mixture. Two cases were also observed where a second sequence ladder was primed six nucleotides past the 3′ end of a string and two cases where a second ladder was displaced five nucleotides. When the ladder was shifted by six bases, one of the hexamers present in the mixture could pair with the six bases immediately past the 3′ end of the string with a single mismatch; when shifted by five bases, the five nucleotides at the 3′ end of the hexamer could pair with the five bases immediately past the 3′ end of the string.

To test the effects of mismatches more systematically, we tested all possible single-base mismatches in each of the three hexamers of string A3 to A1 for their effect on priming under standard conditions (3 μg of SSB). No hexamer with any mismatch in A1, the 3′ hexamer of the string, stimulated priming significantly. On the other hand, 8 of the 18 possible mismatches in A2 and 12 of 18 in A3 primed correct sequence, the intensity of the sequence ladder ranging from weak to moderate. Extending the analysis to other strings, all 18 mismatches in B4 primed correct sequence in combination with hexamers B5 and B3, again ranging from weak to moderate. All 18 mismatches in the middle hexamer of still another string primed sequence ladders almost as intensely as the perfectly matched hexamer.

These observations suggest that a range of contiguous hexamer strings with limited mismatches can prime sequence ladders to some extent in the presence of SSB. Thus, secondary sites where hexamers can by chance cooperate to prime strongly enough to interfere with reading a primary sequence ladder will occur more frequently than if only perfect matches could prime significantly. Our experience suggests that the frequency of such secondary sites is low enough to cause little problem in primer walking with hexamer strings on templates as large as 40 kbp. Such sites

might be expected to become a significant problem at DNA lengths in the range of hundreds of thousands to millions of base pairs.

Priming with strings of pentamers or heptamers. If strings of contiguous pentamers would prime sequencing reactions specifically, the size of the library needed for efficient sequencing would be one-fourth that needed for hexamers. Priming was tested with various combinations of a string of seven contiguous pentamers complementary to M13 DNA at the position of the A string of hexamers. Only weak and ambiguous priming was seen under conditions where hexamer strings primed intensely. Decreasing the reaction temperature to −2.5° or −5°C or increasing the primer concentration tenfold to 50 μM did not provide much improvement. Pentamers appear unlikely to be useful for priming in strings.

Heptamers might be useful for sequencing but probably less so than hexamers. Not only would a larger library be required, but several experiments indicated that priming by individual heptamers is not as well suppressed. On the other hand, a heptamer flanked by two hexamers primed clean sequence ladders in the presence of SSB, and such strings might be more effective than strings of hexamers in some cases. Addition of only a few hundred heptamers to a hexamer library could provide a useful density of such strings.

Mechanism of priming by hexamer strings. SSB is thought to bind single-stranded DNA by wrapping the DNA around an octamer of SSB, protecting about 145 nucleotides from digestion by deoxyribonuclease (DNase) but leaving an average of about 30 unbound nucleotides between DNA-octamer beads (*10*). These unbound nucleotides may be the sites of initial binding of oligonucleotides to the DNA. Random movement of the octamer beads along the DNA strand might expose all potential binding sites in the DNA and also displace weakly bound oligonucleotides before they can prime DNA synthesis.

Masking of individual priming sites is important in the success of priming by hexamer strings and presumably increases the effective concentration of hexamers available for binding in the string. Specific priming by strings of hexamers was only occasionally apparent in the absence of SSB.

Base-stacking interactions between adjacent hexamers must stabilize the binding of strings of hexamers to SSB-coated DNA because priming is not stimulated unless the binding sites are adjacent in the template. Another instance where base stacking between adjacent oligonucleotides enhanced binding has been reported by Khrapko and co-workers (*15*), who analyzed binding to a 17-mer in the absence of any SSB.

Prospects for DNA sequencing. Primer walking with strings of three or four hexamers has the potential to improve the efficiency of genomic sequencing by at least an order of magnitude. The burden of subcloning, multiple template preparation, repetitive sequencing, and assembly of random blocks of sequence in the shotgun strategy would be eliminated, as would other inconvenient and expensive methods to traverse a few kilobase pairs of DNA. Both strands of DNA, up to at least 40 kbp, could be completely sequenced in the minimum number of sequencing reactions. Error rates should be low because optimal priming sites can be selected to resolve ambiguities. Most hexamers are likely to be usable, and the success rate should increase from the 60 to 90% we obtained as more is learned about selection of the hexamer strings likely to prime well.

A major obstacle to using this method is the unavailability of suitable libraries of hexamers. However, 1 μmol of hexamer supplies enough material for 20,000 primings, so the intrinsic cost of hexamers is less than a penny a reaction at current commercial prices. A complete library of 4096 hexamers could be divided into 200 working libraries, each of which could supply hexamer strings for more than 100,000 sequence reactions. In this way, working libraries could become accessible and affordable to many laboratories. Libraries containing as few as 1500 hexamers could also be effective (*16*).

