SHORT TECHNICAL REPORTS

Manuscripts published in the Short Technical Reports section are shorter and less comprehensive in scope than full Research Reports.

Improved Cycle Sequencing of GC-Rich Templates by a Combination of Nucleotide Analogs

BioTechniques 29:268-270 (August 2000)

ABSTRACT

A common problem in automated DNA sequencing when applying the Sanger chain termination method is ambiguous base calling caused by band compressions. Band compressions are caused by anomalies in the migration behavior of certain DNA fragments in the polyacrylamide gel because of intramolecular base pairing between guanine and cytosine residues. To reduce such undesired secondary structures, several modifications of the sequencing reaction parameters have been performed previously. Here, we have applied mixtures of the nucleotide analogs 7-deaza-dGTP and dITP instead of dGTP in the cycle sequencing reaction and in combination with varying buffer conditions. Band compressions were particularly well resolved, and reading length was optimal when a ratio of 7-deaza-dGTP:dITP of 4:1 was used in the in vitro DNA synthesis with AmpliTaqÒ FS DNA polymerase. We conclude that the incorporation of both nucleotide analogs at these particular ratios leads to heterogeneous DNA chains that result in a reduction or elimination of intramolecular base pairing and thus a higher accuracy in the base assignment.

INTRODUCTION

One common problem with the Sanger chain termination method in automated DNA sequencing is that certain DNA fragments show anomalies in their migration behavior during gel electrophoresis through the polyacrylamide gel. These anomalies in the gel migration are detected as compressed bands in the gel.

The phenomenon is caused by the ability of ssDNA to form base-paired hairpin structures in the DNA sequence (5,6). This is generally a result of guanine residues being able to undergo intramolecular interactions that are different from the conventional Watson-Crick base pairing with the proximate bases, so-called Hoogsteen base pairing. During electrophoresis, this reduces the viscous drag of the DNA fragment and leads to band compressions and uncertain base calling.

Previous experiments addressing the resolution of band compressions aimed to reduce or eliminate intramolecular base pairing through the full replacement of dGTP with such nucleotide analogs as dITP (2,3) or 7-deaza-dGTP (1,6). In genome projects in house, we came across numerous band compressions that we were not able to resolve using either 7-deaza-dGTP or dITP alone in the sequencing reaction instead of dGTP. To replace dGTP more efficiently with respect to the resolution of band compressions, we attempted to develop alternative nucleotide mixtures. To this end, we have investigated the use of nucleotide mixtures of 7-deaza-dGTP and dITP at certain ratios in cycle sequencing reactions together with AmpliTaq® FS DNA polymerase

