Nucleic Acids Research, 1993, Vol. 21, No. 11 2709–2714

Elimination of band compression in sequencing gels by the use of N^4-methyl-2′-deoxycytidine 5′-triphosphate

Samantha Li+, Alberto Haces, Linda Stupar, Gulilat Gebeyehu and Reynaldo C.Pless*
Life Technologies, Inc., PO Box 6009, 8717 Grovemont Circle, Gaithersburg, MD 20884-9980, USA

Received December 24, 1992; Revised and Accepted April 20, 1993

ABSTRACT

Taq DNA polymerase, Sequenase, and the large fragment of *E.coli* polymerase I effectively utilize N^4-methyl-2′-deoxycytidine 5′-triphosphate (N^4-methyl-dCTP) in the place of dCTP in dideoxynucleotide terminator sequencing reactions on single-stranded templates. When the resulting fragment mixtures are resolved on sequencing gels, they are found to be free of band compressions even in cases where such compressions remain unresolved by the substitution of 7-deaza-dGTP for dGTP. Sequencing reactions using N^4-methyl-dCTP instead of dCTP are somewhat more prone to false stops than are sequencing reactions using 7-deaza-dGTP instead of dGTP; this difference is more pronounced when sequencing with Sequenase at 37°C than when sequencing with Taq DNA polymerase at 72°C. For the three polymerases investigated, replacement of dCTP by N^4-methyl-dCTP does not fundamentally change the characteristic variations in band intensities seen in the C-lane. N^4-methyl-dCTP can also be used for sequencing double-stranded DNA and for DNA amplification by the polymerase chain reaction.

INTRODUCTION

Resolution of sequencing reaction mixtures by electrophoresis through denaturing polyacrylamide gels is often locally impaired by irregularities in the spacing of the electrophoretic bands. This phenomenon usually occurs as 'band compression', a reduction in the spacing between consecutive bands, which may render their correct reading impossible. Band compression in sequencing gels occurs when the corresponding local nucleotide sequence contains an inverted repeat, and it is thought to arise from formation of hairpin structures in the single-stranded DNA fragments, despite the presence of 7 M urea in the polyacrylamide gel. To alleviate this problem, the dGTP in the synthesis mixtures can be substituted by 7-deaza-dGTP (1,2) or by dITP. However, the use of 7-deaza-dGTP does not lead to complete resolution of band compressions caused by extended inverted repeats (which give rise to particularly stable hairpin loops), and use of dITP often leads to false stops at G-sites.

We have examined substitution of N^4-methyl-2′-deoxycytidine 5′-triphosphate (N^4-methyl-dCTP) for dCTP as another method to destabilize the G·C base pairs, which are the mainstay in the hairpin structures, in order to suppress band compression in sequencing gels. The reduced stability of the G·N^4-methylC pair, compared to the G·C pair, has been documented in short self-complementary oligonucleotides (3,4). When those data are compared with the results obtained in a comparison of the melting temperatures of poly(G)·poly(C) and poly(7-deazaG)·poly(C) (5), the destabilization imparted to the G·C base pair by N^4-methylation of the cytosine appears to be much greater than the destabilization of the G·C pair caused by substituting 7-deazaguanine for guanine. It was expected, therefore, that substitution of N^4-methyl-dCTP for dCTP would be more effective in resolving band compressions than is substitution of 7-deaza-dGTP for dGTP. This is, indeed, borne out by the present study.

MATERIALS AND METHODS

The triethylammonium salt of N^4-methyl-dCTP was prepared in analogy to the published synthesis of N^4-aminoethyl-CTP (6). 100 mg of dCTP (sodium salt, Sigma Chemical Company) was dissolved in 4.38 ml of an aqueous solution containing 5.29 M methylammonium chloride, 0.01 M methylamine, 2.4 M sodium bisulfite, and 0.04% (w/v) hydroquinone. The solution, which had a pH of 6.5 at 25°C, was held for 18 h at 40°C under a blanket of argon. The mixture was adjusted to pH 8.5 with aqueous NaOH and held at 25°C for 2 h. After 100-fold dilution with water, the mixture was loaded onto a DEAE G-25 (bicarbonate form) column, which was then washed with water and eluted with a linear gradient of triethylammonium bicarbonate (0.01 M to 1 M). The appropriate fractions were pooled and desalted by repeated rotary evaporation with ethanol to afford a 35% yield of the desired material. The UV absorption spectra in aqueous solution at pH 10 (λ_{max} = 270 nm, λ_{min} = 247 nm) and at pH 2 (λ_{max} = 279 nm, λ_{min} = 241 nm) were close to those reported for the nucleoside (7). For quantitation of N^4-methyl-dCTP in aqueous solution, the ϵ_{max} value (11,700 M^{-1} cm^{-1} at 270 nm and pH 10) reported for the nucleoside (7) was used.

