Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4241–4245, May 1993
Biochemistry

DNA sequencing: Modular primers assembled from a library of hexamers or pentamers

(base-stacking/primer extension/oligonucleotide annealing/primer walking/walking primer)

LEV E. KOTLER, DINA ZEVIN-SONKIN, IRINA A. SOBOLEV, ALEXANDER D. BESKIN, AND LEVY E. ULANOVSKY*

Department of Structural Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

Communicated by Charles R. Cantor, January 8, 1993 (received for review November 14, 1992)

ABSTRACT Here we report a striking effect displayed by "modular primers," which consist of hexamer or pentamer oligonucleotide modules base-stacked to each other upon annealing to a DNA template. Such a combination of modules is found to prime DNA sequencing reactions uniquely, unlike either of the modules alone. We attribute this effect in part to the increase in the affinity of an oligonucleotide for the template in the presence of an adjacent module. All possible pentamer (or hexamer) sequences total 1024 (or 4096) samples, a manageable size for a presynthesized library. This approach can replace the synthesis of primers, which is the current bottleneck in time and cost of the primer walking sequencing, and can allow full automation of the closed cycle of walking.

In most strategies of DNA sequencing, such as random "shotgun" or nested deletion subcloning, the major bottleneck in efficiency is not the stage of sequencing *per se* but, rather, the "front end" (subcloning and template preparation) and/or the "back end" (integration of sequences from individual shotgun runs). In contrast, sequencing by primer walking minimizes the front and back end problems, using the same template many times in a processive manner, with a new primer for each run. But it presents its own bottleneck in terms of time and cost—the primer synthesis step. This step produces a huge excess, of about six orders of magnitude, of the synthesized primer (0.2–1.0 μmol) over the amount needed for a typical sequencing reaction (0.5 pmol). Studier (1) suggested to eliminate the need for synthesis of walking primers by building a library of presynthesized short oligonucleotides (8-mers and/or 9-mers, the shortest primers expected to be unique for plasmid size templates), but the size of such a potential library, even an incomplete one, is problematic. Szybalski (2) proposed that two hexamers be ligated on the template into a unique 12-mer primer, a 64-fold reduction in the (6-mers vs. 9-mers) library size, but complete ligation of hexamers on the template is yet to be shown.

The present article describes a remarkable manyfold increase in the sequence specificity of priming by short oligonucleotides, such as hexamers or pentamers, when tandemly annealed to the DNA template, as compared to each of them separately. In DNA sequencing reactions this phenomenon results in unique priming by what we term a "modular primer," a tandem string of two or three short oligonucleotides, with no ligation required whatsoever. In contrast, the same pentamers or hexamers show nonunique, multiple priming when used "one at a time" (without adjacent partners). This effect is interpreted here as resulting in part from the increase in the affinity of the oligonucleotides for the template caused by their base-stacking, as they anneal to the template next to each other, in comparison with their annealing alone, with no neighbors. Modular primers showed a 91% success rate in sequencing reactions, which is comparable to the performance of conventional 17-mer primers.

A complete oligonucleotide library of all possible pentamer or hexamer sequences comprises only 1024 or 4096 samples, respectively, and would remove the need for synthesis of new primers for each walking step. Not only time but also cost per walking step is thus reduced, since the scale of oligonucleotide synthesis is sufficient to produce thousands of libraries for users. Furthermore, modular primers would allow full automation of the closed cycle of walking sequencing, which is currently impossible because of the primer synthesis stage. With processive sequencing of continuous DNA templates as long as tens of kilobases at hand, subcloning artefacts, such as DNA rearrangements, and unclonable segments would be minimized.

