

Octamer-primed sequencing technology: effects of dNTP supplementation

A. I. Kraltcheva and S. H. Hardin*

Department of Biology and Biochemistry and Institute of Molecular Biology, University of Houston, Houston, TX 77204-5513, USA

Received December 6, 2000; Revised March 2, 2001; Accepted March 12, 2001

ABSTRACT

Octamer-primed sequencing is a directed DNA sequencing strategy that employs the use of a presynthesized octamer primer library. Together with electronic octamer sequencing technology (eOST) it provides a faster, less expensive way to obtain DNA sequence information and can be used as a valuable tool for gap closure in large-scale genomic sequencing. In this paper we discuss the effect of dGTP/TTP supplementation on octamer sequencing. We show that addition of 75 μ M dGTP and 5 μ M TTP can improve the sequencing success rate by increasing the length and accuracy of generated sequence information. We also discuss the effect of template base composition immediately downstream of the octamer primer on the outcome of octamer sequencing.

INTRODUCTION

Octamer sequencing technology (OST) is a primer-directed DNA sequencing method in which an individual octamer oligonucleotide is selected from a presynthesized primer library and used to prime a DNA sequencing reaction (1–3). The availability of such a library eliminates delays associated with designing and synthesizing gene-specific primers. However, a general problem associated with such libraries is their size (~1000 primers comprise our working octamer library), which often makes primer selection time consuming. To bypass this problem a computer program, electronic OST (eOST), was developed (4). This program leads to optimization of octamer selection through automation of the process. The core of eOST consists of a dynamic folding algorithm that predicts the possible formation of secondary structures within the template region bordering the octamer-binding site and determines the free energy of these potentially inhibitory structures. Using this program, primers located in regions that are predicted to contain such structures can be avoided.

Without additional optimization another potential problem of presynthesized short primer libraries is that short primers are more likely than a traditional length (18–25 base) primer to fail to initiate the sequencing reaction, rendering the libraries

less efficient than acceptable. Possible solutions to the problem are to use modified primers (5) or to improve the reaction conditions (3) in order to increase the success rate of short oligo-primed sequencing reactions.

The octamer primer library was initially developed and optimized for use in isotopic sequencing reactions (1). Later, with the development of fluorescent sequencing technology, this technique was adapted for use with rhodamine fluorescent chemistry (3). Subsequently, a more sensitive dye chemistry, termed BigDye chemistry, was introduced (6). However, we noted that this new BigDye chemistry produces a reduced success rate with octamer primers, compared to the rhodamine chemistry. A direct comparison of the two chemistries was performed using the Re2 template, a 4993 bp upstream region of the *Drosophila melanogaster timeless* gene cloned into a double-stranded plasmid vector (template described in 2). Of the initially tested 11 primers only one did not produce sequence data with the older rhodamine chemistry. However, using the BigDye chemistry three octamers repeatedly failed to prime the sequencing reaction. A notable difference between the two chemistries is their deoxynucleotide (dNTP) composition. Rhodamine chemistry contains dATP, dCTP, TTP and, rather than dGTP, dITP to minimize band compression (7,8). In addition to dITP, BigDye chemistry substitutes TTP with dUTP. Presence of this nucleotide facilitates incorporation of the T terminator, leading to a better peak pattern (8).

This chemistry difference, together with the overall lower success rate of fluorescent octamer sequencing compared to isotopic octamer sequencing, prompted our investigation of possible reasons for octamer-primed reaction failure with either fluorescent chemistry. Two possible reasons are a reduced incorporation efficiency of the non-natural bases dITP and dUTP, relative to the natural bases dGTP and TTP, respectively (9,10), and/or a weakened duplex stability associated with incorporation of dITP and dUTP into the DNA strand (11,12). For these and other reasons it is possible that either or both of the substituted dNTPs (dITP and dUTP) have a net effect of reducing extension from the octamer. We report that supplementing reactions with a combination of dGTP and TTP can increase sequencing reaction success rate and, additionally, that the nucleotides incorporated into the growing DNA chain, as specified by the template, should be used to identify the optimal sequencing primer.

*To whom correspondence should be addressed. Tel: +1 713 743 2686; Fax: +1 713 743 2636; Email: shardin@uh.edu

MATERIALS AND METHODS

Oligonucleotide synthesis

Library DNA oligonucleotides used in regular and supplemented octamer-primed DNA sequencing were synthesized by Genosys Biotech (all primers with prefix 384) or Operon Technologies (all primers with prefixes AB, AT, BB, BT, CB, CT and OP). The oligonucleotide used as a template in the BigDye primer extension experiments was synthesized by MWG-Biotech.

Octamer-primed DNA sequencing

Automated octamer-primed DNA sequencing reactions were performed as described (2). The 10 μ l reaction mix of the control (non-supplemented) reactions contained 50 pmol octamer primer, 200 ng plasmid template and 2 μ l of ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (original version). In the supplementation experiments 15 pmol primer were used and different concentrations of freshly diluted dGTP and TTP, either alone or in combination, were added to the sequencing reaction. The cycling regimen was 99 cycles of 96°C for 10 s, 40°C for 1 min and 60°C for 4 min. The sequencing reactions were precipitated, resuspended in 3 μ l of loading buffer (50 mg/ml Blue Dextran, 5 parts deionized formamide, 1 part 25 mM EDTA pH 8.0). The samples were heated to 96°C for 2 min and half of the reaction was loaded onto a 36 cm (5.3% RapidGel-XL; US Biochemical) sequencing gel. Data was collected in 2X mode on an ABI PRISM 377 DNA Sequencer and was manually edited using Sequencher v3.0 or v3.1 (GeneCodes).

Analysis of DNA sequence read length and data accuracy

The sequence data generated in octamer sequencing was trimmed automatically using Sequencher v3.0 or v3.1. The following trimming parameters were used: (i) the sequences were trimmed until the first 25 bases at the 3'- and 5'-termini contained fewer than three ambiguities; (ii) the leading and trailing ambiguous bases were removed. To obtain the 500 base read all the sequence reads obtained with an individual primer before and after supplementation were aligned in a contig and trimmed beginning at a common starting point. In this analysis data accuracy for the 500 base read was determined using Sequencher v3.0 or v3.1.

Primer extension assays

Primers were 5'-phosphorylated with T4 polynucleotide kinase (Promega). The kinase reaction contained 1 μ g octamer primer, 300 μ Ci [γ -³²P]ATP (NEN Life Science Products) and the appropriate buffer (Promega). All extension reactions were performed in a total volume of 10 μ l using ~50 000 counts of radiolabeled octamer.

Extension reactions with *Taq* polymerase (Promega) also contained dGTP, dITP, TTP or dUTP at a final concentration of 20 or 200 μ M, 2 mM MgCl₂, *Taq* polymerase reaction buffer (Promega) and 200 ng plasmid template. The cycling regimen was 15 or 45 cycles of 96°C for 10 s, 40°C for 1 min and 72°C for 4 min.