* To whom correspondence should be addressed

+ Present address: Hoefer Scientific Instruments, San Francisco, CA 94107-0387, USA

Enzymes were obtained from the following sources: Taq DNA polymerase for sequencing reactions and the large fragment of *E.coli* DNA polymerase I (Klenow fragment) from BRL, Sequenase from U.S. Biochemicals, Taq DNA polymerase for PCR (AmpliTaq) from Cetus.

[α-^{32}P]dATP (3000 Ci/mmol) and [^{35}S]dATPαS (650 Ci/mmol) were from Amersham.

Sequencing gel electrophoresis was performed in 0.4 mm thick 6% or 8% polyacrylamide gels, containing running buffer and 7 M urea, prepared from Gel-Mix 6 or Gel-Mix 8 (BRL). Running buffer was 100 mM Tris, 90 mM boric acid, 1 mM Na_2EDTA, prepared from Gel-Mix Running Mate (BRL). Electrophoresis was conducted at 45°C–50°C, measured on the outer glass surface. Autoradioagraphy was performed with Kodak SB film.

M13mp19(+) strand DNA, pUC19 DNA, the M13 17-base primer, and the M13/pUC Forward 23-Base Sequencing Primer were the products of BRL. 864I DNA and 964I DNA were the gift of Dr Deborah Polayes. HPV16/pT713 contains the entire HPV16 genome (7904 bp, reference 8) cloned into plasmid pT713 (2818 bp, BRL).

Sequencing reactions with Taq DNA polymerase were performed for 12 minutes at 72°C. The mixtures contained 60 μg/ml single-stranded template (i.e. 25 nM template), 0.10 μg/ml 23-base primer (i.e. 13 nM primer), 0.1 μM [α-^{32}P]dATP, 25 mM Tris–HCl (pH 9.0 at 25°C), 5 mM $MgCl_2$, 20 mM KCl, 2 mM dithiothreitol, 0.05% Triton X-100, and 33 U/ml Taq DNA polymerase. In addition, the various reaction tubes contained the nucleotide concentrations listed in Table I.

Sequencing reactions with Sequenase on single-stranded templates were performed according to the Sequenase protocol provided by the United States Biochemical Corporation. The termination reaction mixtures contained 8.4 μg/ml single-stranded template (i.e. 3.5 nM template), 0.58 μg/ml 23-base primer (i.e. 76 nM primer), 0.27 μM [^{35}S]dATPαS, 14 mM Tris–HCl (pH 7.5), 3.5 mM $MgCl_2$, 17.5 mM NaCl, and 113 U/ml Sequenase, and, in addition, the appropriate nucleotides in the following concentrations: for dATP, dCTP, dGTP, and dTTP: 40 μM; for N^4-methyl-dCTP and 7-deaza-dGTP: 80 μM; for ddATP, ddCTP, ddGTP, and ddTTP: 4 μM, except for ddCTP in reactions containing N^4-methyl-dCTP: 0.4 μM and for ddGTP in reactions containing 7-deaza-dGTP: 2 μM.

For sequencing with Sequenase on a double-stranded template, pUC19 DNA was denatured by alkaline treatment (9). Termination reaction mixtures had the same composition as used in sequencing of single-stranded DNA with Sequenase.

Sequencing with the large fragment of *E.coli* DNA polymerase I was performed essentially according to the KiloBase sequencing protocol of BRL. N^4-Methyl-dCTP was used instead of dCTP in the labeling-extension step preparatory to the termination reactions usign N^4-methyl-dCTP. The termination reaction mixtures contained 39 μg/ml M13mp19(+) DNA (i.e. 15.5 nM template), 0.11 μg/ml 17-base primer (i.e. 19.5 nM primer), 22 mM Tris–HCl (pH 8.0), 5 mM $MgCl_2$, 25 mM NaCl, 3.65 mM dithiothreitol, 0.56 μM [^{35}S]dATPαS, 55 U/ml large fragment of *E.coli* DNA polymerase I, and, in addition, the nucleotide concentrations listed in Table II.

