Manganese citrate improves base-calling accuracy in DNA sequencing reactions using rhodamine-based fluorescent dye-terminators

Christopher Korch* and Harry Drabkin

University of Colorado Cancer Center DNA Sequencing and Analysis Core and 1 Division of Medical Oncology, School of Medicine, Box B171, University of Colorado Health Sciences Center, 4200 East Ninth Avenue, Denver, CO 80262, USA

Received November 3, 1998; Revised and Accepted January 8, 1999

ABSTRACT

While dideoxy-terminators labeled with rhodamine-based fluorescent dyes provide the most versatile method of automated DNA sequencing, variation in peak heights reduces base-calling accuracy. We describe a simple approach that uses additions of a manganese salt and the metal buffer sodium citrate (MnCit) to overcome this limitation. This modification reduces peak height variability 2-fold and significantly increases the number of accurately read bases in DNA sequences.

Automated cycle-sequencing with dye-terminators is the most versatile means of sequencing DNA. By labeling the four ddNTPs with different dyes, the sequencing reactions can be performed simultaneously in a single tube and electrophoresed in a single lane. However, a major disadvantage with dye-terminators is an unevenness in peak heights that reduces the accuracy of base-calling. Although sequencing with dye-labeled primers can produce more uniform peaks heights, it is less universally applicable and more cumbersome and expensive to perform.

Recent genetically engineered improvements to DNA polymerases (1–6) and the development of new labeling dyes have partially eliminated differences in peak heights. Two new FS dye-terminator cycle sequencing kits developed by PE Applied Biosystems (PE/ABI, Foster City, CA), which use either 4,7-dichloro-rhodamine-based dyes (d-Rhodamine dyes; 7) or Biosystems (PE/ABI, Foster City, CA), which use either energy-transfer dyes (BigDyes; 7,8), show more uniform peak heights and less suppression of Gs after As. However, because the spectral windows required for these new dyes do not match those of the filters installed in the commonly used PE/ABI’s 373A automated sequencer, they are not usable on this instrument without an expensive upgrade. Moreover, changing the filter wheel to match these newer dyes prevent use of the PE/ABI 373A for genotype analysis, again because of spectral window–dye incompatibilities.

Tabor and Richardson (1–3) and Fuller (9) reported that DNA polymerases, in the presence of manganese, discriminated less against ddNTPs than in the presence of magnesium. Addition of 10–15 mM of the metal buffer sodium isocitrate and 2 mM Mn2+ to a reaction of containing Mg2+ reduced peak height variation ~7-fold. However, Brandis et al. (10), using different Taq DNA polymerases (wild-type and KlenTaq), did not observe similar improvements, although they performed their studies using either magnesium or manganese, but not both. They concluded the improvements by the addition of manganese were not sufficiently significant to warrant its routine use.

Since the optimal Mg2+ concentration for Taq is less than that for T7 DNA polymerase, we suspected that the concentrations of Mn2+ and isocitrate previously used would be inhibitory for Taq. Furthermore, to our knowledge there have been no reports studying the effect of manganese in combination with magnesium on the so-called F667Y family of genetically engineered Taq DNA polymerases (3–6). In this report, we show that modification of the PE/ABI Prism™ Dye Terminator Cycle Sequencing kit, which uses rhodamine-based dyes, to include manganese dramatically improves the uniformity of peak heights, and this is usable on both the PE/ABI 373A and the PE/ABI 377 DNA sequencers. Combining this with other modifications of the DNA sequencing protocol results in a significant increase in base-calling accuracy at minimal cost.