Reactions with AmpliTaq polymerase FS (Perkin Elmer) contained 2 μ l BigDye premix, 10 ng oligonucleotide template, 1.5 μ l 5 \times sequencing buffer (containing 80 mM Tris and 2 mM MgCl₂) and, if supplemented, 75 μ M dGTP and 5 μ M TTP.

The cycling regimen was 15, 45 or 99 cycles of 96°C for 10 s, 40°C for 1 min and 60°C for 4 min. The sample volume was reduced by lyophilization to ~2.5 μ l, resuspended in 2 μ l of stop solution (containing 99% formamide, 0.1% bromophenol blue and 0.1% xylene cyanol) and electrophoresed at 2400 V for 1.5–2 h through a 40 cm, 20% denaturing polyacrylamide gel. Data were collected and the results were quantified with a Fuji MacBas1000 Phosphorimager (Fuji Medical Systems, USA).

RESULTS AND DISCUSSION

Comparison of the relative incorporation efficiencies of natural (dGTP, TTP) and non-natural (dITP, dUTP) nucleotides into the growing DNA chain by *Taq* polymerase

Previous experiments demonstrated that certain DNA polymerases incorporate dITP and dUTP less efficiently than the corresponding natural nucleotides (9,10). We conducted primer extension experiments to compare the relative rates of incorporation of dITP and dGTP and also of dUTP and TTP by *Taq* polymerase, a predecessor of the polymerase currently used in fluorescent sequencing (AmpliTaq FS). The experiments were performed at equal molar concentrations (20 or 200 μ M) of either the natural or the non-natural dNTP. Preliminary experiments showed that higher concentrations of dNTPs (1 mM) inhibited *Taq* polymerase.

The relative incorporation rates for these natural versus non-natural nucleotides were determined in the process of template-directed addition of two (for TTP and dUTP) and three (for dGTP and dITP) consecutive nucleotides. We found that under these conditions the natural bases dGTP and TTP were incorporated by the polymerase more efficiently than the non-natural bases dITP and dUTP, respectively (Fig. 1). Depending on both the amount of dNTP and the number of cycles, the percentage of completely extended product resulting from dITP incorporation ranged from 14 \pm 2% (20 μ M, 45 cycles) to 19 \pm 2% (200 μ M, 45 cycles). The corresponding percentage resulting from dGTP incorporation was between 29 \pm 5% (20 μ M, 15 cycles) and 46 \pm 6% (20 μ M, 45 cycles) (Fig. 1B). A similar trend is observed with dUTP and TTP. The percentage of completely extended product with dUTP ranged from 27 \pm 2% (20 μ M, 45 cycles) to 32 \pm 2% (200 μ M, 45 cycles) and with TTP from 44 \pm 1% (200 μ M, 15 cycles) to 55 \pm 4% (20 μ M, 45 cycles). As per expectation, our experiments indicate that the polymerase does not incorporate the non-natural nucleotides as efficiently as their natural counterparts. The accumulation of incompletely extended products was more pronounced during the incorporation of dITP than that of the other non-natural nucleotide dUTP, as compared to the corresponding natural nucleotides (Fig. 1A). Another difference in the incorporation patterns of the natural and the non-natural nucleotides was that in the second case there was a relatively reduced increase in completely extended product amounts at higher numbers of cycles (Fig. 1B). This may be due either to the accumulation of incompletely extended product as a result of pausing of the polymerase during addition of the non-natural nucleotides or to less efficient extension from a non-natural base.

It is possible that the lower rate of incorporation of dITP and dUTP causes reduced extension from the octamer, which leads

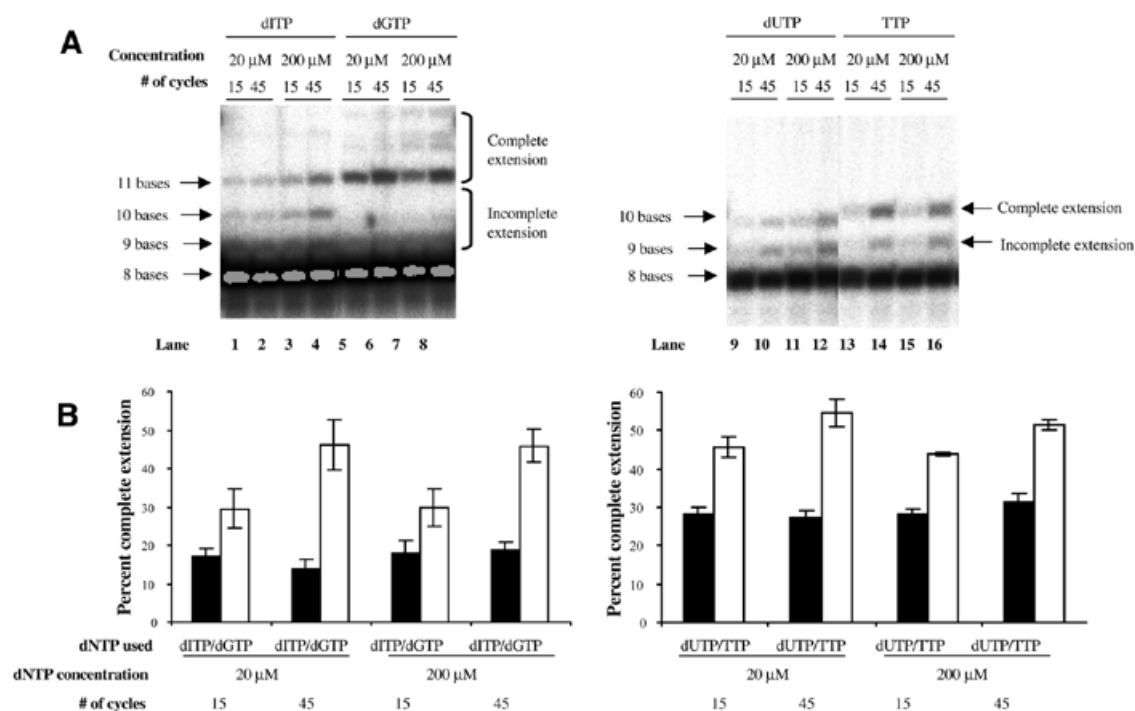


Figure 1. Efficiency of incorporation of non-natural (dITP, dUTP) and natural (dGTP, TTP) nucleotides. Percent completely extended product for reactions containing non-natural (black bars) and natural (white bars) nucleotides are plotted. dNTP, concentration and cycle number are indicated. (A) Primer extension reactions contain only the indicated dNTP. The products generated in these reactions were separated on a 20% denaturing polyacrylamide gel. (B) The results of the primer extension experiments in (A) were quantified with a phosphorimager.

to more rapid dissociation of the primer or minimally extended primer from the template prior to the extension step. Our results indicate that the lower success rate of OST may be due to less efficient incorporation of these nucleotides, hence, supplementation with natural nucleotides could lead to an improvement in the sequencing process.

