 

# Improved fidelity of thermostable ligases for detection of microsatellite repeat sequences using nucleoside analogs

**Monib Zirvi, Donald E. Bergstrom[1], Andrea S. Saurage[2], Robert P. Hammer[2] and Francis Barany***

Department of Microbiology, Box 62, Hearst Microbiology Research Center, Strang Cancer Prevention Center, Joan and Sanford I. Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10021, USA, [1]Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 46208, USA and [2]Department of Chemistry, Louisiana State University, Baton Rouge, LA 70803, USA

Received July 7, 1999; Revised October 8, 1999; Accepted October 17, 1999

## ABSTRACT

**Microsatellite repeats consisting of dinucleotide sequences are ubiquitous in the human genome and have proven useful for linkage analysis, positional cloning and forensic identification purposes. In this study, the potential of utilizing the ligase detection reaction for the analysis of such microsatellite repeat sequences was investigated. Initially, the fidelity of thermostable DNA ligases was measured for model dinucleotide repeat sequences. Subsequently, the effect of modified oligonucleotides on ligation fidelity for dinucleotide repeats was determined using the nucleoside analogs nitroimidazole, inosine, 7-deazaguanosine and 2-pyrimidinone, as well as natural base mismatches. The measured error rates for a standard dinucleotide template indicated that the nitroimidazole nucleoside analogs could be used to increase the fidelity of ligation when compared to unmodified primers. Furthermore, use of formamide in the ligation buffer also increased ligation fidelity for dinucleotide repeat sequences. Using ligation-based assays to detect polymorphic alleles of microsatellite repeats in the human genome opens the possibility of using array-based typing of these loci for human identification, loss-of-heterozygosity studies and linkage analysis.**

## INTRODUCTION

Microsatellites are polymorphic repetitive sequences which occur throughout the human genome. Analysis of these repetitive sequences has proved very useful in human genome studies due to their highly polymorphic nature and the ability to determine chromosomal segregation patterns in human pedigrees. Association studies which identify given microsatellite loci with the transmission of a disease phenotype have been used extensively in the last decade for the positional cloning of a number of genes responsible for inherited human disorders such as cystic fibrosis (CFTR), Duchenne muscular dystrophy, Wilms' tumor (WT1) and familial adenomatous polyposis (1–5). In general, microsatellite repeats are analyzed by PCR amplification followed by electrophoretic separation on gels (6,7); however, detection of variations occurring within microsatellite repeats and closely clustered mutations is complicated by artifacts that occur during PCR.

Our laboratory has developed a combined polymerase chain reaction/ligase detection reaction (PCR/LDR) method for discriminating single-base mutations, as well as mononucleotide repeat polymorphisms (8–12). In this study, detection of alterations in dinucleotide repeat sequences was carried out using ligation mediated by thermostable ligases. The effect of using modified oligonucleotide primers in ligation for this standard template was also determined for various nucleoside analogs. The analogs tested included nitroimidazole, inosine, 7-deazaguanosine and 2-pyrimidinone, as well as natural base mismatches. Many of these nucleoside analogs have the potential to modulate ligation fidelity through a combination of effects on melting temperature and enzyme–template interactions.

## MATERIALS AND METHODS

### Modified nucleoside analogs

The nitroimidazole nucleoside phosphoramidite was synthesized according to previously published procedures (13). The pyrimidin-2-one nucleoside has been prepared previously by a variety of methods including coupling of mercury salt of pyrimidin-2-one to 2-deoxy chloro sugar (14) and deamination of 2′-deoxycytosine (15). We have used a typical silylation/nucleosidation procedure of 3,5-bis-*O*-*p*-toluoyl-1-chloro-β-D-ribose to provide the β-anomer in 40% yield. Standard deprotection, tritylation and phosphitylation protocols were used to convert this to the nucleoside phosphoramidite which could then be incorporated into oligonucleotides (see experimental details below). The inosine, 7-deazaguanosine, dP and 5-fluoro-deoxyuridine phosphoramidites were obtained from Glen Research (Sterling, VA). The various nucleoside analogs

*To whom correspondence should be addressed. Tel: +1 212 746 6509; Fax: +1 212 746 8587; Email: barany@mail.med.cornell.edu