MATERIALS AND METHODS

Gel-purified oligonucleotides were supplied by the synthesis service of the Weizmann Institute of Science and Biotechnology General (Israel). The front modules and the control 15-mers and 18-mers were end-labeled (3) by T4 polynucleotide kinase (Boehringer Mannheim) with [γ-^{32}P]ATP (6000 Ci/mmol; 1 Ci = 37 GBq; DuPont/NEN). The sequencing of M13mp18 single-stranded DNA template (United States Biochemical) was performed with reagents from the sequencing kits for Sequenase version 1.0 (United States Biochemical) and *Bst* polymerase (Bio-Rad, no. 170-3404). For *Bst* polymerase, the ddNTP concentrations were as specified in the "extended sequencing protocol" of the kit, except that the ddNTP/dNTP ratios were as in the "standard sequencing protocol." For both polymerases, the annealing reaction volume was 12 μl, containing 0.5 pmol single-stranded M13mp18 template (United States Biochemical), 0.5 pmol of ^{32}P-end-labeled front module (or 15-mer control), middle and back modules as specified in the figure legends, and 2.0 μl of buffer stock from the Sequenase kit containing 200 mM Tris·HCl (pH 7.5), 100 mM $MgCl_2$, and 250 mM NaCl or 2.0 μl of buffer stock from the *Bst* polymerase kit containing 100 mM Tris·HCl (pH 8.5) and 100 mM $MgCl_2$. For annealing, the beaker with floating microtubes was boiled for 5 min and cooled to $<$10°C as specified in the figure legends. The tubes were spun down at 4°C and kept on ice. Dithiothreitol (for Sequenase) and the enzyme were added and the reaction volume was split into four dideoxy termination mixtures, as described in the relevant kit manual (with no labeling step), except that the reagents were precooled and added on ice. The temperature regimes of the extension reactions are specified in the figure legends. Half of the reaction with modular primers and one-tenth of that with control primer were electrophoresed on 6% sequencing gel.

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.

RESULTS

Modular Primer Concept. A modular primer consists of hexamers or pentamers (modules) that are not connected covalently but base-stack together upon their annealing to the template. We hypothesized (for reasons described in *Discussion*) and indeed found that modular primers can prime strand extension uniquely, as if their modules were linked covalently. Their priming specificity depends on the existence of one and only one site in the template where the modules can anneal next to each other.

The first modular primer we tried consisted of two hexamer modules (see Fig. 1). The "front" module, 3′-ACGCAA-5′ (the one to be extended), was end-labeled with ^{32}P, whereas the "back" module, 3′-AGGAGC-5′, in a large molar excess (5 pmol for Fig. 1), was not. When annealed to the single-stranded M13mp18 template, they base-stack to each other at position 2891 (in GenBank numbering), where the template reads 5′-TGCGTTTCCTCG-3′.

The front module has three complementary sites in single-stranded M13mp18. Only one of them occurs (by design) next to the annealing site of the back module at the targeted position 2891. Indeed, when the front module is used alone, it shows an unreadable band pattern with Sequenase (lanes marked "6-MER" in Fig. 1*a*) typical for multiple priming. In contrast, when the adjacent back hexamer is included, a perfectly readable band pattern (lanes "6+6") emerges identical to that of the 15-mer control primer. *Bst* polymerase gave a similar effect (Fig. 1*b*), except that the front module alone showed no detectable priming. This indicates that the modular primer phenomenon is not limited to a particular DNA polymerase.