Determining the optimal dGTP/TTP concentration in supplementation experiments

Initially we tested supplementation of sequencing reactions with 200 μM dGTP or 200 μM TTP, either separately or in combination, on three non-working and four working primers (Table 1). This approach enabled us to determine both beneficial as well as adverse effects of dNTP supplementation on the resulting sequencing data. A non-working primer was defined as a primer that failed to produce detectable sequence products or repeatedly produced low quality data in a standard (non-supplemented) sequencing reaction. The data generated in non-supplemented control reactions and in dNTP-supplemented reactions were collected with an ABI PRISM 377 DNA Sequencer and analyzed with Sequencher.

Initial primer amounts used in the sequencing reactions varied between 3 and 37 pmol/rxn. The supplemented reactions were run in parallel with non-supplemented controls to determine the effect of dGTP/TTP supplementation on signal intensity and overall data quality. dGTP supplementation alone transformed two of the three previously non-working octamers into working primers, producing a sequence read of 500 or more bases. The third non-working octamer (384.179) generated a read of 94 bases when dGTP was used in combination with

TTP (Table 1). However, supplementation with 200 μM dGTP led to a substantial increase in the background signal that resulted in a decrease in the accuracy of the sequence generated by two of the control working primers (384.609 and 384.116). The other two control primers failed to produce extension products at this concentration of dGTP. TTP alone did not have any effect on non-working primers and negatively affected the working ones, with effects ranging from reduced read length and accuracy to failure to produce detectable extension products.

Based on the increased background seen with 200 μM dGTP and/or TTP, we decided to test lower concentrations of these nucleotides in cycle sequencing reactions. Subsequently, reactions were supplemented with 10–200 μM of either dGTP or TTP to determine the optimal concentration of each nucleotide (Table 2). It was found that concentrations between 50 and 100 μM dGTP had a positive effect on non-working primers and were not deleterious for working primers. Again TTP alone failed to improve the reaction outcome for non-working primers, but at 10 μM concentration increased the accuracy of the read generated by the tested control primer (384.066). This observation, together with the fact that one octamer (384.179) failed to produce acceptable data with only dGTP supplementation, suggested the use of a combination of the two nucleotides in future experiments.

Our next set of experiments was performed to determine which combination of dGTP/TTP concentrations was optimal for reaction supplementation in octamer sequencing. Combinations of different nucleotide concentrations were tested on both working and non-working primers (Table 3). Based on these and subsequent experiments 75 μM dGTP and 5 μM TTP was

Table 1. Initial dNTP supplementation of working and non-working primers

Primer ID	Supplementation	# of Bases Called (8hr collection; 36 cm read)	Errors or Ns (non-edited)	Percent correct (over bases called)	Errors or Ns (over 500 bases)	Percent correct (over 500 bases)
Non-working primers						
384.035	none	0	na	na	na	na
384.035	200uM dGTP	500	46	90.8	46	90.8
384.035	200uM dTTP	0	na	na	na	na
384.035	200uM dGTP/dTTP	0	na	na	na	na
384.179	none	0	na	na	na	na
384.179	200uM dGTP	0	na	na	na	na
384.179	200uM dTTP	0	na	na	na	na
384.179	200uM dGTP/dTTP	94	41	56.38	na	na
384.584	none	0	na	na	na	na
384.584	200uM dGTP	621	74	88.08	51	89.8
384.584	200uM dTTP	0	na	na	na	na
384.584	200uM dGTP/dTTP	0	na	na	na	na
Working primers						
384.609	none	725	32	95.59	3	99.4
384.609	200uM dGTP	705	41	94.18	12	97.6
384.609	200uM dTTP	422	30	92.89	na	na
384.609	200uM dGTP/dTTP	684	49	92.84	20	96
384.066	none	737	17	97.69	0	100
384.066	200uM dGTP	0	na	na	na	na
384.066	200uM dTTP	734	49	93.32	16	96.8
384.066	200uM dGTP/dTTP	0	na	na	na	na
384.116	none	733	30	95.91	2	99.6
384.116	200uM dGTP	722	95	86.84	54	89.2
384.116	200uM dTTP	0	na	na	na	na
384.116	200uM dGTP/dTTP	0	na	na	na	na
384.402	none	742	22	97.03	0	100
384.402	200uM dGTP	0	na	na	na	na
384.402	200uM dTTP	732	58	92.08	15	97
384.402	200uM dGTP/dTTP	0	na	na	na	na

Three non-working (384.035, 384.179, 384.584) and four working (384.609, 384.066, 384.116, 384.402) primers were supplemented with 200 μ M dGTP and 200 μ M TTP alone or in combination. The sequence read length, number of errors or ambiguities (Ns) and the percent data accuracy over the non-edited sequence, as well as over 500 bases of trimmed sequence, are indicated.

identified as the optimal combination for supplementation (Table 3 and data not shown).

To ensure that these results were not specific to a particular primer or template, this combination was tested on a number of working and non-working primers used to sequence seven different templates (Table 4). Library octamers used in these experiments were selected using the eOST program and only primers positioned in template regions not associated with predicted secondary structures ($\Delta G > -5$ kcal/mol) were chosen for analysis (Table 4). Our previous research has shown that there is a correlation between the ΔG value of secondary structures and the success rate of octamer extension in sequencing (2,4). A $\Delta G > -5$ kcal/mol was identified as a good predictor of OST reaction success over a 50 base template window (4).

Sixteen non-working primers were tested in supplemented reactions. Of those, 11 (~69%) were able to successfully prime in cycle sequencing reactions after the addition of 75 μ M dGTP and 5 μ M TTP (Table 4). Five (~31%) of the supplemented non-working primers did not benefit from dNTP supplementation. Addition of dGTP and TTP had a negative effect on three (~38%) of the eight working primers that were tested in these experiments. The reason for this detrimental effect is not known, but the high percentage of working

primers that were negatively affected by nucleotide supplementation prompted us to redirect our efforts towards improving non-working primers only, rather than towards determining universal OST reaction conditions.