**Table 1.** Synthetic templates and nucleoside analog modified LDR primers

| | |
|---|---|
| | **D6S291 Synthetic Templates** |
| CA11 Template | 5' TGG TAA GTG ACT CTG TCT CTC TGT ACA CAC ACA CAC ACA CAC ACA CCC TGT GAA ACC ATC ACC AC 3' |
| TG11 Template | 5' GTG GTG ATG GTT TCA CAG GGT GTG TGT GTG TGT GTG TGT GTA CAG AGA GAC AGA GTC ACT TAC CA 3' |
| CA12 Template | 5' TGG TAA GTG ACT CTG TCT CTC TGT ACA CAC ACA CAC ACA CAC ACA CAC CCT GTG AAA CCA TCA CCA C 3' |
| TG12 Template | 5' GTG GTG ATG GTT TCA CAG GGT GTG TGT GTG TGT GTG TGT GTG TAC AGA GAG ACA GAG TCA CTT ACC A 3' |
| CA13 Template | 5' TGG TAA GTG ACT CTG TCT CTC TGT ACA CAC ACA CAC ACA CAC ACA CAC ACC CTG TGA AAC CAT CAC CAC 3' |
| TG13 Template | 5' GTG GTG ATG GTT TCA CAG GGT GTG TGT GTG TGT GTG TGT GTG TGT ACA GAG AGA CAG AGT CAC TTA CCA 3' |
| | **Standard LDR primers** |
| TG6 Common Primer | 5' p GTG TGT GTG TGT ACA GAG AGA CAG AGT C -Fl 3' |
| TG5 Primer | 5' CAA AAA TGG TTT CAC AGG GTG TGT GTG T 3' |
| TG6 Primer | 5' CAA AAT AGG TTT CAC AGG GTG TGT GTG TGT 3' |
| TG7 Primer | 5' CAA AAT AAG TTT CAC AGG GTG TGT GTG TGT GT 3' |
| **I=Inosine** | **Inosine LDR primers** |
| TG6 Common Primer | 5' p GTI TGT ITG TIT ACA GAG AGA CAG AGT CAC TTA CC -Fl 3' |
| TG5 Primer | 5' AAA GTG ATG GTT TCA CAG GGT ITG TIT GT 3' |
| TG6 Primer | 5' AAA ATG ATG GTT TCA CAG GIT GTI TGT ITG T 3' |
| TG7 Primer | 5' AAA AAG ATG GTT TCA CAG GGT ITG TIT GTI TGT 3' |
| **X=7-Deazaguanosine** | **7-Deazaguanosine LDR primers** |
| TG6 Common Primer | 5' p GTX TGT XTG TXT ACA GAG AGA CAG AGT CAC TTA CC -Fl 3' |
| TG5 Primer | 5' GTG ATG GTT TCA CAG GGT XTG TXT GT 3' |
| TG6 Primer | 5' ATG ATG GTT TCA CAG GXT GTX TGT XTG T 3' |
| TG7 Primer | 5' AAG ATG GTT TCA CAG GGT XTG TXT GTX TGT 3' |
| **Z=2-Pyrimidinone** | **2-Pyrimidinone LDR primers** |
| CA6 Common Primer | 5' p ACA ZAC AZA CAZ CCT GTG AAA CCA TCA C -Fl 3' |
| CA5 Primer | 5' GTA AGT GAC TCT GTC TCT CTG TAC AZA CAZ AC 3' |
| CA6 Primer | 5' ATA AGT GAC TCT GTC TCT CTG TAZ ACA ZAC AZA C 3' |
| CA7 Primer | 5' AAT TGT GAC TCT GTC TCT CTG TAC AZA CAZ ACA ZAC 3' |
| **Q=Nitrimidazole** | **Nitroimidazole LDR primers** |
| CA6 Common Primer | 5' p ACA QAC AQA CAC CCT GTG AAA CCA TCA C -Fl 3' |
| CA5 Primer | 5' GTA AGT GAC TCT GTC TCT CTG TAQ ACA QAC AC 3' |
| CA6 Primer | 5' ATA AGT GAC TCT GTC TCT CTG TAC AQA CAQ ACA C 3' |
| CA7 Primer | 5' AAT TGT GAC TCT GTC TCT CTG TAC ACA QAC AQA CAC 3' |
| **T (Bold)=Mismatch** | **Mismatch LDR primers** |
| TG6 Common Primer | 5' p ACA **T**AC A**T**A CAC CCT GTG AAA CCA TCA C -Fl 3' |
| TG5 Primer | 5' GTA AGT GAC TCT GTC TCT CTG TA**T** ACA **T**AC AC 3' |
| TG6 Primer | 5' ATA AGT GAC TCT GTC TCT CTG TAC A**T**A CA**T** ACA C 3' |
| TG7 Primer | 5' AAT TGT GAC TCT GTC TCT CTG TAC ACA **T**AC A**T**A CAC 3' |

were then incorporated into ligation primers using standard oligonucleotide synthesis conditions. The sequences of the various primers used are shown in Table 1.

### 1-(2′-Deoxy-3′,5′-bis-*O*-*p*-toluoyl-β-D-ribofuranosyl)-pyrimidin-2-one

Hexamethyldisilazane (6.9 ml, 31.4 mmol) was added to a flask containing 2-pyrimidinone hydrochloride (0.93 g, 7.0 mmol) with a few crystals of ammonium sulfate. The reaction mixture was heated until reflux for 1 h resulting in a yellowish solution. Once cool, the excess solvent was removed under diminished pressure. Immediately following the removal of the solvent, freshly distilled chloroform was added under argon, followed by addition of 3,5-bis-*O*-*p*-toluoyl-1-chloro-β-D-ribose (16) (1.40 g, 3.6 mmol). After stirring overnight, the reaction mixture was washed with saturated $NaHCO_3$, $H_2O$ and saturated NaCl, dried with $Na_2SO_4$, filtered and the solvent removed under reduced pressure yielding a creamy white solid. This solid was redissolved in chloroform and crystallized overnight using 3:1 hexanes:EtOAc at reduced temperature. The resulting crystals (>95% β-anomer) were filtered and rinsed with cool 3:1 hexanes:EtOAc yielding a white solid (0.65 g, 40%). $^1$H (200 MHz, $CDCl_3$), 8.65 (t, 1H, H4), 8.05 (dd, 1H, H6), 7.96 (d, 4H, Ar), 7.61 (d, 4H, Ar,), 6.32 (m, 2H, H5, H1′), 5.58 (d, 1H, H3′), 4.93 (t, 1H, H4″), 4.59 (d, 2H, H5′, H5″), 3.10–2.99 (m, 1H, H2′), 2.77 (d, 1H, H2′), 2.42 (s, 3H, $CH_3$), 2.41 (s, 3H, $CH_3$). $^{13}$C (50 MHz, $CDCl_3$) δ 166.17 [**C**(O)], 165.79 (C4), 165.42 [**C**(O)], 155.23 (C2), 144.26 [C(O)-**Ar**], 143.82 [C(O)-**Ar**], 142.8 (C6), 129.66 (Ar), 129.11 (Ar), 126.83 (**Ar**-$CH_3$), 126.55 (**Ar**-$CH_3$), 103.4 (C5), 89.84 (C1′), 82.84 (C4′), 75.06 (C3′), 64.03 (C5′), 37.41 (C2′), 21.66 ($CH_3$).