PCR amplification was carried out in 50-μl solutions containing 200 μM dATP, 200 μM dCTP or N^4-methyl-dCTP, 200 μM dGTP, 200 μM dTTP, two primers (5′-GGTCGATGTATGTCTTGTTG-3′ and 5′-GTCTACGTGTGTGCTTTGTAC-3′, each at 1 μM concentration), HPV16/pT713 at 0.1 ng/ml (i.e. 0.014 pM plasmid), 50 mM KCl, 10 mM Tris–HCl (pH 8.3), 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin, and 6 units/ml AmpliTaq enzyme. The solutions were overlaid with mineral oil and subjected to 25 cycles of either a high-temperature regimen (1 min at 94°C, 2 min at 55°C, 3 min at 72°C) or a low-temperature regimen (1 min at 94°C, 2 min at 37°C, 3 min at 55°C) in a DNA Thermal Cycler from Perkin Elmer Cetus. The resulting mixtures were resolved by electrophoresis through a 2% agarose gel, followed by visualization with ethidium bromide and UV irradiation.

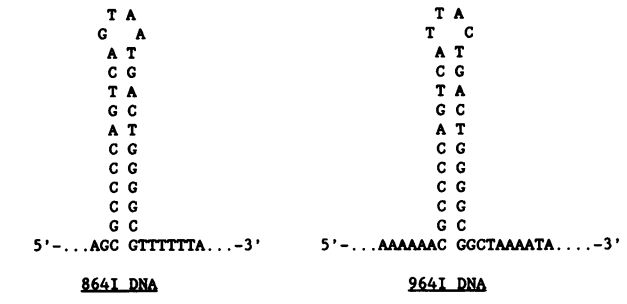

Figure 1. Local nucleotide sequences containing inverted repeats. In each case, the newly synthesized strand is shown.

RESULTS

As examples of templates leading to strong band compressions in the sequencing gels we chose 864I DNA and 964I DNA. In these DNAs, a 133-bp insert containing an inverted repeat is located in the multiple cloning sites of M13mp8/pUC8 and M13mp9/pUC9, respectively; the two constructs contain the identical insert in opposite orientations. In each case, the single-stranded template (and the corresponding newly synthesized fragments formed in the dideoxy terminator sequencing reactions) can form a double-stranded stem of eleven successive correctly matched base pairs (eight of which are G·C pairs), culminating in a four-nucleotide hairpin loop. The local sequences containing the inverted repeats are shown in Fig. 1 for the newly synthesized DNA (i.e. these sequences are complementary to the single-stranded templates).

Dideoxynucleotide terminator reaction mixtures with Taq DNA polymerase at 72°C contained the dNTP concentrations and ddNTP concentrations listed in Table I. The ddNTP/dNTP ratios were adjusted to give long extension (to >500 nucleotides from the labelled 5′-end), with approximately equal length in all four lanes. Fig. 2 shows the sequencing ladders for the region containing the inverted repeat in 964I DNA. For the set of reactions using N^4-methyl-dCTP instead of dCTP, the sequence is readable throughout, and all band spacings are regular. In contrast, the polymerization mixtures obtained with the canonical set of dNTPs (i.e. dATP, dCTP, dGTP, and dTTP) show severe band compression in the sequence GGGGCGGC (i.e. in the putative stem sequence distal to the primer), followed by distinct distention of spacings in the subsequent region (i.e. in the sequence immediately past the hairpin). Qualitatively similar results were obtained in the set of reactions using 7-deaza-dGTP instead of dGTP; band compression is still severe in the sequence GGGCGGC (though some resolution is achieved by the use of the deaza analogue), and band distention is still seen in the sequence TAA following the hairpin.

Table I. Nucleotide concentrations (μM) in the sequencing reactions with Taq DNA polymerase

Canonical set								
	dATP	dCTP	dGTP	dTTP	ddATP	ddCTP	ddGTP	ddTTP
A-lane	16	24	24	24	160	–	–	–
C-lane	24	16	24	24	–	176	–	–
G-lane	24	24	38	24	–	–	90	–
T-lane	24	24	24	8.8	–	–	–	150
N^4-methyl-dCTP								
	dATP	N^4-me-dCTP	dGTP	dTTP	ddATP	ddCTP	ddGTP	ddTTP
A-lane	16	48	24	24	160	–	–	–
C-lane	24	32	24	24	–	58	–	–
G-lane	24	48	38	24	–	–	90	–
T-lane	24	48	24	8.8	–	–	–	150
7-deaza-dGTP								
	dATP	dCTP	7-deaza-dGTP	dTTP	ddATP	ddCTP	ddGTP	ddTTP
A–lane	16	24	48	24	160	–	–	–
C-lane	24	16	48	24	–	176	–	–
G-lane	24	24	38	24	–	–	45	–
T-lane	24	24	48	8.8	–	–	–	150
N^4-methyl-dCTP + 7-deaza-dGTP								
	dATP	N^4-me dCTP	7-deaza-dGTP	dTTP	ddATP	ddCTP	ddGTP	ddTTP
A-lane	16	48	76	24	176	–	–	–
C-lane	24	32	76	24	–	58	–	–
G-lane	24	48	76	24	–	–	45	–
T-lane	24	48	76	8.8	–	–	–	150
dITP								
	dATP	dCTP	dITP	dTTP	ddATP	ddCTP	ddGTP	ddTTP
A-lane	16	24	48	24	160	–	–	–
C-lane	24	16	48	24	–	176	–	–
G-lane	24	24	76	24	–	–	45	–
T-lane	24	24	48	8.8	–	–	–	150