Different primer concentrations were used in supplementation of non-working primers. We observed that the amount of primer used in supplemented reactions affects the quality of the generated sequence. Specifically, a decrease in primer amount to ~10 pmol/rxn was directly correlated with an increase in the sequence information accuracy (Table 5). We determined that 15 pmol of an octamer primer in a 10 μ l reaction provide the best results, as determined by read length and accuracy of the sequence read (Table 5 and data not shown).

dGTP/TTP supplementation of sequencing reactions facilitates primer extension

To monitor the effect of dNTP supplementation on octamer-primed sequencing reactions we performed primer extension experiments under the conditions the standard BigDye sequencing premix. Radiolabeled octamer was used in these experiments to track primer extension. Control (non-supplemented) reactions and reactions supplemented with 75 μ M dGTP and 5 μ M TTP were cycled 15, 45 or 99 times. The major extension product in these experiments results from the

Table 2. Determining the optimal concentrations of dGTP and TTP in supplemented reactions

Primer ID	Supplementation	# of Bases Called (8hr collection; 36 cm read)	Errors or Ns (non-edited)	Percent correct (over bases called)
Non-working primer				
384.035	none	0	na	na
384.035	10 μ M dGTP	615	30	95.12
384.035	50 μ M dGTP	611	19	96.89
384.035	100 μ M dGTP	530	28	94.72
384.035	200 μ M dGTP	603	21	96.52
384.035	10 μ M dTTP	0	na	na
384.035	50 μ M dTTP	0	na	na
384.035	100 μ M dTTP	0	na	na
384.035	200 μ M dTTP	0	na	na
384.179	none	0	na	na
384.179	10 μ M dGTP	0	na	na
384.179	50 μ M dGTP	197	28	85.79
384.179	100 μ M dGTP	293	66	77.47
384.179	200 μ M dGTP	0	na	na
384.179	10 μ M dTTP	0	na	na
384.179	50 μ M dTTP	0	na	na
384.179	100 μ M dTTP	0	na	na
384.179	200 μ M dTTP	0	na	na
Working primer				
384.066	none	738	19	97.43
384.066	10 μ M dGTP	742	36	95.15
384.066	50 μ M dGTP	726	12	98.35
384.066	100 μ M dGTP	722	10	98.61
384.066	200 μ M dGTP	726	12	98.35
384.066	10 μ M dTTP	715	9	98.74
384.066	50 μ M dTTP	725	21	97.10
384.066	100 μ M dTTP	731	31	95.76
384.066	200 μ M dTTP	363	21	94.21

Two non-working (384.035, 384.179) and one working (384.066) primer were supplemented with dGTP or TTP at 10–200 μ M (final concentration). The number of bases called, the number of errors or ambiguous bases present in the non-edited sequence and the percent accuracy over the number of bases called are shown.

Table 3. Determining the optimal dGTP/TTP combination in supplemented reactions

Primer ID	Supplementation dGTP (μM)	TTP (μM)	# of Bases Called (8hr collection; 36 cm read)	Errors or Ns (non-edited)	Percent correct (over bases called)	Errors or Ns (over 500 bases)	Percent correct (over 500 bases)
Non-working primer							
384.035	60	2	622	30	95.18	6	98.82
384.035	70	2	604	19	96.85	3	99.41
384.035	80	2	622	24	96.14	5	99.02
384.035	60	8	618	30	95.15	6	98.82
384.035	70	8	616	26	95.8	6	98.81
384.035	80	8	622	32	94.9	8	98.43
384.179	60	2	322	20	93.79	na	na
384.179	70	2	0	na	na	na	na
384.179	80	2	599	56	90.65	37	92.70
384.179	60	8	624	26	95.8	12	97.64
384.179	70	8	601	40	93.34	19	96.27
384.179	80	8	631	36	94.30	14	97.25
Working primer							
384.066	60	2	618	10	98.38	4	99.32
384.066	70	2	618	5	99.19	2	99.67
384.066	80	2	611	3	99.51	0	100.00
384.066	60	8	623	10	98.39	1	99.83
384.066	70	8	624	17	97.28	5	99.16
384.066	80	8	615	9	98.54	2	99.66

Primer ID, combination of dGTP and TTP assayed, number of bases called, number of errors and ambiguous bases called and percent accuracy over either the non-edited sequence or the 500 base trimmed sequence are indicated.

terminal transferase activity associated with the enzyme in the sequencing premix. Our results indicate that addition of natural dNTPs led to an increase in the rate of octamer extension (Fig. 2). In three independent experiments, after each specified number

Table 4. Summary of the dGTP/TTP supplementation

Primer ID	Primer sequence	Template	Delta G	Primer source	Supplementation effect
Non-working primers (require supplementation)					
384.631	CTCTCCCG	1	-3.90	Genosys	No effect
OP634	CTCCGGCA	2	-4.00	Operon	No effect
OP472	GGACAGGC	3	-2.10	Operon	No effect
CT969	CCAGTGCC	4	-1.70	Operon	No effect
CB1423	CGCACTGG	4	-1.70	Operon	No effect
384.012	CCTGTCCC	5	-3.00	Genosys	Positive effect
384.035	CTCTCGCC	6	-2.20	Genosys	Positive effect
384.179	GAGCCCTC	6	-2.40	Genosys	Positive effect
384.451	GGATCGGC	5	-2.60	Genosys	Positive effect
384.584	GACTCGGG	6	-2.50	Genosys	Positive effect
BT2238	CGGTCCG	7	-2.40	Operon	Positive effect
CB495	CTCCGCTG	4	-4.60	Operon	Positive effect
CT1897	CAGCGGAG	4	-4.60	Operon	Positive effect
OP090	GGGTGTCC	2	-2.60	Operon	Positive effect
OP109	CTGGAGC	5	-2.50	Operon	Positive effect
OP493	GCTCCAGG	5	-2.50	Operon	Positive effect
Working primers (do not require supplementation)					
384.062	GCTGTGCC	6	-3.10	Genosys	Negative effect
384.268	CAGACCCG	1	0.50	Genosys	Negative effect
384.402	GGCTGTGG	6	-3.30	Genosys	Negative effect
384.066	GCCGATCC	6	0.80	Genosys	No effect
384.068	CCCACTCC	6	-0.90	Genosys	No effect
384.116	CCAGAGGC	6	-0.60	Genosys	No effect
OP571	GAGCTGCG	5	-2.40	Operon	No effect
384.609	CTGTGGGG	6	-0.30	Genosys	No effect

Primer ID, sequence, template project, eOST predicted free energy (ΔG) in kcal/mol, primer source and supplementation effect are indicated. The last column summarizes the effect of supplementation with 75 μ M dGTP and 5 μ M TTP on primers that are either classified as non-working (i.e. they do not produce quality sequence extension products in non-supplemented reactions) or working (i.e. they do not require supplementation to produce sequence extension products).