### 1-(2′-Deoxy-β-D-ribofuranosyl)-pyrimidin-2-one

1-(2′-Deoxy-3′, 5′-bis-*O*-*p*-toluoyl-β-D-ribofuranosyl)-pyrimidin-2-one (0.686 g, 1.49 mmol) was deprotected using ammonia-saturated methanol (100 ml) overnight. The solvent was removed under diminished pressure. A solution of 9:1 chloroform:methanol was added to the dried reaction mixture to precipitate a white solid which was in 43% yield (0.136 g, 0.641 mmol). $^1$H (200 MHz, $CD_3OD$) δ 8.52 (t, 1H, H4), 8.22 (dd, 1H, H6), 6.47 (t, 1H, H5), 5.97 (d, 1H, H1′), 4.19 (t, 1H, H3′), 4.10 (s, 1H, H4′), 3.36 (d, 2H, H5′, H5″), 2.62–2.51 (m, 1H, H2′), 1.97 (d, J = 11.2, 1H, H2′).

### 1-(2′-Deoxy-5′-dimethoxytrityl-β-D-ribofuranosyl)-pyrimidin-2-one

1-(2′-Deoxy-β-D-ribofuranosyl)-pyrimidin-2-one (0.130g, 0.64 mmol) was co-evaporated with two portions of pyridine then dissolved in pyridine (4 ml). Next, addition of *N,N*-dimethylaminopyridine (0.11g, 0.992 mmol) was followed by diisopropylethylamine (0.8 ml), and dimethoxytrityl chloride (1.05 g, 3.09 mmol) and dimethylaminopyridine (1 mg). The reaction was left stirring overnight under inert atmosphere. Purification of the crude product by flash chromatography yielded 0.136 g (39%) of yellowish oil. $^1$H (250 MHz, $CDCl_3$) δ 8.52 (t, 1H, H4), 8.20 (dd, 1H, H6), 7.42–7.22 (m. 8H, **Ar**-$OCH_3$, **Ar**-$OCH_3$), 6.82 (m, 5H, Ar), 6.30 (t, 1H, H5), 6.22 (d, 1H, H1′), 4.53 (t, 1H, H3′), 4.47 (d, 1H, H4′), 3.78 (s, 6H, Ar-OC**H**$_3$), 3.18–3.13 (m, 2H, H5′, H5″), 2.25–2.18 (m, 1H, H2′), 1.94 (d, 1H, H2′).

### 1-(2′-Deoxy-5′-dimethoxytrityl-3′-*O*-2-cyanoethyl-*N,N*-diisopropylphosphoramidite-β-D-ribofuranosyl)-pyrimidin-2-one

1-(2′-Deoxy-5′-dimethoxytrityl-β-D-ribofuranosyl)-pyrimidin-2-one (0.15 g, 0.309 mmol) was twice co-evaporated with pyridine then dissolved in $CH_2Cl_2$ (3 ml) under argon. Next, diisopropylchlorocyanoethylphosphoramidite (100 µl, 0.61 mmol) was added slowly and the solution was stirred for 1 h. The crude reaction was quenched with methanol and then washed with brine (5.4 M NaCl), dried with $Na_2SO_4$ and concentrated under reduced pressure. Purification by flash chromatography (1:1:1 hexanes:EtOAc:$Et_3N$) provided a yellow foam (38 mg, 16%). $^{31}$P (101.2 MHz, $CDCl_3$) δ 151.20, 150.18 mixture of diastereomers.

### Oligonucleotide synthesis and purification

Oligonucleotides were synthesized on an ABI 394 DNA Synthesizer. (PE Biosystems Inc., Foster City, CA). Oligonucleotides used in LDR were purified by electrophoresis on 10% polyacrylamide/7 M urea gels. Bands were visualized by UV shadowing and excised from the gel. They were eluted overnight at 64°C in TNE buffer (100 mM Tris–HCl pH 8.0, 500 mM NaCl, 5 mM EDTA). The oligonucleotides were recovered from the eluate using C18 Sep-Pak cartridges (Waters Corp., Milford, MA) following the manufacturer's instructions. Oligonucleotides were resuspended to ~1 mM in 100 µl TE (10 mM Tris pH 8.0, 1 mM EDTA). For LDR, gel-purified stock solutions were diluted to 100 µM (100 pmol/µl).