In Fig. 2, in the set of lanes corresponding to the experiment involving N^4-methyl-dCTP, a series of false stops is seen at the beginning of the proximal complementary sequence of the hairpin (sequence CGCCCCAG); these false stops, however, are too weak to interfere with the reading of the sequence. False stops in this region are also seen for the reaction mixtures obtained with the canonical set of dNTPs or with the use of 7-deaza-dGTP, but they are less pronounced than the false stops seen in the mixtures obtained with N^4-methyl-dCTP. These stops probably arise from the hesitation of the DNA polymerase as it meets the hairpin structure in the template.

Fig. 3 shows the sequencing autoradiogram for the hairpin region in 864I DNA. Again, severe band compression (sequence GGGGCG) and band distention (sequence TTT) are seen for the reaction mixtures obtained with the use of 7-deaza-dGTP, while the use of N^4-methyl-dCTP results in completely regular band spacings throughout the region. Use of both analogues, N^4-methyl-dCTP and 7-deaza-dGTP, in the same set of mixtures results in regular spacing, as expected, but also in an increased propensity for false stops and highly variable band intensities within the lanes. Replacement of dGTP by dITP produces regular spacing throughout, but leads to a much increased background and pronounced false stops.

For a more quantitative evaluation, Fig. 4 presents the peak-to-peak distances of successive electrophoretic bands for sequencing reaction mixtures obtained on the 964I DNA template. For the reaction mixtures obtained with N^4-methyl-dCTP,

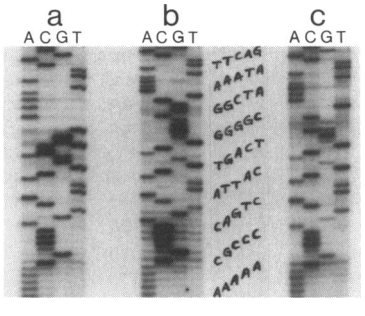

Figure 2. 6% sequencing gel obtained with Taq DNA polymerase on a single-stranded template of 964I DNA. Set **a** contains dATP, dCTP, dGTP, and dTTP; set **b** dATP, N^4-methyl-dCTP, dGTP, and dTTP; set **c** dATP, dCTP, 7-deaza-dGTP, and dTTP.

regular spacing is observed throughout the potential hairpin region. Reaction mixtures obtained with the canonical set of dNTPs show regular spacing for fragments extending through the proximal segment of the potential stem region and the loop region, but show drastically reduced spacing as the fragments extend into the distal stem sequence and hence can form hairpin structures; band interpretation here is so unclear that only an average mobility decrease (barely larger than zero) can be given for a sequence of ten nucleotides. In the region immediately following the hairpin, band spacing is dramatically increased for

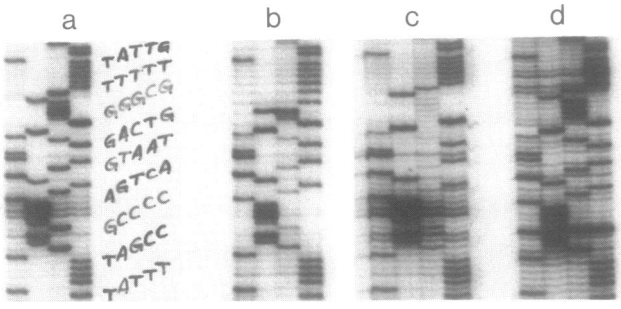

Figure 3. 6% sequencing gel obtained with Taq DNA polymerase on a single-stranded template of 864I DNA. Set **a** contains dATP, N^4-methyl-dCTP, dGTP, and dTTP; set **b** dATP, dCTP, 7-deaza-dGTP, and dTTP; set **c** dATP, N^4-methyl-dCTP, 7-deaza-dGTP, and dTTP; set **d** dATP, dCTP, dITP, and dTTP.