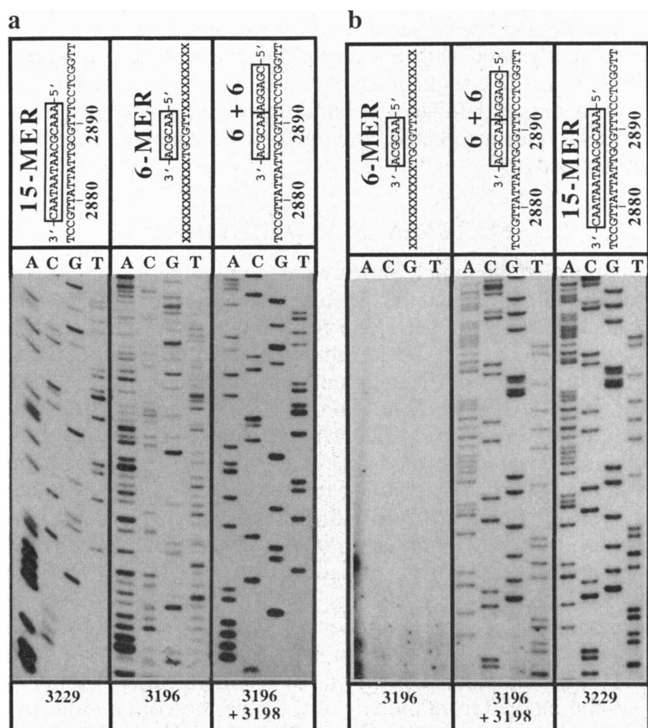

FIG. 1. Simple modular primer. Two hexamer modules, 3′-ACGCAA-5′ (front module) and 3′-AGGAGC-5′ (back module), can base-stack at position 2891 of the single-stranded M13mp18 template when annealed to it. (*a*) Sequencing with Sequenase. (*b*) Sequencing with *Bst* polymerase. Lanes are marked in accordance with the primers: front module alone, "6-mer;" front and back modules together, "6 + 6;" control, "15-mer." The amounts of ^{32}P-labeled front module and unlabeled back module (if relevant) were 0.5 pmol and 5.0 pmol, respectively, except the reaction in the 6-mer lanes of *b* contained 2 pmol of primer, a 4-fold excess over the template. The temperature regimes were as follows. (*a*) For annealing, the beaker with floating microtubes was boiled for 5 min, cooled to 35°C on the bench in 40 min while stirring, and placed into ice water, still stirring, its temperature reaching 10°C in about 10 min. The extension reactions were gradually warmed from 4°C to 37°C in 20 min and stopped. (*b*) After 5 min of boiling the annealing samples were cooled from 100°C to <10°C in 10 min. The extension reactions were gradually warmed from 4°C to 65°C in 20 min. The identification numbers of oligonucleotides are shown under each gel.

Six Tandem Hexamer Modules Tested. Fig. 2 shows the first walking step, 223 bases downstream, where six hexamers were synthesized complementary to six tandem annealing sites to try several shifted modular primers. These six hexamer modules except the last one were selected to have purine–purine-type base-stacking between them. We found that modular primers consisting of three tandem hexamer modules, 6+6+6 structure, generally worked better than two-module 6+6 structures (see Fig. 2). All four modular primers of the 6+6+6 type gave perfectly readable band patterns (Fig. 2 *b–e*) as clear as those of the control 18-mer primer. (The signal strength achieved with modular primer was on the average about 20% of that achieved with 15/18-mer controls.) Whereas Fig. 2 *b–e* show these reactions with Sequenase, Fig. 2*f* shows reactions with *Bst* polymerase primed by modular structures 6, 6+6, 6+6+6, 6+8.

Front Module Tested in Each of Its Four Sites in M13mp18. As a further test of the modular primers, one of these hexamer modules, 3′-AAGGGA-5′, was used as a common front module for four 6+6+6 type primers in each of all four of its complementary sites in single-stranded M13mp18 template (Fig. 3). Three of them worked well. They happened to have purine–purine base-stacking in the first and the second junctions between the hexamers in the 6+6+6 structure, whereas the fourth primer, in position 3104, which produced an unreadable band pattern, had non-purine–purine stackings in both junctions. This experiment showed that the same end-labeled module could be extended by polymerase at one site or another, depending on the presence of adjacently annealed auxiliary modules.

Primers of 6+6+6 Type Showed 91% Success Rate. The purine–purine base-stacking was found superior at other sites as well. Altogether, throughout the M13 genome, we tried 31 modular primers of the 6+6+6 type (see Table 1). Of these, 22 had purine–purine stacking between the front and the middle modules. Of those, 20 gave readable band patterns with Sequenase (91%), and 17 did so with *Bst* polymerase. Such success rates are not much different from those of conventional 17-mer primers. Non-purine–purine base-stacking types between the front and the middle modules in the 6+6+6 structure were also checked (see Table 1). Only 3 of 9 such primers worked well with Sequenase, and only 2 did so with *Bst* polymerase. The purine–purine base-stacking is known to have a higher energy of interaction than the other three combinations (4). The second junction, between the middle and the back modules, seems less demanding as to its base composition. Indeed, of the 12 tested 6+6+6 primers with purine–purine in the first junction and non-purine–purine in the second junction, 10 were successful with Sequenase, and 9 were with *Bst* polymerase.