The upstream, or discriminating, oligonucleotides had fluorescent reporter groups at their 5′ termini. The fluorescent oligonucleotides were synthesized using Fam and Tet phosphoramidites (PE Biosystems Inc., Foster City, CA). The downstream, or common, oligonucleotides were phosphorylated at the 5′ end using a chemical phosphorylation reagent, and blocked at the 3′ end using a 3′-spacer C3 CPG (Glen Research, Sterling, VA). The use of chemical phosphorylation reagent is an alternative to the enzymatic techniques for oligonucleotide phosphorylation, with the advantage of allowing phosphorylation efficiency to be determined.

### LDR conditions

LDR reactions were carried out in a 20 µl mixture containing 20 mM Tris–HCl pH 7.6, 10 mM $MgCl_2$, 100 mM KCl, 10 mM DTT, 1 mM $NAD^+$, 25 nM (500 fmol) of the detecting primers and mixtures of PCR products from cell lines or tumor samples. The reaction mixture was heated for 1.5 min at 94°C prior to adding 25 fmol of the wild-type or mutant *Tth* DNA ligase. Ligases were overproduced and purified as described previously (8,17). LDR reactions were thermally cycled for 20 cycles of 15 s at 94°C and 2 min at 65°C. Reactions were stopped by chilling the tubes in an ethanol–dry ice bath, and adding 0.5 µl of 0.5 mM EDTA. Aliquots of 2.5 µl of the reaction products were mixed with 2.5 µl of loading buffer (83% formamide, 8.3 mM EDTA and 0.17% Blue Dextran). The mixture was supplemented with 0.5 µl Rox-1000, or TAMRA 350 molecular weight marker, denatured at 94°C for 2 min, chilled rapidly on ice prior to loading on an 8 M urea–10% polyacrylamide gel, and electrophoresed on an ABI 373 DNA sequencer at 1400 V for 2 h. Fluorescent ligation products were analyzed and quantified using the ABI Gene Scan 672 software. The amount of product is calculated from a calibration curve (one fmol = 600 peak area units).

## RESULTS AND DISCUSSION

### Detection of dinucleotide repeat polymorphisms using thermostable ligase

The initial model microsatellite repeat chosen for this study is the polymorphic D6S291 sequence for which the three most frequent alleles contain CA11, CA12 and CA13 dinucleotide repeats. LDR primers were designed and synthesized for this sequence containing repeats with 11, 12 or 13 CA dinucleotides (Fig. 1, top). These primers were then ligated using synthetic templates and LDR products were analyzed on an ABI 373 DNA sequencer. The results (Fig. 1, bottom) demonstrate that ligation yielded predominantly the correct size product, but misligations on incorrect templates were up to 25% of the correct product signal. When using primers to distinguish CA18 from CA19 synthetic templates, the amount of product formed for both templates was equal regardless of the LDR primers used (data not shown). This indicates that the thermostable ligase loses its fidelity for discriminating dinucleotide slippage in repeats >36 nt in length. Table 2 summarizes the amount of misligation found when using *Tth* DNA ligase to detect various length mono- and dinucleotide microsatellite repeats.

**Table 2.** Summary of misligation error rate and ligation fidelity of *Tth* ligase for various length microsatellite repeats

| Type of mutation | Repeat length | Ligation error rate (%) |
|---|---|---|
| Point mutation | 1 | 0.1–0.2 |
| A10→A9 | 10 | 1 |
| A16→A15 | 16 | 4 |
| A19→A18 | 19 | 10 |
| CA13→CA12 | 26 | 25 |
| CA19→CA18 | 38 | 100 |

### Use of nucleoside analogs to improve discrimination of dinucleotide repeat sequences

A variety of primers containing nucleoside analogs were tested to determine if such modifications could help increase the