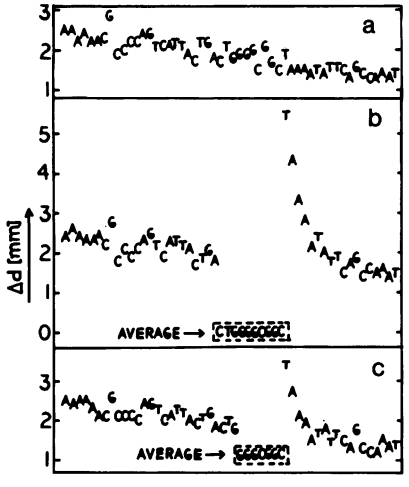

Figure 4. Spacing of the electrophoretic bands measured for the region containing the inverted repeat in 964I DNA. For each band, the separation (Δd) from the next lower band in the ladder is given. The bands are identified by the base for which they stand. The 5′→3′ direction is from left to right. In panel **a**, the sequencing laders were obtained with dATP, N^4-methyl-dCTP, dGTP, and dTTP; in panel **b**, with dATP, dCTP, dGTP, and dTTP; in panel **c**, with dATP, dCTP, 7-deaza-dGTP, and dTTP.

several bands, reaching in one instance almost thrice the normal value. The set of reaction mixtures obtained with the 7-deazaguanine nucleotide again shows a low average band spacing in the unresolved region (GGGCGGC) and distinct band distention after it, but these effects are not as pronounced as in the reactions obtained with the canonical set of nucleotides.

Other DNA polymerases used in dideoxynucleotide terminator sequencing were also able to accept the N^4-methylcytosine analogue. Fig. 5 shows results obtained with Sequenase on M13mp19 single-stranded DNA at 37°C, using [^{35}S]dATP. Clear sequence is obtained with the canonical set of dNTPs, with the substitution of 7-deaza-dGTP for dGTP, and with the substitution of N^4-methyl-dCTP for dCTP; but in the latter case there is a greater propensity for false stops. Simultaneous use of 7-deaza-dGTP and N^4-methyl-dCTP leads to increased incidence of false stops.

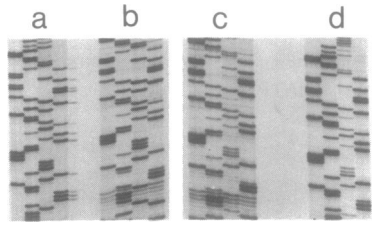

Figure 5. 8% sequencing gel obtained with Sequenase on a single-stranded template of M13mp19 DNA. Set **a**: dATP, dCTP, dGTP, and dTTP; set **b**: dATP, N^4-methyl-dCTP, dGTP, and dTTP: set **c**: dATP, N^4-methyl-dCTP, 7-deaza-dGTP, and TTP. Set **d**: dATP, dCTP, 7-deaza-dGTP, and dTTP.

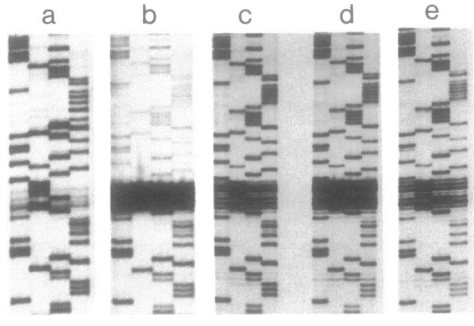

Figure 6. 6% sequencing gel obtained with Sequenase on a single-stranded template of 864I DNA. Set **a**: dATP, dCTP, dGTP, and dTTP; sets **b**, **c**, **d**, and **e**: dATP, N^4-methyl-dCTP, dGTP, and dTTP. In set **b** the nucleotide concentrations specified in Materials and Methods were used; in sets **c**, **d**, and **e**, all nucleotide concentrations were raised 5-fold, 10-fold, and 30-fold, respectively.

Given this tendency, it was not surprising to find (Fig. 6) that use of the N^4-methylcytosine analogue led to very strong false stops when Sequenase was used at 37°C to sequence the single-stranded 864I template, with its strong potential for hairpin formation. In this case, false stops were even apparent with the canonical set of dNTPs. Increasing the nucleotide concentrations in the N^4-methyl-dCTP set fivefold did reduce the false stops and increase the signal intensity in the region beyond the false stops; further increase in the nucleotide concentration to tenfold or thirtyfold the original values brought no further benefit.