Table 5. Effect of primer concentration in reactions supplemented with 75 μ M dGTP and 5 μ M TTP

Primer ID	Primer pmol/rxn	# of Bases Called (8hr collection; 36 cm read)	Errors or Ns (non-edited)	Percent correct (over bases called)	Percent correct (over 500 bases)	Signal Intensity				
						G	A	T	C	
384.035	5	565	46	92	95	21	40	35	67	
384.035	14	755	168	78	96	39	114	77	156	
384.035	25	121	12	90	na	92	272	192	402	
384.035	50	26	didn't assemble with consensus	na	na	149	250	272	194	
384.179	3	784	244	69	77	22	41	36	57	
384.179	10	680	17	98	99	84	224	135	359	
384.179	25	559	21	96	96	170	477	279	726	
384.179	50	595	40	93	94	222	619	368	1024	

Primer ID, primer amount, read length, number of errors or ambiguous base calls, percent accuracy and signal intensity for two representative non-working primers that benefit from nucleotide supplementation are indicated.

of cycles the amount of completely extended product was consistently higher for the supplemented reactions. With supplementation this amount ranged from $49 \pm 2\%$ (15 cycles) to $74 \pm 1\%$ (99 cycles) and without supplementation from $25 \pm 2\%$ (15 cycles) to $60 \pm 4\%$ (99 cycles). The effect of dGTP/TTP supplementation is most pronounced in the initial 15 cycles. It is possible that addition of dGTP and TTP, due to their higher rates of incorporation, facilitates initial extension of the octamer during the sequencing reaction and prevents its dissociation from the template, leading to a reduced amount of incompletely extended product. Specifically, after supplementation

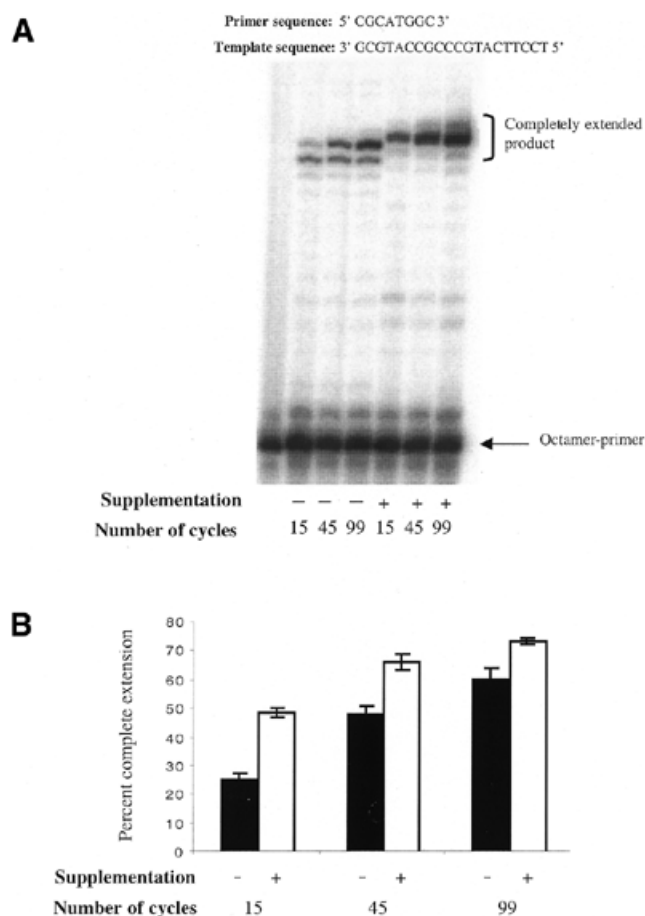


Figure 2. Effect of dNTP supplementation on the extension of a representative octamer. (A) Control (non-supplemented) sequencing reactions were performed as described (2). The supplemented reactions contained 75 μ M dGTP and 5 μ M TTP. Products generated in these reactions were separated on a 20% denaturing polyacrylamide gel. Number of cycles and presence of supplementation are indicated. (B) The results of three independent sets of primer extension experiments were quantified with a phosphorimager. Percent completely extended product for control (black bars) and supplemented (white bars) reactions are plotted.

the percentage of incompletely extended products decreased from ~75 to ~51% during the initial 15 cycles. The observed incorporation and extension rate differences could be crucial for the outcome of the sequencing reaction.

Effect of dNTP supplementation on signal intensity, read length and data accuracy

dNTP supplementation of non-working primers leads to an increase in signal intensity and sequence data quality, compared to non-supplemented control reactions with these primers (Fig. 3). Comparison of chromatograms of sequences generated by a representative primer before and after supplementation reveals that supplementation dramatically improves the quality of the sequencing data. The peak pattern is more uniform and peak height is increased, leading to fewer ambiguities and errors in base calling (Fig. 3A and B). Additionally, as a result of the increase in signal intensity, detection and tracking of sequencing reaction products is facilitated (Fig. 3C).

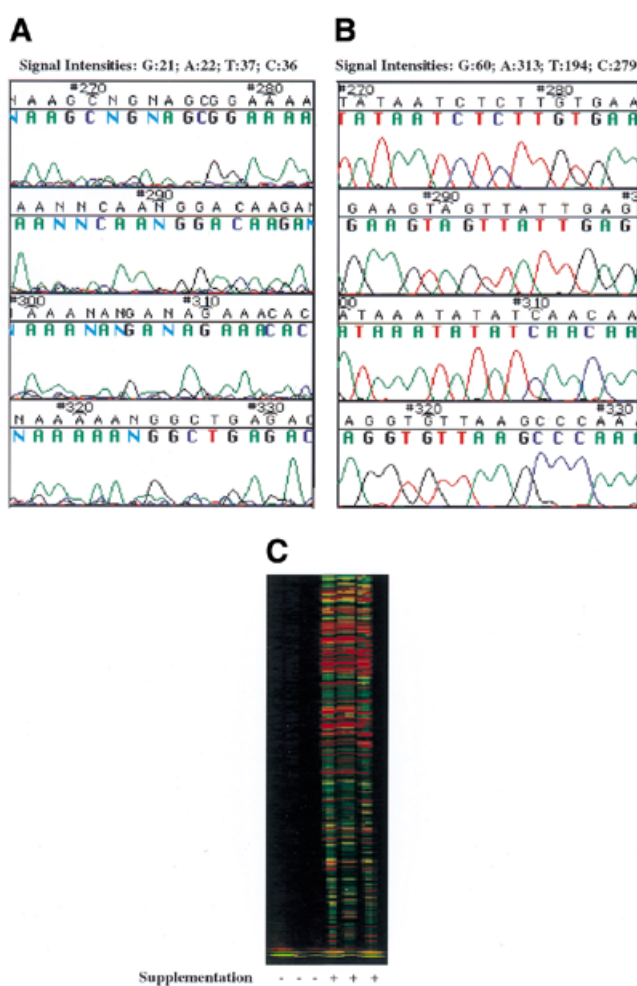


Figure 3. Sample chromatograms and compressed gel file of sequence data generated by a representative non-working primer before and after supplementation. Sequencing reactions were performed as described (2). The reactions were run in triplicate; the first set of three reactions was generated in the absence and the second set in the presence of nucleotide supplementation. (A) Control (non-supplemented) reaction. (B) Reaction supplemented with 75 μ M dGTP and 5 μ M TTP. (C) Compressed gel file of the reactions using the primer in (A) and (B).

dNTP supplementation of 16 octamers that initially failed in fluorescent sequencing led to an increase in the length of the sequencing information generated by 11 of these primers. Read length increases ranged from moderate (34 bases, primer CB495) to substantial (786 bases, primer BT2238) (Fig. 4A). Of these 11 primers, eight generated reads of 500 bases or more with dGTP/TTP supplementation. Primer 384.451 produced a high quality sequence read of 491 bases. Sequence accuracy was also positively influenced by nucleotide supplementation. Six of the nine octamers that generated at least 491 bases of sequence information after supplementation produced this data at an average of >98% accuracy (Fig. 4B).

