Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 11460–11462, December 1993
Biochemistry

Strings of contiguous modified pentanucleotides with increased DNA-binding affinity can be used for DNA sequencing by primer walking

(modified oligonucleotides/primers/genome walking/single-stranded DNA-binding protein)

TATYANA AZHIKINA, SVETLANA VESELOVSKAYA, VICTOR MYASNIKOV, VICTOR POTAPOV, OLGA ERMOLAYEVA, AND EUGENE SVERDLOV*

Laboratory of Human Genome Structure and Functions, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya Street 16/10, 117871 Moscow V-437, Russia

Communicated by Matthew Meselson, August 25, 1993

ABSTRACT **Modified oligonucleotides containing 5-methylcytidine and/or 2-aminoadenosine form tighter hybrids with DNA and are, therefore, more efficient primers for DNA sequencing as compared to their natural counterparts. Strings of contiguous modified pentanucleotides can be used for DNA sequencing by primer walking.**

One of the promising approaches to speeding up large-scale DNA sequencing is a recently described "primer walking" technique (1). In this technique, the first primer complementary to a known portion of DNA under study is used for sequencing the adjacent region. The new sequence information obtained is then used to prompt a new primer for further sequencing beyond the known regions. By repeatedly performing this procedure, the sequence of long DNA stretches can be established. However, the approach includes expensive and time-consuming synthesis of new and rather long primers after each sequencing step. This shortcoming was elegantly circumvented by Studier's group (2). Instead of synthesizing long sequencing primers, they proposed to use strings of short contiguous oligonucleotides constituting together the sequence of a desirable primer. Oligonucleotide constituents of such composite primers could be simply chosen from a premade oligonucleotide library. The authors (2) have demonstrated that strings of three contiguous hexamers can serve as specific sequencing primers in the presence of single-stranded DNA-binding protein (SSB). The full library of all possible hexanucleotides should contain as many as $4^6 = 4096$ individual components. Still this figure seems too high for routine sequencing, and a further decrease in size of the library is highly desirable.

In principle, the decrease could be achieved by shortening oligomers that form composite primers. However, it was convincingly demonstrated (2) that six is the lowest limit of the length of natural oligomers capable of functioning as components of composite primers, whereas pentamer combinations provide only "weak and ambiguous" priming.

We have previously described the use of modified oligonucleotides containing 5-methylcytidine (m^5C) and 2-aminoadenosine (n^2A) instead of their natural counterparts as primers for sequencing (3). Oligonucleotides with m^5C (4) or n^2A (see the discussion of the problem in ref. 5) or both modified residues (3) have higher affinity to their complementary sequences as compared with identical but unsubstituted oligonucleotides; therefore, the modified oligonucleotides are more efficient sequencing primers (3). Here we demonstrate that modified pentanucleotides can be successfully used as constituents of composite primers for the primer walking strategy of DNA sequencing.

MATERIALS AND METHODS

Phosphoramidites. 5-Methyl-2′-deoxycytidine and 2-amino-2′-deoxyadenosine were prepared according to refs. 6 and 7; protected using the 4,4′-dimethoxytrityl group for the 5′ position, the benzoyl group for the N^4 in 5-methyl-2′-deoxycytidine, and the dimethylacetamidine group for N^2 and N^6 in 2-amino-2′-deoxyadenosine (8); and converted to the corresponding 3′-cyanoethyldiisopropyl phosphoramidites (9). Other phosphoramidites were synthesized by a standard procedure.

Oligonucleotide Synthesis. Oligonucleotides were synthesized on a controlled porous glass support with a Milligene 7500 DNA synthesizer. After deprotection by aqueous ammonia at 60°C for 20 h, the oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis.

Melting Points of Duplexes. Melting point (t_m) values were determined for duplexes formed by modified oligonucleotides and a complementary unmodified 24-mer d(GCACTGGCCGTCGTTTTACAACGT) by measuring melting curves in TM buffer (10 mM Tris·HCl, pH 8.0/5 mM $MgCl_2$) on a Jasco 7850 spectrophotometer.

Sequencing. Single-stranded M13 DNA template was prepared as described (10).

