KRISTEL

FINAL REPORT DOE/ER/61831--T1

DOE Grant No. DE-FG02-94ER61831 with the Weizmann Institute of Science, ending 3/31/96.
Title: DNA Sequencing Technology, Walking With Modular Primers

Levy Ulanovsky, Principal Investigator

ABSTRACT

The success of the Human Genome Project depends on the development of adequate technology for rapid and inexpensive DNA sequencing, which will also benefit biomedical research in general. We are working on DNA technologies that eliminate primer synthesis, the main bottleneck in sequencing by primer walking. We have developed modular primers that are assembled from three 5-mer, 6-mer or 7-mer modules selected from a presynthesized library of as few as 1000 oligonucleotides (=4,=5,=7). The three modules anneal contiguously at the selected template site and prime there uniquely, even though each is not unique for the most part when used alone. This technique is expected to speed up primer walking 30 to 50 fold, and reduce the sequencing cost by a factor of 5 to 15. Time and expense will be saved on primer synthesis itself and even more so due to closed-loop automation of primer walking, made possible by the instant availability of primers . Apart from saving time and cost, closed-loop automation would also minimize the errors and complications associated with human intervention between the walks.

We have also developed two additional approaches to primer-library based sequencing. One involves a "branched" structure of modular primers which has a distinctly different mechanism of achieving priming specificity. The other introduces the concept of "Differential Extension with Nucleotide Subsets" as an approach increasing priming specificity, priming strength and allowing cycle sequencing. These approaches are expected to be more robust than our original version of the modular primer technique.

INTRODUCTION

The success of the Human Genome Project depends on the development of adequate technology for rapid and inexpensive DNA sequencing, which will also benefit biomedical research in general. The currently favored shotgun strategy for DNA sequencing has two main bottlenecks: template preparation (front end), and assembly of the sequence contigs (back end). Primer walking minimizes both of these problems and in addition reduces the redundancy of sequencing several fold. However, the primer walking strategy has its own bottleneck, primer synthesis, which is expensive, slow, and most critically, complicates full automation. It was proposed to eliminate the primer synthesis step by the use of presynthesized libraries of primers of different sequences (=1). A common rationale behind most library based priming techniques is that since the scale of primer synthesis exceeds the amount used in a conventional sequencing reaction by a factor of million or so, thousands of usable copies of the library can be aliquoted from a single synthesis. More importantly, the instant availability of primers makes complete

DISCLAIMER

automation of the closed cycle of primer walking possible. Closed-end automation would speed up primer walking by a factor of several dozens and decrease the cost of DNA sequencing by about an order of magnitude. However, the problem is that even the shortest primer expected to be unique in a plasmid-size template, a nonamer, has a library of unmanageable size (262,144 possible nonamer sequences). To reduce the library to a manageable size (e.g. 4,096 possible hexamers), a number of approaches have been developed in which individual short oligonucleotides (each too short to prime uniquely when alone) were either ligated (=2a,=2b,=2c,=2d,=2e), or assembled without ligation (=3,=4,=5,=6a,=6b,=6c,=6d,=7) to give unique (long) primers.

We are working on DNA sequencing technologies, such as modular primer walking, that eliminate primer synthesis, the main bottleneck in primer walking. Modular primers that are assembled from three 5-mer, 6-mer or 7-mer modules selected from a presynthesized library of as few as 1000 oligonucleotides (=4,=5,=7). In one version of the technique (contiguous modular primers), the three modules anneal contiguously at the selected template site and prime there uniquely, even though each is not unique for the most part when used alone. The primer library approaches are expected to speed up primer walking 30 to 50 fold, and reduce the sequencing cost by a factor of 5 to 15. Time and expense will be saved not only on primer synthesis itself but, more importantly, because the instant availability of the primers enables closed-loop automation of the complete cycle of walking sequencing. Apart from saving time and cost, closed- loop automation would largely minimize the errors and complications associated with human intervention between the walks.

AUTOMATED SEQUENCING WITH MODULAR PRIMERS

We have demonstrated that modular primers can be used with dye-terminators of the ABI 373A automated sequencer (=5). Also, reactions with modular primers and dye terminators were run successfully on the replaceable matrix capillaries of Barry Karger (Northeastern University) in 60 min, and read beyond 500 bases with as few as 2 base-calling errors. Because no protein needs to be removed, no precipitation or phenol extraction (an obstacle to closed-end automation) is required. The success rate and quality of automated sequencing with modular primers are similar to those with conventional primers 17-20 bases long. For the most part, few (if any) base-calling errors are found within the first 400 bases of the sequence run. In those experiments we preferred pentamer-based primers of the 5+7+7 structure with Pu-Pu base stacking between the 5-mer (to be extended) and the adjacent 7-mer. Both heptamers were 3'-end modified to prevent their extension by the polymerase; each having two degenerate positions and thus the same size library as the pentamers (512 sequences). Even though modular primers of this type used with dye-terminators for Sequenase showed a success rate similar to that of conventional long primers, these dye-terminators are avoided by most users as unreliable.

MECHANISM OF THE CONTIGUOUS MODULAR PRIMER EFFECT

Initially it was largely a mystery why modular primers prime uniquely, while a single module used alone under the same conditions, often shows alternative priming of comparable strength. In a puzzling way, the single module, even in a large excess over the template, no longer primes at the alternative sites, when modules with which it can form a contiguous string

are also present. We performed a special study (=7) which revealed that this phenomenon cannot be explained by cooperative annealing of the modules at the intended site in the template. Instead, the mechanism seems to involve competition between different primers for the available polymerase. In this competition, the polymerase is preferentially engaged by longer primers, whether modular or conventional, at the expense of shorter primers, even though the latter can otherwise prime with similar or occasionally higher efficiency.