The modular primers differed greatly in the resulting intensities of their sequencing reactions (signal strength). The secondary structure of the template around the priming site seems crucial for the modular primer efficiency. It was found that the only two 6+6+6 primers that were unsuccessful with Sequenase (of the 22 with purine–purine first stacking, see Table 1) were both located on one arm of an imperfect hairpin, centered at position 7148. When we blocked the other arm of the hairpin by annealing to it a complementary 18-mer, both previously unsuccessful 6+6+6 primers worked well. Therefore, when criteria are developed to avoid unfavorable secondary structures, the modular primer effi-

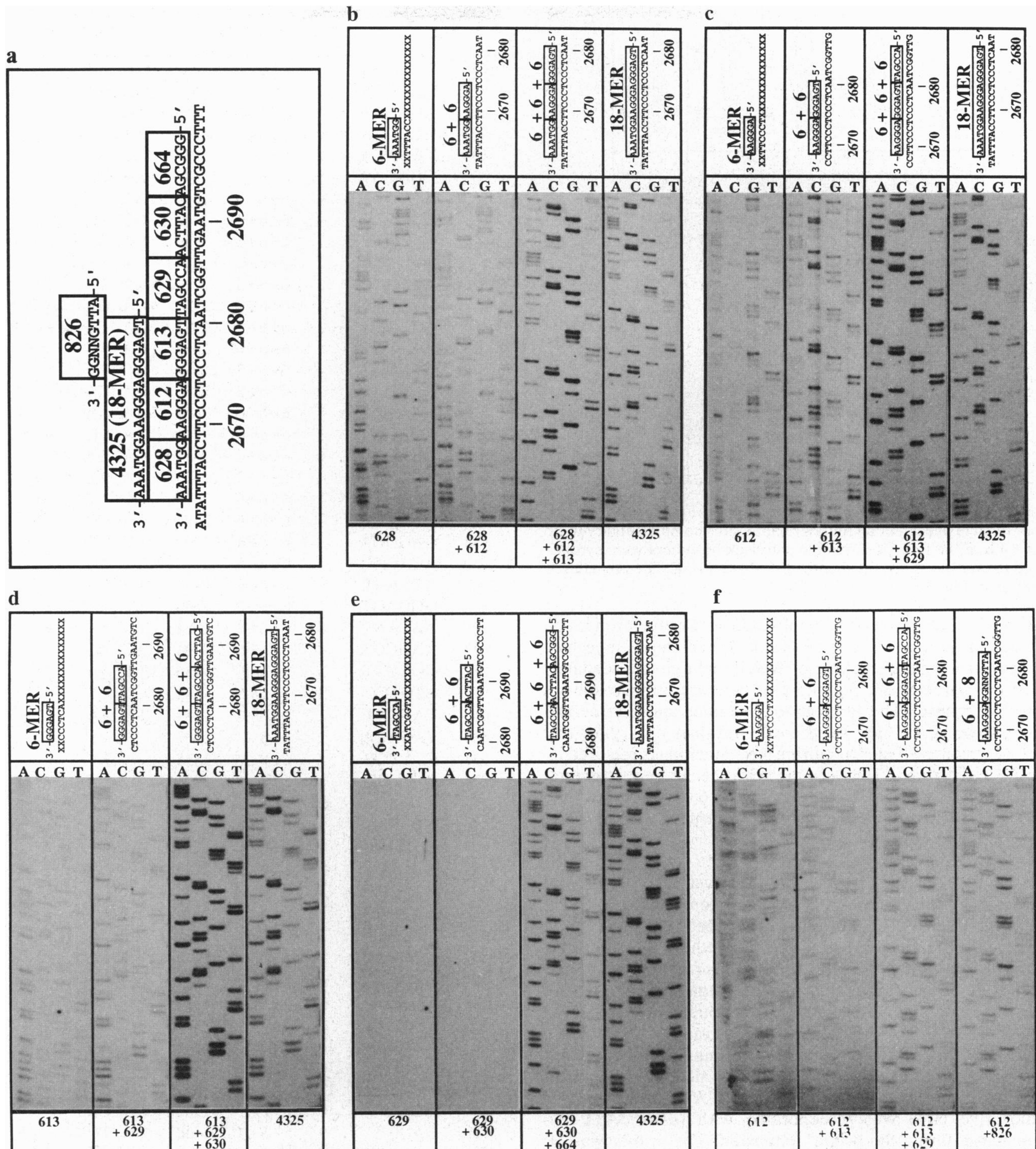

FIG. 2. First walking step. (*a*) Six hexamer modules were synthesized complementary to six tandem annealing sites in the single-stranded M13mp18 template. Modular primers were constructed from these modules as shown on top of each gel. (*b–e*) DNA sequencing reactions performed with Sequenase. (*f*) Reactions with *Bst* polymerase (procedures as in Fig. 1*b*). In *b–f* the annealing was as in Fig. 1*b*, except that the amounts of the middle and back modules were 40 pmol for hexamers and 200 pmol for the degenerate octamer; the amount of the front module in 6-mer lanes was 0.5 pmol for Sequenase (*b–e*) and 2.0 pmol for *Bst* polymerase (*f*). In *b–e* after precooled reagents and Sequenase were added on ice, the microtubes were transferred immediately into a water bath kept at 37°C (*b, c,* and *e*) or 48°C (*d*). Note that the retardation shifts of a few bases between the band patterns of modular and control primers match the staggers between the 5′ ends of the front module and those of the control primer shown on top of the gels.

ciency can be optimized by shifting the priming site. Parameters for further optimization may include primer structure, its G/C content, stacking types, polymerase nature, temperature and salt regimes of annealing and polymerization, etc.