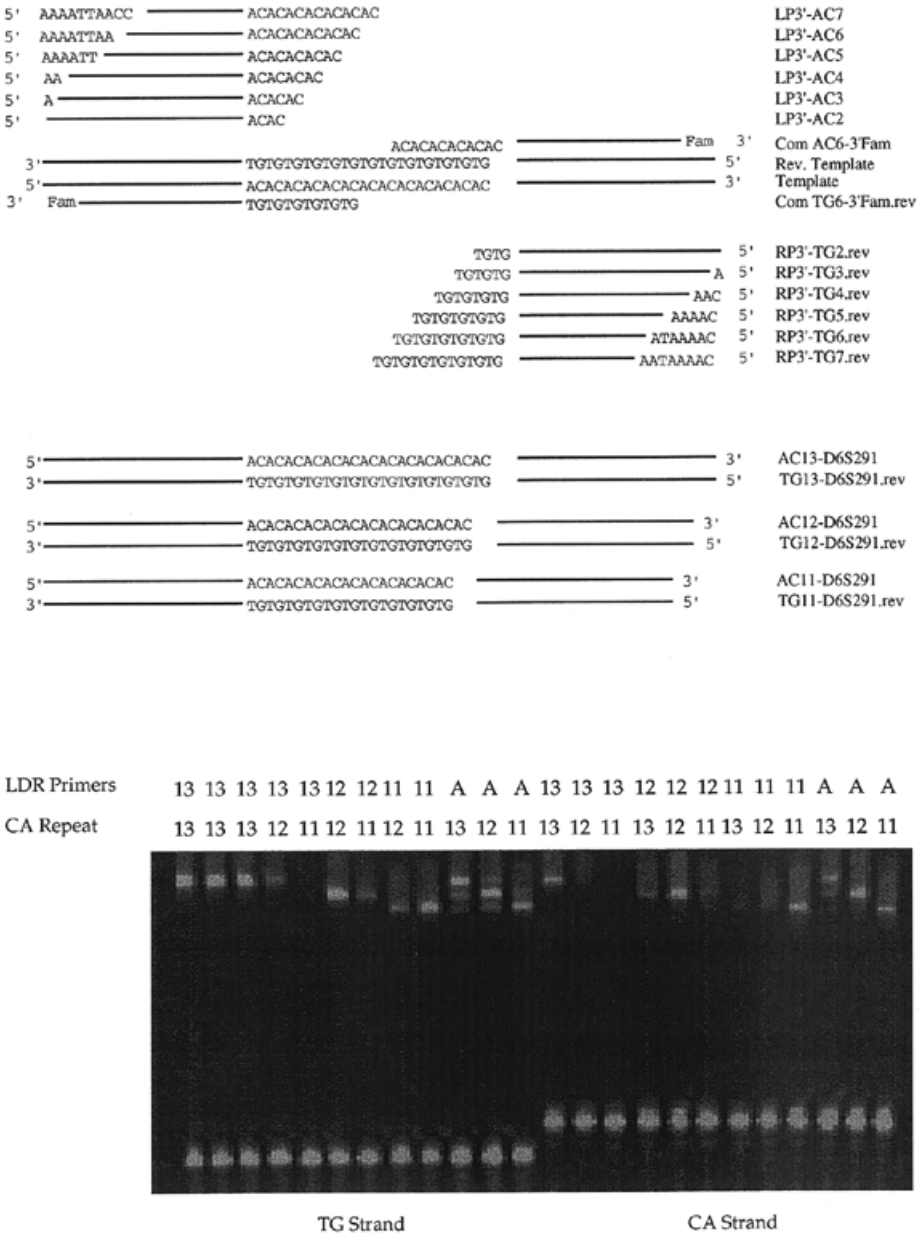


**Figure 1.** LDR results for dinucleotide repeat sequences. (Top) Primers designed to detect CA11, CA12 and CA13 templates. The ligase reaction was carried out using 500 fmol of each primer and 50 fmol of template DNA in a 20 µl reaction with 25 fmol of *Tth* ligase per reaction. After ligation, the primers were separated and analyzed on an ABI 373 automated DNA sequencer. (Bottom) Ligation results for the detection of synthetic CA repeat templates using matched and mismatched LDR primer pairs. The 'A' columns represent lanes in which a multiplex ligation reaction was carried out using primers that could detect CA13, CA12 and CA11. The crosstalk from misligation of LDR primers, which differed by one CA unit, can be seen as off-diagonal bands.

specificity of ligation. The LDR primers synthesized had from five to seven dinucleotide repeat sequences and ligation fidelity was compared to the standardized assay on CA11–CA13 templates. The various analogs tested and their special properties are listed in Table 3. Structures of the various nucleoside analogs compared with the natural bases are shown in Figure 2.

Use of nucleoside analogs such as nitroimidazole (13) and pyrimidin-2-one in place of cytidine had previously been shown to lower the overall $T_m$ contribution of a G:C base pair by ~10°C/modification while at the same time maintaining the specificity of hybridization. Analogs of guanosine that have been shown to decrease duplex stability include inosine and 7-deazaguanosine (18). However, these nucleoside analogs are also capable of mispairing as a result of reduced specificity during hybridization. Preliminary studies have shown that the fidelity of ligation can be improved by incorporating nucleoside analogs such as 3-nitropyrrole two bases upstream of the ligation junction (19).

In general, ligation of primers hybridized to dinucleotide microsatellite repeat sequences showed greater slippage than