Readable sequence was also obtained using N^4-methyl-dCTP on single-stranded M13mp19 template with the large fragment of *E.coli* DNA polymerase I and [^{35}S]dATPαS. With the dNTP and ddNTP concentrations given in Table II, clear, well balanced sequencing signal was obtained in all four lanes, both in the experiment with N^4-methyl-dCMP and in the control experiment with the canonical nucleotides.

Sequenase and pUC19 DNA were used to test the N^4-methylcytosine analogue in sequencing of double-stranded DNA. As seen in Fig. 7, clear sequence information was obtained with the canonical nucleotides, with the substitution of 7-deaza-dGTP for dGTP, and with the substitution of N^4-methyl-dCTP for dCTP. Fig. 7 also shows that weak band compressions seen with the canonical set of nucleotides were fully resolved using either the N^4-methyl-dCTP or the 7-deaza-dGTP substitution.

Replacement of dCTP by N^4-methyl-dCTP results in a distinct reduction in the electrophoretic mobility of the newly synthesized

Table II. Nucleotide concentrations (μM) in the termination reactions with the large fragment of *E.coli* DNA polymerase I

Canonical set								
	dATP	dCTP	dGTP	dTTP	ddATP	ddCTP	ddGTP	ddTTP
A-lane	10	100	100	100	120	–	–	–
C-lane	100	10	100	100	–	20	–	–
G-lane	100	100	100	100	–	–	30	–
T-lane	100	100	100	100	–	–	–	100
N^4-methyl-dCTP								
	dATP	N^4-me-dCTP	dGTP	dTTP	ddATP	ddCTP	ddGTP	ddTTP
A-lane	10	100	100	100	120	–	–	–
C-lane	100	10	100	100	–	2	–	–
G-lane	100	100	10	100	–	–	60	–
T-lane	100	100	100	10	–	–	–	300

polynucleotides. A direct side-by-side comparison of electrophoretic lanes in an 8% polyacrylamide gel showed the polynucleotides containing N^4-methylcytosine to be offset from the corresponding cytosine containing polynucleotides by one nucleotide unit at the 100-mer stage, by three nucleotide units at the 250-mer stage, and by six nucleotide units at the 400-mer stage, with each of these polynucleotides containing about 25% C. In line with this observation, direct measurement of the total spacing in C-runs unaffected by band compression or band distention gave values which were about 8% higher in electrophoretic lanes obtained with N^4-methyl-dCTP, compared to the lanes obtained with the canonical set of dNTPs. In contrast to the N^4-methylcytosine analogue, the use of the 7-deazaguanine analogue does not noticeably affect band mobilities.

Substitution of N^4-methyl-dCTP for dCTP does not cause a major alteration in the general rules which govern the variation of signal intensity along the electrophoretic lane. With Taq DNA polymerase, C-bands immediately following other C-bands stand out in the C-lane; this rule remains unchanged by the substitution of N^4-methyl-dCTP for dCTP or of 7-deaza-dGTP for dGTP. Sequencing ladders obtained with Sequenase are characterized by more uniform band intensities along the lanes; this remains unaffected by the use of the N^4-methylcytosine or 7-deazaguanine analogues. With the Klenow fragment of *E.coli* DNA polymerase I, the intensification of the second band in a succession of C-bands which is commonly observed with the canonical set of dNTPs is exaggerated with the use of N^4-methyl-dCTP.

To assess the potential for using N^4-methyl-dCTP in PCR amplification, reaction mixtures containing the canonical nucleotide set were compared to mixtures in which dCTP was replaced by the N^4-methylcytosine analogue, in a PCR experiment designed to amplify a 293-bp sequence of HPV16 DNA. Using a high-temperature regimen (1 min at 94°C, 2 min at 55°C, 3 min at 72°C, 25 cycles) the desired fragment was obtained with the canonical dNTPs, but not with N^4-methyl-dCTP. A low-temperature regimen (1 min at 94°C, 2 min at 37°C, 3 min at 55°C, 25 cycles), conducted with dCTP or with N^4-methyl-dCTP in the reaction mixture, cleanly produced the expected fragment as the sole amplification product. The PCR products obtained with dCTP or with N^4-methyl-dCTP showed identical band intensity when visualized in the agarose gel with ethidium bromide. No attempt was made to estimate the base fidelity achieved in our PCR experiments with N^4-methyl-dCTP.