Sequencing reactions primed by traditional length oligonucleotides can also benefit from dGTP/TTP supplementation. Preliminary data show that the presence of 75 μ M dGTP and 5 μ M TTP led to an ~20% increase in sequence read length obtained from sequencing GT-rich templates (M.Reddy, personal communication).

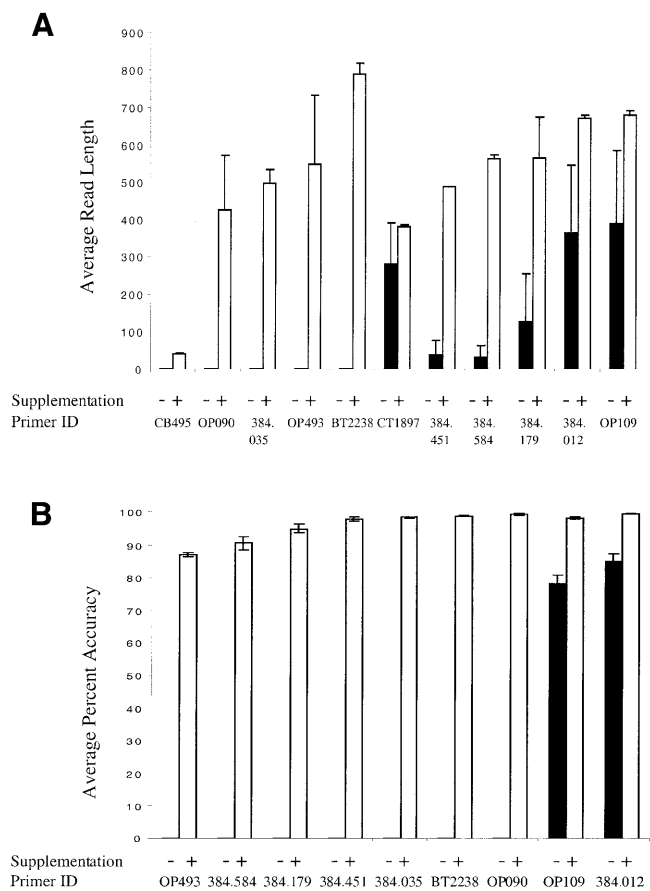


Figure 4. Effect of dGTP/TTP supplementation on sequencing information quality. Black bars correspond to non-supplemented and white bars to supplemented reactions. Plus indicates presence and minus indicates absence of supplementation. (A) The average sequence read length and standard error generated by library octamers are shown. Individual primers were tested in control (non-supplemented) reactions and in reactions supplemented with 75 μ M dGTP and 5 μ M TTP. Each reaction was repeated at least three times. (B) Sequence data accuracy obtained by individual primers. Accuracy is calculated as the percentage of the correctly called bases over 500 bases. In the case of primer 384.451 accuracy was determined over 491 bases.

Effect of template base composition on OST success or failure prior to supplementation

There are several factors that influence the outcome of an octamer-primed sequencing reaction. We have demonstrated that sequencing reaction success is influenced by the stability of predicted template secondary structures (2,4) and by the stability of the primer–template duplex, which is determined by base composition and base order. As we now report, another key factor in OST sequencing reaction success or failure is the identity of the bases incorporated into the growing DNA chain.

According to our hypothesis an octamer sequencing reaction is negatively impacted by replacement of the natural dGTP and TTP with the non-natural dITP and dUTP, respectively. Therefore, we predicted that the number of dGTP and/or TTP specified for incorporation into the growing DNA chain would influence the outcome of the sequencing reaction. We expected that higher numbers of dGTP and TTP will result in unsuccessful sequencing reactions due to the replacement of

Table 6. Bases specified for incorporation into the growing DNA chain immediately 3' of the octamer

Primer ID	12-base sequence 3' from primer	Base composition			
		G	A	T	C
Non-working primers (require supplementation)					
<i>Do not benefit from supplementation</i>					
384.631	ACCGATTCCGGT	3	2	4	3
OP634	AGCCCGTCGTGG	5	1	2	4
OP472	GGCGTTTAAACAT	3	3	4	2
CT969	CTTCGCCGAGGT	4	1	3	4
CB1423	CTGCCCTGAGGTT	4	1	4	3
Total # of bases		19/60	8/60	17/60	16/60
Percent of total # of bases		31.7	13.3	28.3	26.7
<i>Benefit from supplementation</i>					
384.012	TGGACGAGCAGG	6	3	1	2
384.035	GCGCTGCTCCGT	4	0	3	5
384.179	GATTCGGTTTCGT	4	1	5	2
384.451	GTGAGAACATG	4	4	3	1
384.584	CAGCTTGACCCG	3	2	2	5
BT2238	TGTTGTTTCTTA	2	1	9	0
CB495	CCGAGCCGGCCG	5	1	0	6
CT1897	TATCGGTTGCAT	3	2	5	2
OP090	GGGGTGGATCCG	7	1	2	2
OP109	TGCTTCCAAGG	3	2	4	3
OP493	TGTTGTCCTTG	4	0	6	2
Total # of bases		45/132	17/132	40/132	30/132
Percent of total # of bases		34.1	12.9	30.3	22.7
Working primers (do not require supplementation)					
384.062	ATCACTGCGCAG	3	3	2	4
384.268	AAAATGGACGAC	3	6	1	2
384.402	TATTAACAACAA	0	7	3	2
384.066	GCCCAAAGTTTC	2	3	3	4
384.068	ATCCCTCACAAA	0	5	2	5
384.116	ACCCTCACCTGA	1	3	2	6
OP571	CAAGTATTTCAA	1	5	4	2
384.445	TCCCGCACCGTT	2	4	5	1
384.507	TCAATTGCGATG	3	3	4	2
384.617	TGGGACAGGTGT	6	2	3	1
384.730	CCAGACTTCGCT	2	2	3	5
384.123	CGCCCATCTCCA	1	2	2	7
384.141	GGGCATGAAGGA	6	4	1	1
384.474	TATGCCGTATAC	2	3	4	3
384.507	CACPTCTCTTG	1	1	6	4
384.525	CITTAATCTTCCA	0	3	5	4
384.179	GCAGATTGATGG	5	3	3	1
384.371	TCTCGCATCTGT	1	1	6	4
384.473	CCTCTCCTTTAC	0	1	5	6
384.699	CTTCAGCGACAG	3	3	2	4
384.737	CGAGGCCAGCAT	4	3	1	4
384.49	CCAAATGTGAGC	3	4	2	3
384.353	ATAAAACCTTCA	0	6	3	3
384.433	AAAAGCTGATGA	3	6	2	1
384.732	AAGTACAGGGAG	5	5	1	1
BB2136	TGGTGTCTCAT	3	1	6	2
BB1207	TTCCAAGTGATC	2	3	4	3
BT1992	AGTTTTTGGTAA	3	3	6	0
BB154	ACCCAAAAAATT	0	7	2	3
BB1858	TACAAAAACTGC	1	6	2	3
BT256	TTGAGCAACCCA	2	4	2	4
BT1185	AGATCACAGAAA	2	7	1	2
BB400	TACCCAAATGCC	1	4	2	5
BT534	AAGCCTTTACTG	2	3	4	3
AB1059	TGTGTTGTCTGT	4	0	7	1
AB694	AGCTTGTCCGGAT	4	2	4	2
AB857	CCTGGTAGTTGA	4	2	4	2
AB437	CCGCGGAGGCTT	5	1	2	4
AB605	AAGCGTCCGTCT	3	2	3	4
AT684	GCAATCAGCATGA	3	4	2	3
AT521	GCCGGTGTGCAC	5	1	2	4
T941	CTTTAGTATGT	2	2	7	1
AT773	GCCAAAGTCAAC	2	5	1	4
AT319	GCCAAAGTCAAC	1	4	2	5
384.609	AGCGTATATATT	2	4	5	1
Total # of bases		108/540	153/540	143/540	136/540
Percent of total # of bases		20	28.3	26.48	25.19