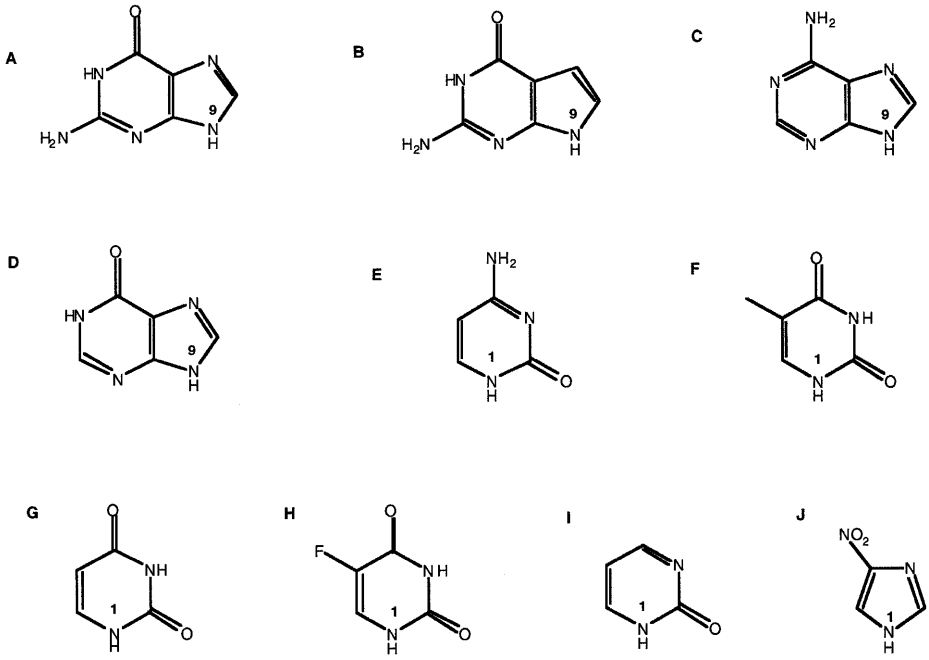


**Figure 2.** Structure of nucleotide bases tested in LDR primers for ligation efficiency in dinucleotide repeats. The nucleotide bases include: A, guanine; B, 7-deazaguanine; C, adenine; D, hypoxanthine; E, cytosine; F, thymine; G, uracil; H, 5-fluorouracil; I, pyrimidin-2-one; J, 4-nitroimidazole. The bases are linked to C1′ of deoxyribose through the numbered nitrogen atom.

**Table 3.** Table of nucleoside analogs incorporated in LDR primers to test for reduction of mismatch ligation on dinucleotide repeat targets

| |
|---|
| **Guanine Analogues** |
| **Hypoxanthine**: replacing guanine on TG strand<br>• (guanine analogue which forms only two hydrogen bonds with cytosine) |
| **7-deazaguanine**: replacing guanine on TG strand<br>• (guanine analogue which forms weaker hydrogen bonds with cytosine) |
| **Cytosine Analogues** |
| **Nitroimidazole**: replacing cytosine on CA strand<br>• (base pairs with guanine but destabilizes hybridization) |
| **Pyrimidin-2-one**: replacing cytosine on CA strand<br>• (cytosine analogue which forms only two hydrogen bonds with guanine) |
| **Thymine**: replacing cytosine on CA strand<br>• (cytosine "analogue" which forms mismatch with guanine) |
| **5-Fluorouracil**: replacing cytosine on CA strand<br>• (base pairs with guanine slightly better than thymine) |

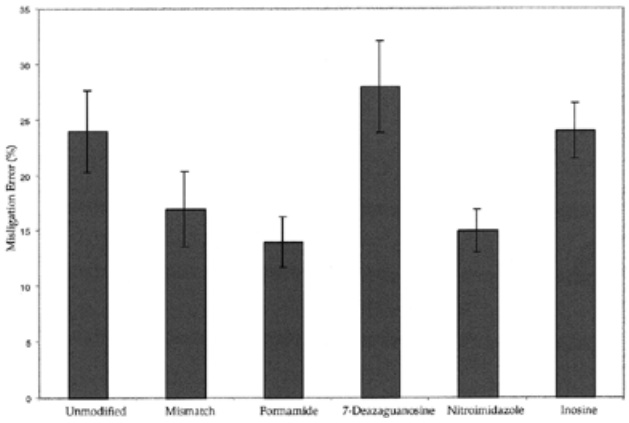


**Figure 3.** Graph of average misligation error rate for dinucleotide repeats using modified oligonucleotides. Ligation reaction conditions were modified by using primers containing nucleoside analogs, mismatches or the addition of formamide to the ligation buffer. The *y*-axis in the figure indicates the average percentage of misligation error for CA11, CA12 and CA13 templates generated using multiple trials with each modification. The bars indicate which modifications were introduced at alternating dinucleotide repeats in the standard LDR primers (see Table 1 for sequences).

mononucleotide repeats (12). Analysis using modified primers indicated that misligation could be reduced on average to 15% of the correct signal (Fig. 3). Modifications which improved fidelity included increasing the ligation temperature, adding formamide to the ligation buffer, or introducing deliberate mismatches or the nitroimidazole nucleoside analog into the LDR primers. Altering location or spacing of the nitroimidazole analogs in the primers did not significantly change ligation fidelity compared to the set of nitroimidazole-containing primers shown in Table 1 (data not shown). The strongest signals, however, were obtained using primers without analogs. *Tth* DNA ligase could maintain high fidelity in a solution

containing >10% formamide and, furthermore, ligation fidelity increased under these conditions (Fig. 3).

The 7-deazaguanosine and inosine containing primers did not improve the ligation fidelity for the different repeats tested (Fig. 3). This result may be due to mispairing of these nucleoside analogs with adjacent bases on the opposite strand (18). Primers containing the nitroimidazole analog did improve the ligation fidelity for the microsatellite repeats tested. Finally, polyacrylamide gel electrophoresis demonstrated that the 2-pyrimidinone-containing primers were cleaved instead of ligating during thermocycling (data not shown). The instabililty of pyrimidin-2-one nucleosides to acidic conditions has been previously noted (20). Given the high $\Delta pK_a$ of Tris buffer with temperature ($\Delta pK_a = -0.033/°C$) (21,22), the LDR reactions at 95°C are approximately pH 5, which is sufficient to remove the pyrimidin-2-one nucleobase leading to abasic sites and strand scission under the LDR conditions. Future studies will focus on pyrimidin-2-one analogs with electron-withdrawing groups to stabilize them to acidic hydrolysis. Overall, ligation fidelity could be increased by the introduction of nitroimidazole within the repeat region or the addition of mismatches near the ligation junction. Other conditions either diminished the total signal from ligation or increased the misligation error compared to unmodified ligation primers. Using combinations of formamide and the nitroimidazole-containing LDR primers, the misligation error was reduced to 10% of the correct signal for repeats up to 13 dinucleotides in length (data not shown).