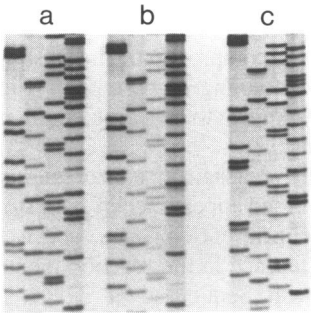

Figure 7. 6% sequencing gel obtained by double-stranded sequencing of pUC19 DNA with Sequenase. Set **a**: dATP, dCTP, dGTP, and dTTP; set **b**: dATP, dCTP, dCTP, 7-deaza-dGTP, and dTTP, set **c**: dATP, N^4-methyl-dCTP, dGTP, and dTTP.

DISCUSSION

The duplex poly(7-deazaG)·poly(C) shows only slightly lower thermal stability, compared to poly(G)·poly(C), with T_m values of 74°C and 75°C, respectively, measured in 0.2 M sodium EDTA at pH 5.3 (5). In contrast to this moderate effect, the destabilization of the G·C pair achieved by the substitution of N^4-methylcytosine for cytosine is more substantial: for the duplex formed from a fully selfcomplementary dodecadeoxyribonucleotide, conversion of only two base pairs from G·C to G·N^4-methylC resulted in a change of T_m from 60°C to 55°C, measured in 0.1 M NaCl, 0.2 mM EDTA, at 0.24 mM total nucleotide concentration (4), and for the duplex formed from the hexamer d-CGCGCG, change of the two central base pairs from G·C to G·N^4-methylC is reported to depress the T_m by about 19°C, measured in D_2O solution containing 150 mM NaCl, 10 mM phosphate (pH 7.4), and 0.2 mM EDTA at 48 mM total nucleotide concentration (3). At the polynucleotide level, complete replacement of cytosines in poly(I)·poly(C) by N^4-methylcytosines lowers the T_m by at least 50°C, while 39% replacement of cytosines by N^4-methylcytosines results in a T_m depression of 15°C (10).

On the basis of the lower thermodynamic stability of the G·N^4-methylC base pair, compared to the G·C and 7-deazaG·C pairs, one would expect two main effects in using N^4-methyl-dCTP in DNA sequencing: 1. reduced tendency for band compressions, due to the destabilization of the G·C pairs forming hairpin structures, and 2. an increased potential for false stops, if the polymerase proves unable to efficiently displace

hairpin structures in the template while it is newly forming only weak G·N^4-methylC pairs. On the electrophoretic ladder both effects will appear in close proximity; for the rules of complementarity dictate that if the newly synthesized strand contains a region conducive to hairpin formation (causing band compression) the corresponding region in the template strand has a similar propensity for hairpin formation (causing false stops).

Qualitatively speaking, these expectations were borne out in the present study. The DNA fragments synthesized in the presence of N^4-methyl-dCTP appear to be fully denatured under the conditions of sequencing gel electrophoresis, as seen by the resolution of band compression even in severe cases (potential hairpins containing 11 base pairs). With hairpin structures of weak stability, both the 7-deazaguanine substitution and the N^4-methylcytosine substitution effected full resolution of band compressions. With stronger hairpin structures, however, the 7-deazaguanine analogue is ineffective in bringing about resolution of the band compression. This is shown in the present work on hairpins containing 11 base pairs (Fig. 2 and Fig. 3), but it is already well known to be a problem in much less extreme cases (e.g. reference 2).

Compared to the 7-deazaguanine analogue, use of N^4-methyl-dCTP leads to increased appearance of false stops. The intensity of these varies for different enzymes, and is probably mainly affected by the difference in the reaction temperatures used. In chain extension with Sequenase at 37°C, N^4-methyl-dCTP caused severe false stops with the 11-base-pair hairpin; but in this case false stops were seen even with the canonical set of nucleotides. In chain extension with Taq DNA polymerase at 72°C, the use of N^4-methyl-dCTP caused only weak false stops in the case of the 11-base-pair hairpin. In sequencing with Taq DNA polymerase, for a wide range of hairpin stabilities the N^4-methylcytosine substitution effects resolution of band compression without undue formation of false stops, and with hairpins where false stops become a serious problem with N^4-methyl-dCTP, the 7-deazaguanine analogue is already far beyond its capability for resolution.

Assessing the deoxyribonucleotide analogues which are widely being used to counteract band compressions, the 7-deazaguanine substitution is not destabilizing enough in many instances, while the hypoxanthine substitution, in contrast, is too strongly destabilizing. The N^4-methylcytosine substitution is intermediate in this regard, and should be beneficial in most cases.