The 5+5+7 and 6+8 Structures. Apart from 6+6 and 6+6+6 type primers, we also tested other structures (see Table 1), such as 5+5+7 and 6+8, where the heptamers and the octamers contained two degenerate base positions each (Figs. 4 and 2*f*). The libraries of such degenerate heptamers and octamers are the same size as those of pentamers and hexamers, respectively. Of six 5+5+7 primers with purine–purine in both stackings four worked well with Sequenase,

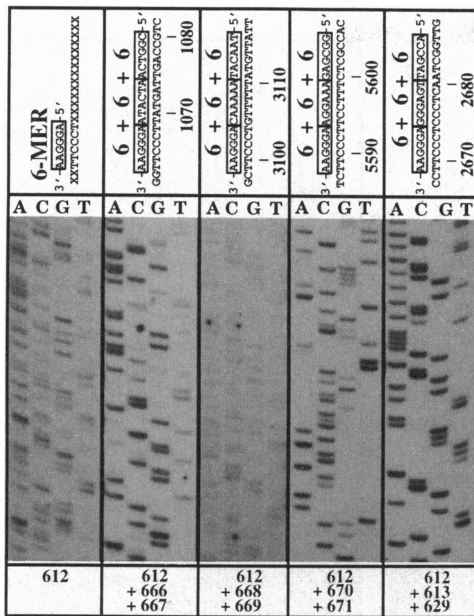

FIG. 3. One hexamer (the front module of Fig. 2*c*) used as the front module in each of all four M13mp18 sites complementary to it. For each of the four sites, middle and back hexamers were synthesized to anneal next to the identical six-base string. The procedures were as in Fig. 2*c*.

but only one did so with *Bst* polymerase. Twelve primers of the 6+8 structure were tried (Table 1). Of these, nine had purine–purine base-stacking of which eight gave a readable band pattern with *Bst* polymerase, whereas only six did so with Sequenase. The libraries of hexamer and degenerate octamer would be half the size of a hexamer library for the 6+6+6 structure if purine–purine base-stacking is required in the 6+8 structure. The reasons why the 5+5+7 structure worked better with Sequenase, while 6+8 preferred *Bst* polymerase, are not clear. (The 5+5+5 structure often failed with both polymerases.)

Modifications in the Technique.

(*i*) A 6+6+6 primer (position 2668) was chosen for sequencing of the double-stranded M13 template, while the single-stranded M13 control was sequenced in parallel (both as described for Fig. 2*d*, except that the annealing was performed by transferring the microtubes from boiling water directly to ice water; neither alkaline denaturation nor ethanol precipitation was needed). The double-stranded template gave results similar to the single-stranded templates.