In an attempt to accentuate the differences in melting temperature between the CA repeat regions and the flanking unique sequences, primers were synthesized with long tails or propynyl groups in the flanking sequence. Even when such ligations were performed at higher temperatures and/or in the presence of formamide, these modifications decreased ligation fidelity compared with the standard conditions (data not shown). This result suggests that it is not the relative $T_m$ values which predict ligation fidelity, but the substrate structure near the ligation junction as well as the overall length of the repeat sequences.

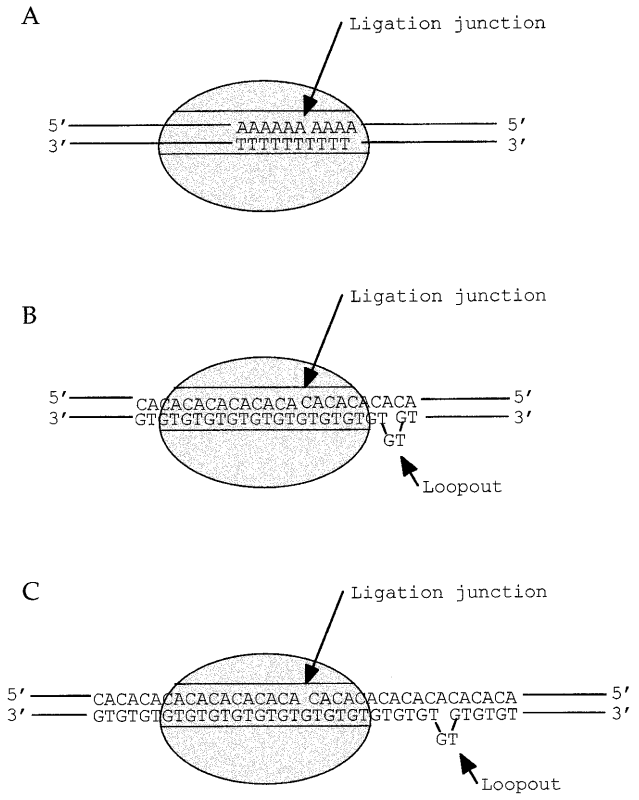


**Figure 4.** Schematic diagram of theoretical model for the fidelity of thermostable DNA ligase for microsatellite repeat sequences. (**A**) For short mononucleotide repeat sequences, the ligase footprint is large enough to span the entire repetitive region, i.e. one helical turn for an A10 repeat. (**B**) For intermediate length dinucleotide repeat sequences such as CA13, the LDR primers can sporadically loopout a single dinucleotide unit outside or at the border of the ligase footprint, therefore increasing the misligation rate. (**C**) For long dinucleotide repeat sequences (CA18 and CA19), the loopout of a single repeat unit occurs frequently and therefore eliminates specificity during ligation.

## CONCLUDING REMARKS

PCR/LDR has been used extensively in the past for the analysis of disease-associated point mutations in non-repetitive DNA sequences (1,8–11). However, the capabilities of PCR/LDR for analysis of microsatellite repeat sequences were not known. In this study, the limits of using ligation for the analysis of a dinucleotide repeat were determined for the *Tth* DNA ligase. For dinucleotide repeats, the amount of noise due to misligation increased by >20-fold when compared to the results obtained for mononucleotide repeats (compare A10→A9 to CA13→CA12 in Table 2). This result suggests that the ligase footprint is close to the span of 26 nt of a CA13 repeat and certainly larger that the span of 10 adenosine nucleotides. The *Tth* DNA ligase most likely stabilizes the DNA duplex directly below its footprint, but may allow nucleotides to loopout of the double helix outside its footprint (Fig. 4). This model correlates with the observed increase in misligation errors observed above CA13 repeat length. Ligation of longer dinucleotide repeats, such as CA19 and CA18, lacked specificity altogether, and produced equal amounts of ligation product on each template. There is experimental evidence that *Tth* DNA ligase can also loopout a base in non-repetitive sequences at low efficiency (23).

By using modified nucleosides within the repeat region, the misligation errors were reduced to <15% of signal obtained for the correct product. The mechanism by which a nucleoside analog can exert its effect on ligation is through the destabilization of the repeat region by lowering the $T_m$ of the repetitive portion of the LDR primer. Due to the alteration of the stability of the repetitive tract, the ligase has more difficulty in aligning the primers at the nick. This reduces the amount of product generated from misligations when compared to primers that have stable repetitive regions. Similarly, deliberate mismatches near the ligation junction makes it more difficult for the ligase to form incorrect products resulting from misligation during LDR.