The DNA template, pGEM3Zf(+) (Promega Corporation, Madison, WI), was prepared as described by Feliciello and Chiniha (11), precipitated with polyethylene glycol 8000, rinsed with 70% ethanol, dried, and dissolved in water. Alternatively, it was used as provided in the sequencing kits by PE/ABI. The sequencing primer labeled –21/pUC (5’-TGTTAAAAGGCGGCCAGT-3’) was that either supplied in the kit or a primer of the same sequence (labeled M13F), but synthesized by Gibco BRL Life Technologies (Gaithersburg, MD) and dissolved in water. Dye-terminator cycle sequencing was performed using ddNTPs labeled with rhodamine-based dyes (Rhodamine kit; PE/ABI 373A and the PE/ABI 377 DNA sequencers. Combining this with other modifications of the DNA sequencing protocol results in a significant increase in base-calling accuracy at minimal cost. The DNA template, pGEM3Zf(+) (Promega Corporation, Madison, WI), was prepared as described by Feliciello and Chiniha (11), precipitated with polyethylene glycol 8000, rinsed with 70% ethanol, dried, and dissolved in water. Alternatively, it was used as provided in the sequencing kits by PE/ABI. The sequencing primer labeled –21/pUC (5’-TGTTAAAAGGCGGCCAGT-3’) was that either supplied in the kit or a primer of the same sequence (labeled M13F), but synthesized by Gibco BRL Life Technologies (Gaithersburg, MD) and dissolved in water. Dye-terminator cycle sequencing was performed using ddNTPs labeled with rhodamine-based dyes (Rhodamine kit; PE/ABI Prism™ Dye Terminator Cycle Sequencing Ready Reaction kit with AmpliTaq DNA polymerase, FS) as recommended by the manufacturer or by the modifications described below. Sequencing primer (3–5 pmol) was mixed with 400 ng of DNA template and 3 µl of AmpliTaq DNA polymerase FS mixture in a total volume of 10 µl with double-deionized ultrafiltered (Milli-Q), sterilized water. Reaction mixtures were incubated in a GeneAmp® PCR System 9600 (Perkin-Elmer, Norwalk, CT) for 5 min at 94°C followed by 25 cycles of 10 s at 96°C, 5 s at 50°C and 4 min at 60°C with a final incubation at 4°C. The samples were processed by ethanol precipitation to remove the excess dye-labeled terminators and buffers, electrophoresed in either 4.2% polyacrylamide, bis-acrylamide (19:1) or 5.0% Long Ranger™ polyacrylamide matrix (FMIC Bioproducts, Rockland, ME), both with urea and TBE buffer, and the data were collected and analyzed with both the PE/ABI 373A and the PE/ABI 377 automated sequencers with...
accompanying software. Peak height variability was determined either by visual inspection of the electropherogram or by measuring the height of 100 individual peaks on the printed electropherograms, corresponding to nucleotides 55 (base 1 of the XhoI site) through 154 nucleotides from the 3' end of the −21/pUC-M13F primers on the plasmid pGEM3Zf(+). Coefficients of variations (CVs = 100% (standard deviation/mean)) of the peak heights were determined for all bases in these regions as well as for each of the four bases. Base-calling accuracy was determined by aligning the sequence called by the PE/ABI software to the known sequence of pGEM3Zf(+) and comparing increments of 50 bases from the 3' end of the primer initiation site to the end of the sequencing data. At the beginning of the sequence data, bases not called due to weak signal strength were excluded from the accuracy calculations.

Testing the addition of MnCl2 concentrations from 1 nM to 3 mM to a reaction containing 1.5–2.0 mM magnesium showed that the range between 0.1 and 1.0 mM most significantly affected the peak profiles produced by AmpliTaq DNA polymerase FS (an F667Y variant of Taq DNA polymerase). Figure 1A shows the results using standard rhodamine-labeled dye-terminators and magnesium (1.5–2.0 mM) alone, which contrasts with a typical peak profile (Fig. 1B) obtained by the addition of MnCl2 and tri-sodium citrate (MnCit) at concentrations of 0.5 mM each. Peak heights are distinctly more uniform with the addition of MnCit, with equimolar concentrations of both components between 0.5 and 0.6 mM being optimal (data not shown). A common problem with ddNTPs labeled with rhodamine-based dyes occurs where a G follows an A. The size of the dye-labeled G peaks is ‘suppressed’, i.e., less than both the average height of G peaks and of the height of the preceding A peak in the sequence trace. Also, the base following a suppressed G is often taller than average. These two effects can obscure the suppressed G, resulting in a base-calling error. This phenomenon is demonstrated in Figure 1A with the marked sequence 5'-CGGGGAGGGCCG-3'. Optimal additions of MnCit significantly reduced G suppression and in many cases, the G peak following an A was as tall as the A peak (see corresponding region marked in Fig. 1B).