Sequencing with pentamers was carried out by using Sequenase version 2.0 (United States Biochemical) according to a two-step Sequenase protocol with [α-^{33}P]dATP (11). An equilibrium mixture contained 0.5 μg of M13mp18 DNA, 3 μg of SSB protein, and 50 pmol of each pentanucleotide in a standard buffer solution. Adverse effects of SSB protein in running the gels were prevented by adding 0.2% SDS to the stop solution.

RESULTS AND DISCUSSION

We have previously shown that oligonucleotides with a modified C and/or A (m^5C and n^2A, respectively) are considerably more efficient DNA sequencing primers than unmodified ones (3). The efficiency was evaluated by the quantity of the primer incorporated in newly synthesized nested DNA fragments giving rise to sequencing patterns. Table 1 briefly summarizes the results obtained in ref. 3 and, in particular, presents relative efficiencies of priming with the standard universal primer and two of its modified analogs (containing either m^5C or both m^5C and n^2A, respectively). The data suggested to us that the lengths of oligonucleotide

Abbreviations: SSB, single-stranded DNA-binding protein; m^5C, 5-methylcytidine; n^2A, 2-aminoadenosine.

*To whom reprint requests should be addressed.

Table 1. Relative efficiencies of priming by modified oligonucleotides in sequencing experiments

Primer	t_m, °C	Relative efficiencies*
GTAAAACGACGGCCAGT	68	1, 0
GTAAAACGACGGCCAGT	73	3, 7
GTAAAACGACGGCCAGT	78	5, 2

Modified bases are shown in boldface type.

*Relative efficiencies were estimated as total densities of the bands of sequencing ladders determined by densitometry (3). Generally speaking, relative efficiency is a function of several parameters (concentration, primer-to-template ratio, conditions of annealing); the values in the table illustrate a typical range of differences. The annealing temperature and the primer-to-template ratio were 42°C and 2:1, respectively.

constituents of Studier's composite primers (1) could be reduced by using modified oligonucleotides.

To test this, we analyzed priming with mixtures of three modified or unmodified pentanucleotides contiguously covering 15 3′-terminal residues of the standard M13 sequencing primer as shown in Table 2. Sequencing experiments were carried out in the presence of *Escherichia coli* SSB under conditions described by Kieleczawa *et al.* (2). In accordance with their data, natural pentanucleotides could not prime the DNA synthesis either individually or as strings of three (data not shown). In contrast, mixtures of three analogous but modified pentamers used as primers resulted in clear-cut patterns of intense bands in sequencing gels similar to those obtained with the standard sequencing primer (Fig. 1).

It might be assumed that the higher stability of DNA–modified pentamer duplexes could increase the probability of priming by individual pentanucleotides. However, although M13 DNA includes many binding sites for the pentamers used (10 for CCAGT, 4 for GACGG, and 7 for AAAAC), we have not noticed any signs of individual priming. Consequently, the priming with contiguous modified pentanucleotides is specific enough for DNA sequencing.

These findings strongly suggest that combinations of contiguous modified pentanucleotides can be used as primers for DNA sequencing in the primer walking strategy. The quality of primer pentamers will obviously depend upon the content of modified bases, and therefore there is a question of how many pentamers are suitable for priming. Assuming that all of the possible modified pentanucleotides containing at least one C or A residue can be used, a complete priming library should contain $4^5 - 2^5 = 992$ pentamer elements. Such a library could be complemented if necessary with a few hexa- and/or heptanucleotides to meet the needs of sequencing experiments as proposed for the libraries of hexamers (2).

There are at least two reasons to believe that a universal priming library can be of reduced size and constructed solely of modified pentanucleotides:

(*i*) Primers for consecutive sequencing steps in the walking procedure must not necessarily correspond to the extreme 15–30 3′-terminal residues of each newly sequenced area of DNA. This region can be extended to about 100 residues, allowing one to find unique regions complementary to strings of three contiguous pentamers available in an existing library. This seems reasonable, even assuming that each sequencing step yields on the average as few as 500 residues. Kieleczawa *et al.* (2) considered this case and found that a library of 1500 hexanucleotides was sufficient to find at least one string of three available hexamers complementary to a DNA region of 100 residues with a probability of 0.99.