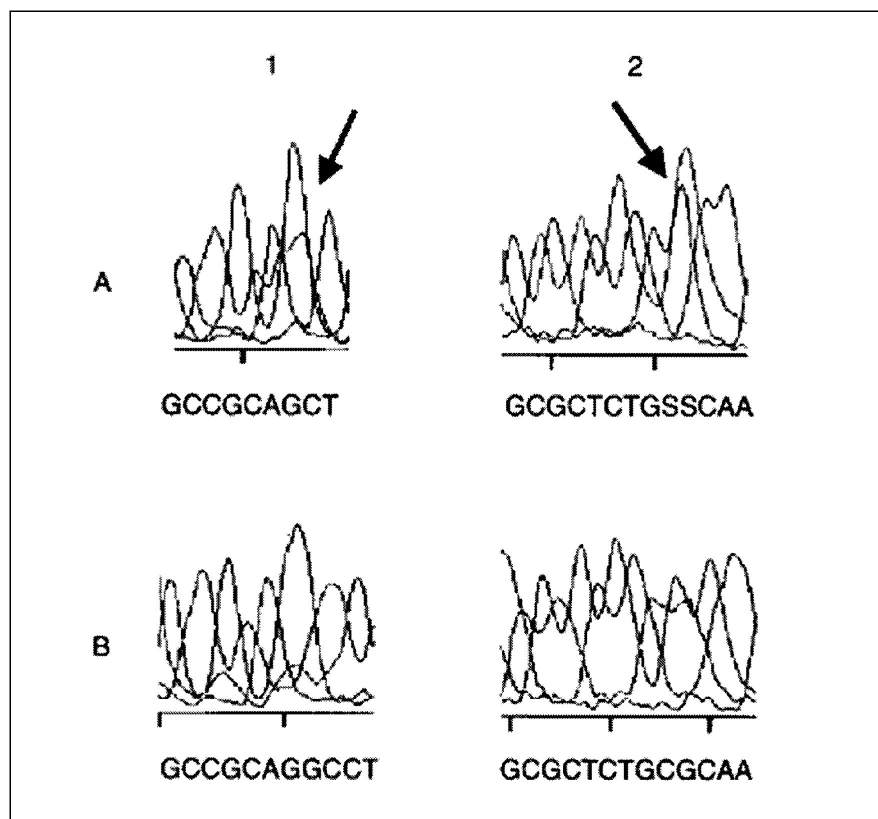

Figure 1. Application of a 7-deaza-dGTP:dITP mixture at a ratio of 4 to cycle sequencing reactions improves base calling. DNA sequence electropherograms produced by sequencing two different templates are compared to two different nucleotide mixtures. In the first case (A), the plasmid DNA was sequenced with 100% 7-deaza-dGTP replacing dGTP. This resulted in typical band compressions in contrast to case (B), in which the same templates were sequenced with a nucleotide-mixture of 7-deaza-dGTP and dITP at a ratio of 4, instead of dGTP, and resulted in a readable sequence. In (B), the DNA sequences, AGGCCT (template 1) and TGCGCA (template 2) were obtained correctly, while in panel (A), the electropherograms show two overlapping peaks (marked with arrows) that are caused by the compressed bands in the gel and lead to the errors in the nucleotide sequence.

Table 1. DNA Sequence of Compression-Prone Regions and Their Resolution

DNA Sequence	100% deaza-dGT	Solution with 4:1 mixture of 7-deaza-dGTP and dITP
1	GCCGCAGCT	GCCGCAG**GC**CT
2	GCGCTCTGSSC	GCGCTCTG**CG**C
3	GCGAAGGCCT	GCGAAG**C**GCCT
4	GAGCTGCTC	GAGCTGC**C**TC
5	GCTCTGCGC	GCTCTGC**G**GC
6	CCGGAGGT	CCGGA**C**GGT
7	GCGGACTAC	GCGCAC**G**TAC
8	GCCCGCAGCCT	not resolved right sequence: AGGCCT
9	TTGACCCCGCT	not resolved right sequence: ACCCCCGC

Seven of nine band compressions could be resolved using the nucleotide mixture 7-deaza-dGTP and dITP at a ratio of 4.

(PE Biosystems, Foster City, CA, USA) (4) and a reaction buffer containing ammonium sulfate.

MATERIALS AND METHODS

The sequencing reactions were carried out using 1 µg plasmid DNA with a 2-kb insert in a pTZ18R-vector (Amersham Pharmacia Biotech, Piscataway, NJ, USA), 2 pmol each upstream and downstream primer (doublex sequencing), 10 U AmpliTaq FS DNA polymerase, 2 µL sequencing buffer (containing 125 mM Tris-HCl, pH 9.3, 150 mM ammonium sulfate and 50 mM $MgCl_2$) in a total volume of 20 µL. The sequences of the primers, synthesized by MWG (Ebersberg, Germany), are (5′→3′): pTZW 5′-CATGCCT-GCAGGTCGACTC-3′; pTZC 5′-AC-GACTCACTATAGGGAATT-3′. Five microliters of this reaction mixture were individually combined with 2 µL of the nucleotide mixtures containing 1 mM dATP, dCTP, dTTP and combinations of 7-deaza-dGTP and dITP (all nucleotides are from Roche Molecular Biochemicals, Mannheim, Germany) in ratios varying from 1–0.1 in steps of 10% at a constant total dITP/7-deaza