If hexamer strings could prime sequencing reactions with a DNA polymerase and SSB that are both thermostable, repeated cycles of synthesis and denaturation might be used to obtain sequence ladders from much smaller concentrations of template DNA, perhaps allowing direct sequencing of DNAs much larger than cosmids. Experiments indicate that SSB can stimulate template-directed ligation of strings of short oligonucleotides, which might be useful for cycle sequencing or in other applications.

The entire process of sequencing by primer walking with hexamer strings could be automated and put under computer control. A battery of templates plus an array of primers would allow sequencing reactions to be assembled rapidly enough to saturate any current or easily foreseeable means of analysis. Sequencing machines based on these principles could operate with little requirement for skilled human intervention and could provide the capacity and efficiency needed for the success of the Human Genome Project.

REFERENCES AND NOTES

1. *Understanding Our Genetic Inheritance. The U.S. Human Genome Project: The First Five Years FY 1991–1995* (U.S. Department of Health and Human Services and U.S. Department of Energy, Washington, DC, 1990); T. Hunkapiller, R. J. Kaiser, B. F. Koop, L. Hood, *Science* **254**, 59 (1991).
2. F. W. Studier, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6917 (1989).
3. W. Szybalski, *Gene* **90**, 177 (1990).
4. B. Weiss, *Methods Enzymol.* **21**, 319 (1971).
5. S. Tabor and C. C. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4767 (1987); *J. Biol. Chem.* **264**, 6447 (1989); the T7 polymerase used was Sequenase 2.0 (U.S. Biochemical Corp.).
6. J. W. Chase and K. R. Williams, *Annu. Rev. Biochem.* **55**, 103 (1986); R. R. Meyer and P. S. Laine, *Microbiol. Rev.* **54**, 342 (1990).
7. P. M. G. F. van Wezenbeek, T. J. M. Hulsebos, J. G. G. Schoenmakers, *Gene* **11**, 129 (1980); R. Ebright, Q. Dong, J. Messing, *ibid.* **114**, 81 (1992).
8. Hexamers were synthesized (1-μmol scale) on a DNA synthesizer (MilliGen 8750, MilliGen/Biosearch, Novato, CA), deblocked, lyophilized, and dissolved in water. Further purification is not needed although more highly purified hexamers were used initially and in much of the work reported here.
9. We performed labeling at limiting dNTP concentrations by adding 6 μl of an ice-cold solution containing 2.5 units of Sequenase; 313 nM each of dCTP, dGTP, and dTTP; and ~3.5 μCi of α-^{35}S–labeled dATP (a slight molar excess over the unlabeled dNTPs) in 10 mM dithiothreitol, 10 mM tris-HCl (pH 7.5), and 0.1 mM EDTA. The termination reaction was for 5 min at 37°C in the standard Sequenase protocol.
10. S. Chrysogelos and J. Griffith, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5803 (1982); J. D. Griffith, L. D. Harris, J. Register III, *Cold Spring Harbor Symp. Quant. Biol.* **49**, 553 (1984).
11. Several procedures gave equivalent results in denaturing double-stranded DNAs for sequencing: 2 min at 100°C before adding reaction buffer; 2 min at 100°C in the presence of SSB; or 2 min at 100°C or 5 min at room temperature in 50 mM NaOH and 50 mM NaCl, followed by neutralization at 0°C. Usually, DNA was denatured by heating.
12. J. J. Dunn and F. W. Studier, *J. Mol. Biol.* **166**, 477 (1983).
13. L. A. Sherman and R. Haselkorn, *J. Virol.* **6**, 841 (1970).
14. K. J. Breslauer, R. Frank, J. Blöcker, L. A. Marky, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3746 (1986); R. S. Quartin and J. G. Wetmur, *Biochemistry* **28**, 1040 (1989).
15. K. R. Khrapko *et al.*, *FEBS Lett.* **256**, 118 (1989); K. R. Khrapko *et al.*, *DNA Sequence* **1**, 375 (1991).
16. $P = 1 - (1 - f^S)^N$ is the probability, P, of finding at least one string of S hexamers in a stretch of N potential priming sites in a template DNA of random sequence, where f is the fraction of total hexamers in the library (that is, the number of hexamers in the library divided by 4096). A library of 1500 hexamers would have a 99% chance of providing at least one string of three hexamers within a stretch of 100 potential priming sites and an 84% chance of providing a string of four.

17. We thank W. Crockett, M. Randesi, and L. Butler for providing oligonucleotides and for help with sequencing; S. Zhang for computer support; B. McGrath for cloning *E. coli* SSB; and M. Blewitt for providing M13 DNA and T7 gene *2.5* protein. Supported by the Office of Health and Environmental Research of the U.S. Department of Energy.

20 July 1992; accepted 15 October 1992