Primer ID, DNA sequence of the 12 base window immediately 3' of the octamer primer (the first 12 incorporated bases) and summary of the number of each nucleotide are indicated. All of the listed non-working primers were supplemented with dGTP and TTP. Most of the listed working primers were not assayed in dNTP supplementation reactions.

these nucleotides with dITP and dUTP, which have a lower incorporation efficiency.

To test our hypothesis we examined the numbers and identities of nucleotides immediately 3' of the octamer used to prime the sequencing reaction. Specifically, the number of occurrences of each nucleotide in this 12 base window were counted and these data plotted (Table 6 and Fig. 5). As

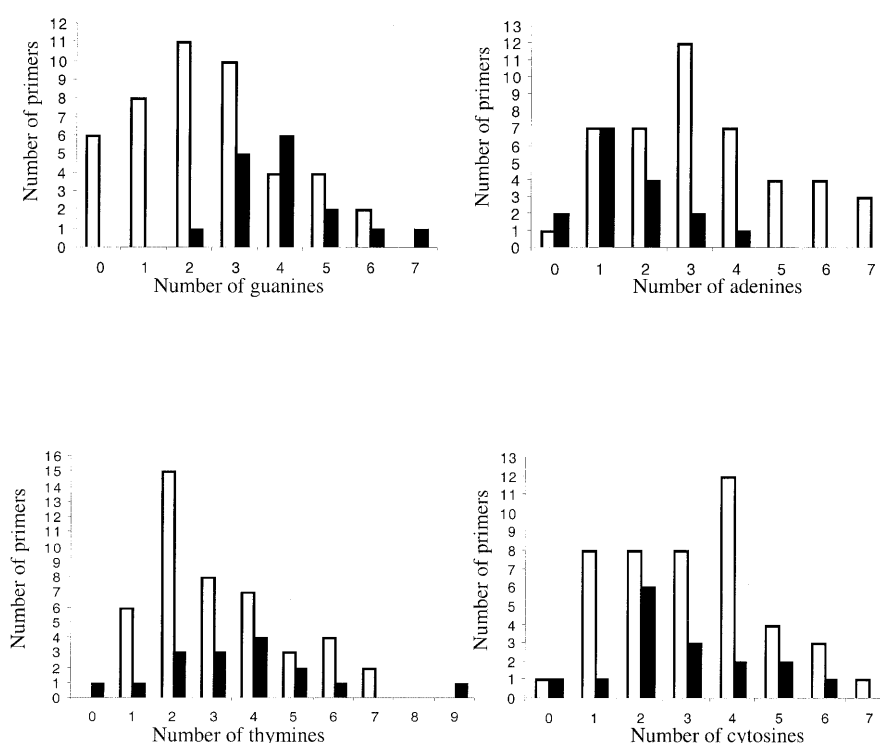


Figure 5. Working and non-working primers classified by the number of guanine, adenine, thymine or cytosine occurrences in the first 12 incorporated bases. Data from 45 working and 16 non-working primers were analyzed and are plotted. White bars indicate working primers (do not require supplementation) and black bars indicate non-working primers (require supplementation).

evidenced by the graph, the most striking difference between working and non-working primers is the number of guanine residues among the nt in this narrow window. Almost all non-working primers contained at least three guanine residues in this 12 base window. Guanines represent >30% of the first 12 nucleotides incorporated by non-working octamers. For the working primers guanines constitute 20% of the total number of bases (Table 6). The increased representation of guanine for non-working primers supports our hypothesis.

Interestingly, the number of guanine bases is more likely to influence the outcome of octamer sequencing than the number of thymine bases (Figs 5 and 6). This observation correlates with the lower amount of TTP used in supplemented reactions; 5 μ M TTP versus 75 μ M dGTP. Our observations are reinforced by the clustering patterns of the two groups of octamers. Working primers tend to cluster towards lower numbers of guanine and thymine bases (Fig. 6A). Most of them are followed by no more than four thymines and three or fewer guanines in the 12 bases immediately 3' of the octamer-binding site (extended DNA strand). The exceptions to this are associated with a higher free energy of the template secondary structures (and are, therefore, less likely to form secondary structure), compared to non-working primers that contain the same number of a particular base type in the 12 base window (data not shown).

Unexpectedly, the number of adenines specified for incorporation into the growing DNA chain differs for working and non-working primers. For the 45 screened working primers adenines represent 28.3% of the total number of bases in the

12 base window. For the non-working primers this percent is ~13% (Table 6). Three or more adenine residues follow in two-thirds of the working primers, while for 13 of the 16 non-working primers this number is two or fewer (Figs 5 and 6). The higher number of adenines among the first 12 nt incorporated 3' of the working primers suggests that the template region where the octamer is located is less likely to be associated with the formation of stable secondary structure. There is also abundant data that adenine residues are incorporated preferentially and more efficiently by different polymerases (the 'A rule') (13,14). This suggests that a larger number of adenines specified for incorporation in the vicinity of the primer terminus might facilitate extension from the octamer, or any primer, during DNA polymerization.