We quantified the effect of increasing amounts of MnCit on peak height variations by measuring the height of 100 nucleotides of pGEM3Zf(+) plasmid DNA using the −21/pUC primer (both as supplied by PE/ABI). The coefficient of peak height variation decreased almost 3-fold from 66% in the absence of MnCit to 24% at 0.5–0.6 mM MnCit (Fig. 2). The overall tendency was to reduce variations of those bases that gave a strong response in the absence of MnCit and to increase the relative strength of weakly labeled bases progressively with increasing MnCit concentration. Depending on the particular base, some increases were slight, whereas others increased as much as 10-fold. Analysis of the CVs for individual nucleotides at different MnCit concentrations showed that between 0.2 and 0.6 mM MnCit, the CV for all peaks decreased, but above 0.6 mM MnCit, the CV for the G peaks uniquely increased ~1.5-fold, accounting for the slight rise in the overall CV. These results indicated that ~0.5 mM MnCit was optimal when using the PE/ABI control pGEM3Zf(+) DNA and the −21/pUC sequencing primer from the kit, which were dissolved in a Tris–EDTA buffer. In contrast, the optimal MnCit concentration in the absence of EDTA was ~0.25 mM (C.Korch and H.Drabkin, unpublished results). Therefore, to obtain the most consistent MnCit results, it is critical to avoid dissolving DNA templates and primers in buffers containing metal chelators, such as EDTA.

The relative intensity of banding patterns in fluorescent gel images and the sum of the average signal strengths for the bases indicated that MnCit could inhibit incorporation of dye-labeled ddNMPs up to 40% (data not shown). Compensation for these losses can be achieved by a combination of: (i) increasing the amount of DNA template 1.5–2-fold; (ii) increasing the number of cycles used in the sequencing reactions; and (iii) dissolving the sequencing reactions in smaller volumes.

Using pGEM3Zf(+) and the M13F (−21/pUC) primer and comparing unedited base-calls with the known sequence of this plasmid, we found that optimal additions of MnCit significantly increased the accuracy of sequence determinations by the PE/ABI computer base-calling algorithm. Addition of MnCit to the rhodamine-based sequencing reagent kit increased the total number of correctly called bases throughout the sequence traces.
by an average of 33 nucleotides within the first 650 nucleotides, and increased the length of sequence with equivalent base-calling accuracy by ~50–150 nucleotides (Fig. 3). In the commonly analyzed range of 51–450 nucleotides from the 3′ end of the primer, there was less than one base-calling error in the presence of MnCit as compared to 12 errors in its absence. The large standard deviations seen in the absence of MnCit are due to the inability of the PE/ABI sequence analysis program to accurately identify small peaks. For instance, the peak of mis-called bases between 301 and 350 nucleotides observed in the absence of MnCit corresponds to the G-rich sequence marked in Figure 1 that shows suppression of Gs by preceding As. These were correctly called in the presence of MnCit. Furthermore, both methods showed decreases in base-calling accuracy and increases in the standard deviations at both ends of the optimal range (51–450 bases). In the first 50 nucleotides this was due to weak signal strength, and above 450 nucleotides, it was due to decreased resolution of peaks by the gel matrix (data not shown).

In summary, we demonstrate that the addition of MnCit significantly improves peak height uniformity using PE/ABI rhodamine-labeled dye-terminators and AmpliTaq DNA polymerase FS in a reaction mixture containing 1.5–2.0 mM Mg²⁺. To our knowledge, this effect has not been reported earlier with the family of F667Y mutant DNA polymerases. The most dramatic effect of MnCit addition is to reduce the discrimination by AmpliTaq FS against dye-labeled ddGTPs, increasing their signal strength, and above 450 nucleotides, it was due to decreased resolution of peaks by the gel matrix (data not shown).

ACKNOWLEDGEMENTS

We are indebted to Ms Efang Li and Ms Christina Dessev for their excellent technical assistance. This work was supported by the NIH National Center for Human Genome Research grant 5R01 HG00358 to H.D. and by the NIH National Cancer Institute grant CA46934 to the University of Colorado Cancer Center for support of the UCCC DNA Sequencing and Analysis Core.

REFERENCES