BRANCHED PRIMERS

We have also developed a "branched" type of modular primer with a distinctly different mechanism of action, where the modules anneal both to the target and to each other, forming a branched structure of the 3-way junction type. While contiguous modular primers are made specific by the polymerase preference for longer primers (=7), branched primers, in contrast, owe their specificity to cooperative annealing of their modules to the intended site in the template. This cooperativity of annealing to the template is provided by mutually complementary "constant" segments of the two modules that bind to each other. In addition to the constant segment, each of the two modules also contains a "variable" segment whose sequence is not complementary to the other module, but rather is selected from the library to match the respective half of the intended site in the template.

DENS (DIFFERENTIAL EXTENSION WITH NUCLEOTIDE SUBSETS)

Furthermore, we have introduced a conceptually different technique for unique priming using short oligonucleotides that normally prime at multiple sites in the target (template) DNA. We have termed this conceptually new technique DENS (Differential Extension with Nucleotide Subsets), Nucleic Acids Research, in press. DENS works by selectively extending a short primer, making it a long one at the intended site only. The procedure starts with a limited initial extension of the primer (at 20-30°C) in the presence of only 2 out of the 4 possible dNTPs. The primer is extended by 6-9 bases or longer at the intended priming site, which is deliberately selected, (as is the two-dNTP set), to maximize the extension length. The subsequent termination reaction at 60-65°C (with all 4 dNTPs present), then accepts the extended primer at the intended site, but not at alternative sites, where the initial extension (if any) is generally much shorter. DENS allows the use of primers as long as 8-mers (degenerate in 2 positions) which prime much more strongly than modular primers involving 5-7 mers and which (unlike the latter) can be used with thermostable polymerases, thus allowing cycle-sequencing with dye-terminators compatible with Taq DNA polymerase, as well as making double-stranded DNA sequencing more robust.

SUCCESS RATE OF DENS AND ITS IMPROVEMENT

In our initial experiments DENS showed a 70% success rate with most failures being due to the signal being either too weak or undetectable. Interestingly, we have hardly seen a result that

was intermediate between high quality and failure. It has been shown that modular primer failures can be caused by unfavorable local secondary structure in the template (=4), and we are currently working on a computer program which can identify the sites to be avoided. The failures due to weak or undetectable signal are not the fault of DENS per se, and can be remedied by more sensitive detectors. The ABI sequencers we used show signal below a certain threshold as "blank". Radio-labeled sequencing shows most such apparent failures to be in fact successes. Most of the parameters and the procedures in the DENS technique are yet to be optimized, which is expected to further improve the success rate. Other possible future improvements include duplex-stabilizing base- modifications which have been found to enhance modular primer performance (=15). The addition of an inosine at the 5'-end of a single heptamer primer was also found to have a signal-enhancing effect. Possibly, it could result from increasing either the primer length (and thus acceptance by the polymerase) or the annealing stability through an extra base stacked to the 5'-end, not unlike ref. (=16), or both.

REFERENCES

1. F.W.Studier, Proc. Natl. Acad. Sci. USA 86, 6917 (1989).
2a. H.Blocker PCT/EP89/00579 (24.05.1989), filed by GBF, priority date 24.05.1988 (P 38 17 591.6).
2b. W.Szybalski, Gene 90, 177 (1990).
2c. H.Blocker . & Lincoln, D. N. 1994. The 'shortmer' approach to nucleic acid sequence analysis. I: Computer simulation of sequencing projects to find economical primer sets. *CABIOS* 10:193-197.
2d. T.Kaczorowski and W.Szybalski, Analytical Biochem . 221, 127 (1994).
2e. J.J. Dunn, L.-L.Butler-Loffredo and F.W.Studier, Anal. Biochem. 228, 91 (1995).
3. J.Kieleczawa, J.J.Dunn, and F.W.Studier, Science 258, 1787 (1992). This method uses contiguous "hexamer strings" in the presence of SSB which, under certain conditions, displaces individually annealed hexamers but not strings.
4. L.Kotler, D.Zevin-Sonkin, I.Sobolev, A.Beskin, and L.Ulanovsky, Proc. Natl. Acad. Sci. USA 90, 4241 (1993). This method uses contiguous oligonucleotide strings which prime uniquely without the difficult-to-remove SSB. Nor does DENS require SSB.
5. L.Kotler, I.Sobolev, and L.Ulanovsky, BioTechniques 17, 554 (1994).
6a. W.R.McCombie, and J.Kieleczawa, BioTechniques 17, 574 (1994).
6b. W.Hou and L.M.Smith, Nucleic Acids Res. 21, 3331 (1993).
6c. W.Hou and L.M.Smith, Analytical Biochem . 221, 136 (1994).
6d. M.A.Lodhi and W.R.McCombie, Genome Res. 6, 10 (1996).
7. A.Beskin, D.Zevin-Sonkin, I.Sobolev & L.Ulanovsky, Nucleic Acids Res. 23, 2881 (1995).
9. C.Yanisch-Perron, J.Vieira, and J.Messing, Gene 33, 103 (1985).
15. T.L.Azhikina et al., Proc. Natl. Acad. Sci. USA 90, 11460 (1993).
16. B.A.Schwitzer and E.T.Kool, J.Am.Chem.Soc. 117, 1863.(1995).