(*ii*) With fluorescent terminators in mind, we tested two methods of radioactive labeling. In parallel with the end-labeling of the front module with [γ-^{32}P]ATP as above, the radioactive label was incorporated with [α-^{32}P]dATP by Sequenase during the primer extension. The primer was of the 6+6 type in position 2891 (the same as in Fig. 1). For [α-^{32}P]dATP incorporation, we prevented possible extension of the back module by adding dideoxy ATP to the 3′ end of the corresponding pentamer, 3′-GGAGC-5′, with terminal transferase. (A similar modification of the 3′ end may be required with fluorescent terminators or internal labeling in automated sequencing, so that two libraries may be used—unmodified for the front module and modified for the other modules.) Both methods of labeling gave readable band patterns. Modular primers were extended even in a final dNTP concentration as low as 0.1 μM at the labeling step.

(*iii*) For most structures of modular primers, quick annealing, such as direct placing of the microtubes from boiling to ice water, generally resulted in much cleaner but somewhat weaker band patterns than a slower cooling.

Table 1. Modular primers tested by sequencing M13mp18

1	2	3	4	5	6	7	8
650	TACGGA	2	6+6+6	657	AG, AA	+	+
651	GCATTA	3	6+6+6	663	AA, AA	+	−
652	AGGAAA	7	6+6+6	669	AA, AA	+	−
653	ACCGCA	6	6+6+6	675	AA, TA	+	−
636	TAGACA	1	6+6+6	1037	AG, AA	+	+
637	GGAGAA	5	6+6+6	1043	AA, CA	+	+
612	AAGGGA	4	6+6+6	1068	AA, AA	+	+
635	CGTTCG	4	6+6+6	1415	GG, TG	+	+
672	CTCCCG	1	6+6+6	1736	GA, CA	+	+
640	GTGACA	2	6+6+6	2095	AA, TT	+	+
616	TTGGAG	3	6+6+6	2261	GG, TT	+	+
614	TGTCAG	1	6+6+6	2476	GA, GA	+	+
615	ACTGCG	2	6+6+6	2482	GA, CG	+	+
628	AAATGG	4	6+6+6	2668	GA, AG	+	+
612	AAGGGA	4	6+6+6	2674	AG, TT	+	+
629	TAGCCA	3	6+6+6	2686	AA, CA	+	+
625	ATAATA	5	6+6+6	2885	AA, AA	+	+
3196	ACGCAA	3	6+6+6	2891	AA, CC	+	+
618	CAAAGG	6	6+6+6	2894	GA, AA	+	+
612	AAGGGA	4	6+6+6	5593	AA, AG	+	+
656	GAAGAG	4	6+6+6	7130	GG, TT	−*	−
658	TTCATA	3	6+6+6	7142	AA, CC	−*	−
634	GAGGGA	3	6+6+6	1409	AC, GG	−	−
639	ACAGAC	2	6+6+6	1742	CA, AC	−	−
641	ATGAGT	4	6+6+6	2101	TT, GA	+	−
613	GGGAGT	2	6+6+6	2680	TT, AA	+	+
626	AATAAT	8	6+6+6	2884	TA, AA	+	+
612	AAGGGA	4	6+6+6	3104	AC, AT	−	−
657	GGCGTT	5	6+6+6	7136	TT, AA	−*	−
659	ATGTCC	3	6+6+6	7148	CC, TT	−	−
660	CAGTAT	1	6+6+6	7154	TT, AA	−	−
636	TAGACA	1	6+8	1037	AG	−	−
635	CGTTCG	4	6+8	1415	GG	+	+
672	CTCCCG	1	6+8	1736	GA	+	+
640	GTGACA	2	6+8	2095	AA	−	+
616	TTGGAG	3	6+8	2261	GG	+	+
614	TGTCAG	1	6+8	2476	GA	+	+
615	ACTGCG	2	6+8	2482	GA	+	+
628	AAATGG	4	6+8	2668	GA	+	+
612	AAGGGA	4	6+8	2674	AG	−	+
634	GAGGGA	3	6+8	1409	AC	−	−
639	ACAGAC	2	6+8	1742	CA	+	+
641	ATGAGT	4	6+8	2101	TT	−	+
527	TCGTG	3	5+5+7	213	GG, AA	−	−
525	TCACA	4	5+5+7	641	AA, GA	+	−
521	AGACA	10	5+5+7	1037	AG, AA	+	−
506	AAAGG	11	5+5+7	2894	GA, AA	−	−
508	AGCCA	12	5+5+7	2899	AA, AA	+	+
504	CCAAA	10	5+5+7	2901	AG, GA	+	−
519	GAGGG	9	5+5+7	1408	GA, TC	−	−
516	TGACA	12	5+5+7	2095	AA, GT	+	−
502	GTCAG	2	5+5+7	2476	GA, CG	−	−
501	AGGGA	10	5+5+7	2674	AG, GT	−	−
512	ACAGA	8	5+5+7	2772	AA, CA	+	−
3199	CGCAA	15	5+5+7	2891	AA, GC	+	+