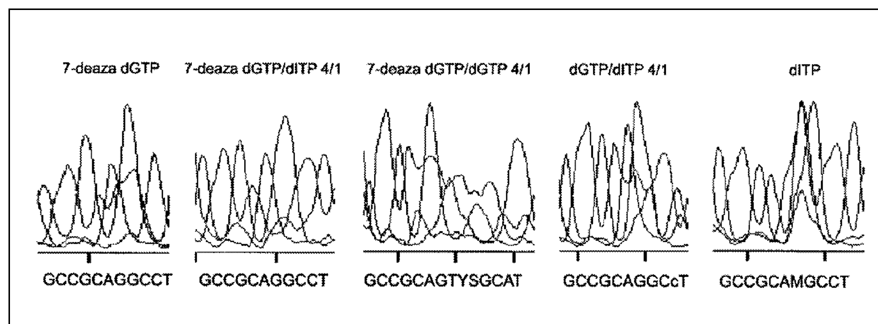

Figure 2. Comparison of sequencing results when applying mixtures of different nucleotide analogs. The replacement of dGTP in the sequencing reaction with various nucleotide analogs and their mixtures (only in the case of 7-deaza-dGTP:dITP at a ratio of 4) gave a correct readable DNA sequence electropherogram. When dGTP is completely replaced by dITP, this led to a stop, which is caused by a compression of all four nucleotide lanes in the gel. When it was partly replaced by 7-deaza-dGTP, the result was totally erroneous sequence data—that is, only the mixture of dGTP with dITP showed the right DNA sequence, but with uncertain base calling in the GCCT motif.

GTP concentration of 2 mM. Each mixture was supplemented with one of the four ddNTPs at a concentration of 2.5 µM. When using 100% 7-deaza-dGTP in the nucleotide mixture instead of dGTP, the buffer conditions were changed to the commercially available 1× sequencing buffer (PE Biosystems). All reactions were run in a Primus 96 Cycler (MWG) under identical cycling conditions applying 30 cycles of 15 s at 97°C (denaturing), 30 s at 55°C (annealing) and 120 s at 68°C (synthesis). The sequencing reaction products were separated on a 66-cm electrophoresis gel containing 4% Page Plus (Ameresco, Solon, OH, USA), 7 M urea and 1× TBE (Merck, Darmstadt, Germany) on a LICOR DNA-Sequencer Long ReadIR[2] 4200™ (LI-COR, Lincoln, NE, USA) under the following electrophoresis conditions: 2500 V, 150 W, 50 mA, 45°C for approximately 10 h. The DNA sequence was then processed with the A.L.F. base calling software, LANE TRACKER (EMBL, Heidelberg, Germany).

RESULTS AND DISCUSSION

The effect of using a combination of 7-deaza-dGTP and dITP, instead of dGTP, was evaluated for the Dye-Primer cycle sequencing method. We have chosen from templates stemming from ongoing genome projects for the analysis. The base contents of all four templates show a high percentage of guanine and cytosine residues. In the case of template (A), the GC content is 59%, template (B) 58.3% GC, template (C) 57.5% GC and template (D) 49.7% GC. In comparison, *E. coli* has a GC content of 50.8%. While sequencing the templates with the reagents containing 100% 7-deaza-dGTP in the nucleotide mixture, we found nine compressed regions (Table 1). We have accordingly modified sequencing reactions of these templates by varying the nucleotide constitution. We totally replaced dGTP with a mixture of 7-deaza-dGTP and dITP, using varying ratios of both nucleotide analogs from 1–0.1 in steps of 10%. Our observation was that, for seven of nine observed compressed regions, the sequence could be processed correctly with the nucleotide mixture containing a 7-deaza-dGTP:dITP quotient of 4 (Figure 1).