Apart from its destabilization of potential hairpin structures, the substitution of N^4-methylcytosine for cytosine also affects the electrophoretic mobility of polynucleotides in another way: it produces a small, but significant retardation. This is not surprising; the band spacings in sequencing ladders are known to correlate with the identity of the base added at each stage (11,12), and the electrophoretic mobility of oligonucleotides in polyacrylamide gels is strongly dependent on the base composition (13). The increased spacing caused by the N^4-methyl substitution increases the resolution in C-runs; this is beneficial because, among the canonical nucleotides, C is the one associated with the shortest band spacings (11,12). The effect of the N^4-methyl substitution on the mobility of the polynucleotides dictates that sequencing reactions should not contain mixtures of dCTP and N^4-methyl-dCTP, as this would result in the formation of isostichs containing varying proportions of cytosines and N^4-methylcytosines, which would cause band broadening.

The sequencing experiments with Taq DNA polymerase showed that this enzyme efficiently utilizes N^4-methyl-dCTP as a substrate for DNA synthesis on a template containing canonical bases. In changing from dCTP to N^4-methyl-dCTP, a twofold increase in concentration was sufficient to give satisfactory chain extension in the different lanes; however, the molar ratio of ddCTP to N^4-methyl-dCTP in the C-reaction had to be reduced considerably (compared to the ddCTP/dCTP ratio in the C-reaction using canonical nucleotides) to avoid premature average termination in the C-lane. Similar ratios were also used when 7-deaza-dGTP substituted for dGTP. The other polymerases, too, required only moderate adjustments in the nucleotide concentrations when the N^4-methylcytosine or 7-deazaguanine analogues were used.

The PCR experiments demonstrated that Taq DNA polymerase tolerates the N^4-methylcytosine moiety not only in the nucleoside triphosphate substrate, but also simultaneously in the template strand. Similar results have been reported with the substitution of 7-deaza-dGTP for dGTP in the PCR reaction (14).

N^4-Methyl-dCTP is easily synthesized. Starting from dCTP, one reaction, followed by a simple purification, produces the compound in acceptable yield. No attempt was made in this work to increase the yield.

Other N^4-alkylcytosine analogues may perform similarly to N^4-methyl-dCTP in DNA sequencing, but their use is not expected to bring any further benefits for resolution of band compressions. The N^4-methyl substitution appears to be sufficient to assure resolution of band compression in all instances, and a more bulky alkyl group may well increase the penchant for false stops in hairpin regions. The N^4,5-dimethylcytosine analogue is not a promising alternative, as poly(N^4,5-dimethylC) failed to interact with poly(dI) (15), probably due to steric interference between the two methyl groups in the configuration required for the formation of two H-bonds in the hypoxanthine·N^4,5-dimethylcytosine base pair.

REFERENCES

1. Barr, P. J., Thayer, R. M., Laybourn, P., Najarian, R. C., Seela, F., and Tolan, D. R. (1986) *BioTechniques 4*, 428–432.
2. Mizusawa, S., Nishimura, S., and Seela F. (1986) *Nucleic Acids Res. 14*, 1319–1324.
3. Fazakerley, G. V., Kraszewski, A., Téoule, R., and Guschlbauer, W. (1987) *Nucleic Acids Res. 15*, 2191–2200.
4. Butkus, V., Klimašauskas, S., PetrauskienR, L., ManelienR, Z., Janulaitis, A., Minchenkova, L. E., and Schyolkina, A. K. (1987) *Nucleic Acids Res. 15*, 8467–8478.
5. Seela, F., Tran-Thi, Q.-H., and Franzen, D. (1982) *Biochemistry 21*, 4338–4343.
6. Draper, D. E. (1984) *Nucleic Acids Res.* 12, 989–1002.
7. Wempen, I., Duschinsky, R., Kaplan, L., and Fox, J. J. (1961) *J. Amer. Chem. Soc. 83*, 4755–4766.
8. Seedorf, K., Krämmer, G., Dürst, M., Sulai, S., and Röwekamp, W. G. (1985) *Virology 145*, 181–185.
9. Chen, E. Y., and Seeburg, P. H. (1985) *DNA 4*, 165–170.
10. Brimacombe, R. L. C., and Reese, C. B. (1966) *J. Mol. Biol. 18*, 529–540.
11. Maxam, A. M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560–564.
12. Ambrose, B. J. B., and Pless, R. C. (1985) *Biochemistry 24*, 6194–6200.
13. Frank, R., and Köster, H. (1979) *Nucleic Acids Res. 6*, 2069–2087.
14. McConlogue, L., Brow, M. A. D., and Innis, M. A. (1988) *Nucleic Acids Res. 16*, 9869.
15. Żmudzka, B., Bollum, F. J., and Shugar, D. (1969) *Biochemistry 8*, 3049–3059.