Based on these observations, we conclude that base composition, either alone or in combination with other factors, affects the outcome of a sequencing reaction. The reasons why five non-working primers did not benefit from dNTP supplementation are not known. However, these non-working primers suggest that, in addition to base composition 3' of the octamer and free energy of the template secondary structures, there are additional parameters that contribute to the success or failure of the extension reaction.

CONCLUSIONS

We have shown that dGTP/TTP supplementation of octamer-primed sequencing reactions increases the sequencing success rate and improves the quality of the sequence information. In

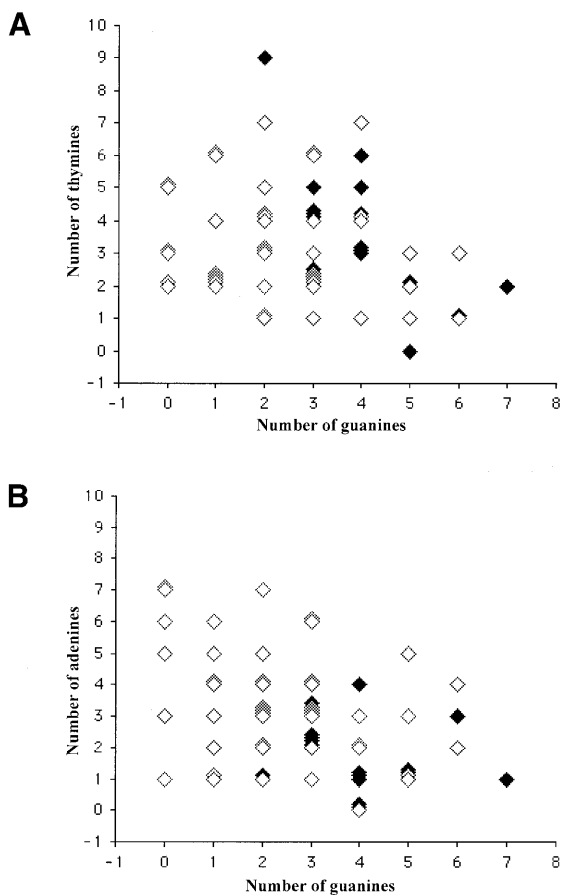


Figure 6. Effect of base composition on the success of octamer sequencing. Numbers of guanine, adenine and thymine residues in the first 12 incorporated bases are plotted for each template region. Working primers that were not tested in the dGTP/TTP supplementation experiments are also shown. White diamonds correspond to working primers and black diamonds indicate non-working primers. The numbers of thymine and guanine bases (A) or adenine and guanine bases (B) in this 12 base region are shown.

our experiments 61 primers were used to sequence seven different templates. Without supplementation 16 of these primers (~26%) failed to produce high quality fluorescent sequence information. However, after supplementation 11 of these initially non-working octamers primed sequencing reactions and nine of these generated long and accurate sequence reads. dNTP supplementation increased the number of successful octamer-primed reactions from 45 working primers (~74% success) to 54 working primers (~89% success). However, because of the negative effect of nucleotide supplementation on 38% of the working primers, we recommend that supplementation be used to produce or improve sequence data from an identified non-working primer, rather than routinely added into every sequencing reaction.

Importantly, template characteristics are an indicator of whether the reaction should be supplemented with the natural nucleotide cocktail. If three or more guanine and four or more thymine bases are specified for incorporation immediately 3' of the primer, supplementation with 75 μ M dGTP and 5 μ M dTTP may lead to an increase in sequence information quality. Furthermore, the base composition 3' of the octamer influences reaction success. Thus, in addition to template secondary structure analysis (2,4), the occurrence frequency of different nucleotide residues in this region should be used to identify the optimal sequencing primer. The data reported in this paper will be incorporated into the eOST program to further enhance the success rate of OST.

ACKNOWLEDGEMENTS

The authors are grateful to Hardin laboratory members for helpful discussions. This work was supported by the NIH (grant R29 HG01151 to S.H.H.).

REFERENCES

- Hardin, S.H., Jones, L.B., Homayouni, R. and McCollum, J.C. (1996) Octamer-primed cycle sequencing: design of an optimized primer library. *Genome Res.*, **6**, 545–550.
- Jones, L.B. and Hardin, S.H. (1998) Octamer-primed cycle sequencing using dye-terminator chemistry. *Nucleic Acids Res.*, **26**, 2824–2826.
- Jones, L.B. and Hardin, S.H. (1998) Octamer sequencing technology: optimization using fluorescent chemistry. *ABRF News*, **9**, 6–10.
- Mei, G. and Hardin, S.H. (2000) Octamer-primed sequencing technology: development of primer identification software. *Nucleic Acids Res.*, **28**, e22.
- Ball, S., Reeve, M.A., Robinson, P.S., Hill, F., Brown, D.M. and Loakes, D. (1998) The use of tailed octamer primers for cycle sequencing. *Nucleic Acids Res.*, **26**, 5225–5227.
- Rosenblum, B.B., Lee, L.G., Spurgeon, S.L., Khan, S.H., Menchen, S.M., Heiner, C.R. and Chen, S.M. (1997) New dye-labeled terminators for improved DNA sequencing patterns. *Nucleic Acids Res.*, **25**, 4500–4504.
- Kumari, D. and Usdin, K. (1999) Sequencing errors in reactions using labeled terminators. *Biotechniques*, **27**, 648–650.
- Applied Biosystems (1998) *Automated DNA Sequencing*, Ch. 2, *ABI PRISM DNA Sequencing Chemistries*. Applied Biosystems Foster City, CA, pp. 2–6.
- Myrnes, B., Guddal, P.H. and Krokan, H. (1982) Metabolism of dITP in HeLa cell extracts, incorporation into DNA by isolated nuclei and release of hypoxanthine from DNA by a hypoxanthine-DNA glycosylase activity. *Nucleic Acids Res.*, **10**, 3693–3701.
- Mosbaugh, D.W. (1988) Purification and characterization of porcine liver DNA polymerase γ : utilization of dUTP and dTTP during *in vitro* DNA synthesis. *Nucleic Acids Res.*, **16**, 5645–5659.
- Martin, F.H., Castro, M.M., Aboul-ela, F. and Tinoco, I., Jr (1985) Base pairing involving deoxyinosine: implications for probe design. *Nucleic Acids Res.*, **13**, 8927–8938.
- Li, Y., Zon, G. and Wilson, W.D. (1991) NMR and molecular modeling evidence for a G-A mismatch base pair in a purine-rich DNA duplex. *Proc. Natl Acad. Sci. USA*, **88**, 26–30.
- Clark, J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.*, **16**, 9677–9686.
- Paz-Elizur, T., Takeshita, M. and Livneh, Z. (1997) Mechanism of bypass synthesis through an abasic site analog by DNA polymerase I. *Biochemistry*, **36**, 1766–1773.