A typical sequencing reaction resolving compressions caused by the use of the 4:1 ratio of 7-deaza-dGTP:dITP was a read of 1156 bases with eight ambiguities. However, no common sequence motif in the compressed regions could be found to lead to a more general rule for compression-prone DNA sequences (7). We also applied nucleotide mixtures containing a 7-deaza-dGTP:dITP at various ratios, whereas the ratio of 0.2 also led to a resolution of five of these compressed regions. The nucleotide mixture containing both analogs at equal rates showed a resolution of only four of these compressed regions; the other ratios applied to the reaction did not lead to a further resolution of the band compressions compared to the 7-deaza-dGTP:dITP ratio of 4. When one of the compressed regions was sequenced with mixtures of different nucleotides and mixtures of nucleotide analogs (Figure 2), only the mixture 7-deaza-dGTP:dITP at the ratio of 4:1 led to the correct DNA sequence. The replacement of dGTP did not result in any deterioration of the quality of sequence patterns with respect to readable length and band uniformity.

These findings demonstrate that the use of mixtures of 7-deaza-dGTP and dITP at a defined ratio leads to improved sequencing results over conventional methods by taking advantage of the destabilizing effects on intramolecular base pairing of both nucleotide analogs, 7-deaza dGTP (1,6) and dITP (5). We therefore speculate that the incorporation of both nucleotide analogs of the above-given ratios leads to heterogeneous DNA chains that cause a reduction or elimination of intramolecular base pairing and thus a higher accuracy in base assignment. These improvements result in a significant decrease in ambiguous base assignments in base identification methods.

REFERENCES

1. **Barr, P.J., R.M. Thayer, P. Laybourn, R.C. Najarian, F. Seela and D.R. Tolan.** 1986. 7-deaza-2′-deoxyguanosine-5′-triphosphate: enhanced resolution in M13 dideoxy sequencing. BioTechniques *4*:428-432.
2. **Fuller, C.W.** 1992. Modified T7 DNA polymerase for DNA sequencing. Methods Enzymol. *216*:329-354.
3. **Hirao, I., Y. Nishimura, Y.-I. Tagawa, K. Watanabe and K.-I. Miura.** 1992. Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effect on DNA sequencing. Nucleic Acids Res. *20*:3891-3896.
4. **Innis, M.A., K.B. Myambo, D.H. Gelfand and M.A. Brow.** 1988. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. Proc. Natl. Acad. Sci. USA *85*:9436-9440.
5. **Mills, D.R. and F.R. Kramer.** 1979. Structure-independent nucleotide sequence analysis. Proc. Natl. Acad. Sci. USA *76*:2232-2235.
6. **Mizusawa, S., S. Nishimura and F. Seela.** 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res. *14*:1319-1324.
7. **Weinshenker, B.G., D.D. Hebrink, A.M. Gacy, C.T. McMurray.** 1998. DNA compression caused by an upstream point mutation. BioTechniques *25*:68-72.

The excellent technical assistance by Sylvia Ellwanger and Uli Bauer is gratefully acknowledged. This work was initiated by Dr. Hartmut Voss, LION Bioscience AG, Heidelberg, Germany, and financially supported by Roche Diagnostics, Mannheim, Germany, under the direction of Dr. Gregor Sagner. We also thank Dr. Ingo Kober, LION Bioscience AG, Heidelberg, Germany, and especially Lyle Middendorf, LI-COR Biotechnology Division, for their helpful comments.
Address correspondence to Michael Motz, Im Neuenheimer Feld 515-517, LION Bioscience AG, D-69120 Heidelberg. e-mail: michael.motz@lionbioscience.com

Received 6 December 1999; accepted 27 March 2000.

Michael Motz[1,2], Svante Pääbo[1] and Christian Kilger[2]
[1]MPI for Evolutionary Anthropology
Leipzig
[2]LION Bioscience AG
Heidelberg, Germany