Table 2. Sequences of the universal M13 primer and pentanucleotides used as components of a composite primer

No.	Oligonucleotide structure
1	GTAAAACGACGGCCAGT
2	CCAGT
2a	CCAGT
3	GACGG
3a	GACGG
4	AAAAC
4a	AAAAC

Nucleotides with modified bases are shown in boldface type.

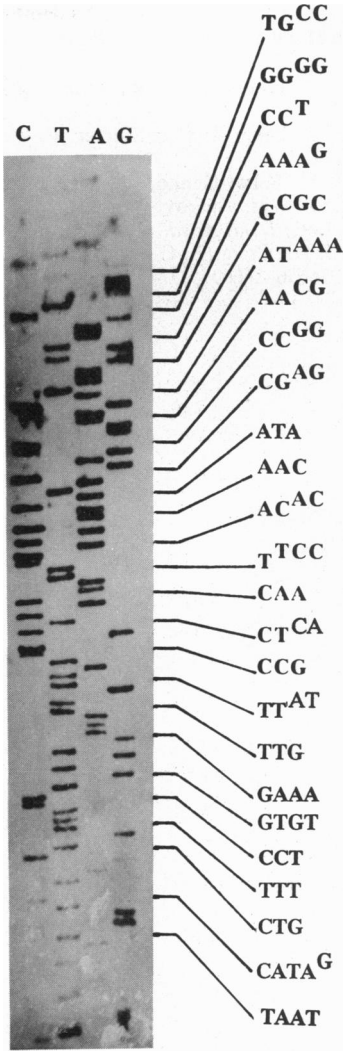

FIG. 1. M13mp18 DNA sequencing by a primer composed of three modified pentanucleotides (2a + 3a + 4a) in the presence of SSB.

(*ii*) The occurrence of different pentanucleotides in the human genome is not uniform and should be taken into account in the choice of the library components. This argument is supported by our analysis of 1.5×10^7 bp of human DNA sequences available in the European Molecular Biology Laboratory data base.

Taken together, these considerations allow one to hope that a library of about 400 modified pentamers would provide a probability close to 1 of finding a target for a string of three pentanucleotides within any DNA fragment of 100 nucleotides.

We thank Dr. A. Carrano and his colleagues from the Lawrence Livermore National Laboratory for the generous gift of enzymes and reagents without which this work could hardly have been done. We are also thankful to Dr. B. Glotov and S. Volik for discussions and

help in preparation of the manuscript. This work was financially supported by the Human Genome State Program of Russia.

1. Studier, F. W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6917–6921.
2. Kieleczawa, J., Dunn, J. J. & Studier, F. W. (1992) *Science* **258,** 1787–1791.
3. Azhikina, T. L., Schevchenko, Yu. O., Lebedev, Yu. B., Veselovskaya, S. V., Myasnikov, V. A., Potapov, V. K. & Sverdlov, E. D. (1993) *Dokl. Acad. Nauk. SSSR* **330,** 642–645.
4. Hoheinsel, J. D., Craig, A. G. & Lebrach, H. (1990) *J. Biol. Chem.* **265,** 16656–16660.
5. Cheong, C., Tinoco, I., Jr., & Chollett, A. (1988) *Nucleic Acids Res.* **16,** 5115–5122.
6. Sung, W. L. (1981) *Nucleic Acids Res.* **9,** 6139–6151.
7. Fathi, R., Goswami, B., Kung, P.-P., Gaffney, B. L. & Jones, R. A. (1990) *Tetrahedron Lett.* **31,** 319–322.
8. Froehler, B. C. & Matteucci, M. D. (1983) *Nucleic Acids Res.* **11,** 8031–8036.
9. Barone, A. D., Tang, J.-Y. & Caruthers, M. H. (1984) *Nucleic Acids Res.* **12,** 4051–4061.
10. Zinder, N. D. & Boeke, J. D. (1982) *Gene* **19,** 1–10.
11. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.