Columns: 1, front module identification numbers; 2, front module 3′ → 5′ sequence; 3, number of occurrences of the complementary site in the single-stranded M13mp18 template; 4, structure type of the modular primer (see text); 5, position in single-stranded M13mp18 (GenBank numbering) of the labeled 5′ end of the front module; 6, base order in the stacking between the modules (from front to back modules); 7 and 8, success of the modular primer with Sequenase and *Bst* polymerase, respectively (determined by whether a readable band pattern is obtained, regardless of the band intensity).

*These primers did work when the template secondary structure was suppressed (see text).

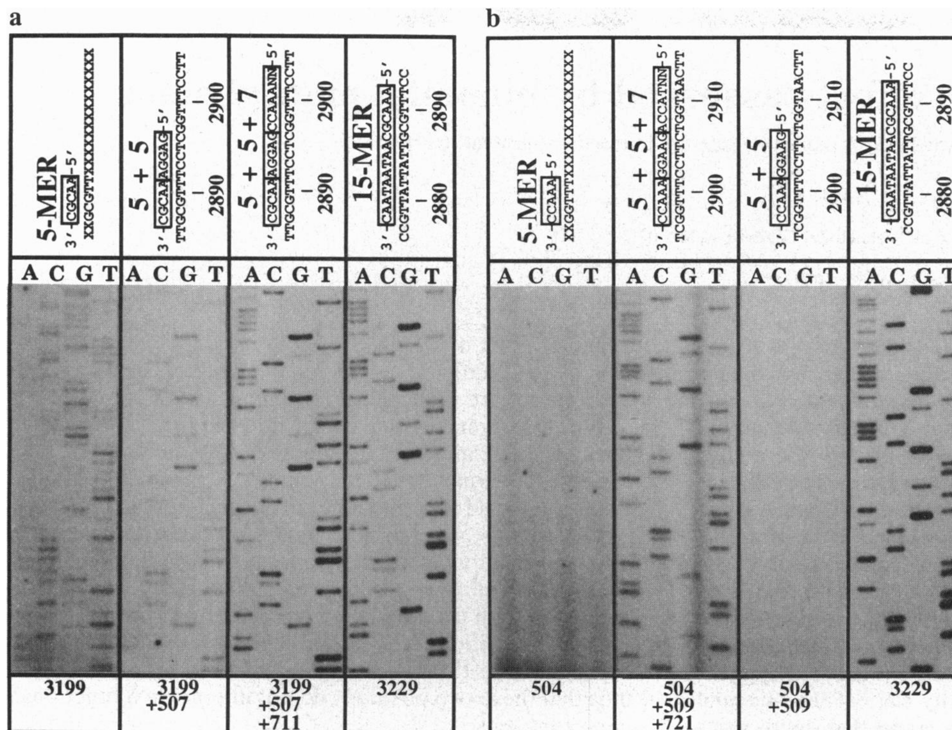

FIG. 4. Modular primers of the 5+5+7 structure. The back heptamers were degenerate in two positions denoted N. The annealing and amounts of middle pentamer and back degenerate heptamer were as in Fig. 2. The amount of the front module, when alone, was 2 pmol, a 4-fold excess over the template. The extension reaction temperature regime was as in Fig. 1*a*.

DISCUSSION

Cooperativity Mechanism. Central to the modular primer mechanism is its cooperativity, a manyfold increase in the sequence specificity and efficiency of priming by a module in the presence of an adjacent module. The following are two of the possible reasons for this cooperativity phenomenon.

(*i*) The affinity of an oligonucleotide for the DNA template increases due to the annealing of an adjacent oligonucleotide (5, 6). The base-stacking interaction between the oligonucleotides may play a role here, as may their cooperative suppression of the secondary structure of the template. The temperature of dissociation/association of an oligonucleotide and the template is increased by several degrees in the presence of an adjacent oligonucleotide (5, 6).

(*ii*) DNA polymerases may prefer more than one module annealed in tandem—e.g., longer primers are known to be more efficient substrates for polymerases (7). This selection occurs as the polymerase interacts with the primer, regardless of the annealing preferences of the primers themselves. DNA polymerases may differ in this respect. Indeed, in most experiments, as in Fig. 1*b*, *Bst* polymerase, unlike Sequenase, yielded a blank gel in reaction with a single hexamer primer, whereas the addition of the back modules resulted in a readable band pattern.

The principal difference between explanations *i* and *ii* is that the auxiliary modules boost the priming efficiency at the targeted site by stimulating annealing (in *i*) and/or extension (in *ii*) of the front module. Accordingly, the effect would take place at the stage of annealing (*i*) and/or polymerization (*ii*).

Closing the Cycle of Automated Walking Sequencing. The relatively small size of the libraries for the hexamer/pentamer modules makes the library approach feasible. This approach can alleviate the synthesis bottleneck in the walking primer sequencing and, perhaps more importantly, make it possible to close the automated walking cycle. In the cycle of walking sequencing, automation is already commercially available at the stages of (*i*) electrophoresis and reading the band pattern (Applied Biosystems; Pharmacia LKB; Millipore, Bedford, MA; and Li-Cor, Lincoln, NE), (*ii*) performing a sequencing reaction (Applied Biosystems and Beckman), and (*iii*) loading gels (Bio-Rad). The only stage that prevents full automation of the cycle is the synthesis of primer for each walking step. This stage can be made redundant by modular primer libraries. The closure of the automation cycle achieves an important goal: continuous advance along large templates in a rapid and inexpensive fashion. A closed cycle robotic system would store a module library. In the proposed system, modules would be selected for each walking cycle by computer to form a primer on the basis of the sequence read in the previous cycle. As ultrathin (8–10) and reusable capillary gels (11, 12) reduce the sequencing electrophoresis time to <1 hr and the detection limit to a few thousand fluorescent molecules per band (10^{-20} mol), the modular primers can be expected to decrease the time per walking cycle to 1 or 2 hr from the time of several days it now takes.

Note Added in Proof. A somewhat different technique of modular primers has been independently developed by Studier and colleagues (13), who use single-stranded DNA binding protein to suppress alternative priming.

We warmly thank Marina Goldberger for assistance.

1. Studier, F. W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6917–6921.
2. Szybalski, W. (1990) *Gene* **90,** 177–178.
3. Sambrook, J., Fritsch, E. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Plainview, NY).
4. Bloomfield, V. A., Crothers, D. M. & Tinoco, I. (1974) *Physical Chemistry of Nucleic Acids* (Harper & Row, New York).
5. Lin, S., Blake, K. R., Miller, P. S. & T'so, P. (1989) *Biochemistry* **28,** 1054–1061.
6. Abramova, T. V., Gorn, V. V., Kutyavin, I. V., Lebedev, A. V., Lochov, S. G. & Podyminogin, M. A. (1990) *Dokl. Biochem.* **315,** 316–319.
7. Nevinsky, G., Veniaminova, A., Levina, A., Podust, V., Lavrik, O. & Holler, E. (1990) *Biochemistry* **29,** 1200–1207.
8. Brumley, R. L. & Smith, L. M. (1991) *Nucleic Acids Res.* **19,** 4121–4126.
9. Stegemann, J., Schwager, C., Erfle, H., Hewitt, N., Voss, H., Zimmermann, J. & Ansorge, W. (1991) *Nucleic Acids Res.* **19,** 675–676.
10. Kostichka, A. J., Marchbanks, M. L., Brumley, R. L., Drossman, H. & Smith, L. M. (1992) *Bio/Technology* **10,** 78–81.
11. Cohen, A. S., Najarian, D. R. & Karger, B. L. (1990) *J. Chromatogr.* **516,** 49–60.
12. Smith, L. M. (1991) *Nature (London)* **349,** 812–813.
13. Kieleczawa, J., Dunn, J. J. & Studier, F. W. (1992) *Science* **258,** 1787–1791.