An analysis of the 5700 microsatellites in the CEPH database (http://www.cephb.fr/bio/ceph-genethon-map.html ) was performed and it was determined that the majority of the CA repeats contained in these mapped amplicons were greater than 20 repeat units in length. However, a group of 40 loci with heterozygosities of ranging from 0.60 to 0.80 have been identified with the longest allele consisting of CA15. Furthermore,

over 1000 dinucleotide repeat sequences with at least one allele less than CA15 were found in sequences from the Whitehead Institute sequence tagged sites database. (http://www.genome.wi.mit.edu/ ) Such dinucleotide repeats would be amenable for analysis using ligase-based detection.

The average single-nucleotide polymorphism (SNP), has a typical heterozygosity ranging between 0.40 and 0.50. To yield the analytical power of 1000 dinucleotide repeats for positional cloning studies, analysis of at least 1750 SNPs would be required. The CEPH database contains an inherent bias towards dinucleotide repeats which are greater than CA20 in length. This is a result of the screening process which utilized long CA probes and hybridization to genomic libraries to identify highly polymorphic microsatellite loci. However, as the human genome sequence is completed, many shorter dinucleotide repeat loci will be found with heterozygosities which are significantly greater than those obtained using SNPs for genetic analysis. Therefore, in the future, it is anticipated that there will be a number of loci that could be analyzed by the approach used in this study. There is also evidence that polymorphic mononucleotide repeats are commonly found in plant and mammalian genomes (24–28). These regions could also be analyzed using LDR (12). Notable microsatellite repeats which are not amenable to ligase-based detection are the extremely long triplet repeat stretches related to genes responsible for Fragile X syndrome, myotonic dystrophy and Huntington's chorea (29).

An important advantage of the use of ligase-based methods for the analysis of mutations and short repeat polymorphisms is the ability to eliminate the need for gel electrophoresis. By using a DNA array, it is possible to identify multiplex LDR products by hybridization instead of gel electrophoresis. In our laboratory, a prototype universal array has been developed that detects LDR products based on addressable sequence tags attached to the LDR primers (30). This array combined with PCR/LDR analysis was able to correctly identify mutations within cell lines known to carry specific mutations in the K-*ras* gene (11,30). In the future, such an array should be capable of typing both single nucleotide and microsatellite repeat sequence polymorphisms at multiple loci simultaneously. This would allow for more rapid methods of human identification, loss-of-heterozygosity detection and linkage analysis.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge Norman Gerry, Reyna Favis, Weiguo Cao, Jie Tong, Joseph Day, Andrew Grace, Jie Zhou and Matt D'Alessio for their helpful suggestions and technical assistance. M.Z. was supported by the William M. Keck Fellowship and the Tri-Institutional Medical Scientist Training Program (MSTP) Fellowship during the course of this study. Support for this work was provided by the National Cancer Institute (P01-CA65930 and RO1-CA81467).

## REFERENCES

1. Eggerding,F.A., Iovannisci,D.M., Brinson,E., Grossman,P. and Winn-Deen,E.S. (1995) *Hum. Mutat.*, **5**, 153–165.
2. Weber,J.L. (1990) *Curr. Opin. Biotechnol.*, **1**, 166–171.
3. Cawkwell,L., Lewis,F.A. and Quirke,P. (1994) *Br. J. Cancer*, **70**, 813–818.
4. Powell,S.M., Zilz,N., Beazer-Barclay,Y., Bryan,T.M., Hamilton,S.R., Thibodeau,S.N., Vogelstein,B. and Kinzler,K.W. (1992) *Nature*, **359**, 235–237.
5. Kinzler,K.W. and Vogelstein,B. (1996) *Cell*, **87**, 159–170.
6. Mansfield,D.C., Brown,A.F., Green,D.K., Carothers,A.D., Morris,S.W., Evans,H.J. and Wright,A.F. (1994) *Genomics*, **24**, 225–233.
7. Ziegle,J.S., Su,Y., Corcoran,K.P., Nie,L., Mayrand,P.E., Hoff,L.B., McBride,L.J., Kronick,M.N. and Diehl,S.R. (1992) *Genomics*, **14**, 1026–1031.
8. Barany,F. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 189–193.
9. Barany,F. (1991) *PCR Methods Appl.*, **1**, 5–16.
10. Day,D., Speiser,P.W., White,P.C. and Barany,F. (1995) *Genomics*, **29**, 152–162.
11. Khanna,M., Park,P., Zirvi,M., Paty,P. and Barany,F. (1999) *Oncogene*, **18**, 27–38.
12. Zirvi,M., Nakayama,T., Newman,G., McCaffrey,T., Ostrer,H., Paty,P. and Barany,F. (1999) *Nucleic Acids Res.*, **27**, e40.
13. Bergstrom,D.E., Zhang,P. and Johnson,W.T. (1997) *Nucleic Acids Res.*, **25**, 1935–1942.
14. Kohler,P. and Volz,E. (1978) *Nucleic Acid Chem.*, 283–289.
15. Gildea,B. and McLaughlin,L.W. (1989) *Nucleic Acids Res.*, **17**, 2261–2281.
16. Huttle,R. and Bucheke,F. (1955) *Chem. Ber.*, **88**, 1586–1590.
17. Barany,F. and Gelfand,D. (1991) *Gene*, **109**, 1–11.
18. Brown,D.M. and Lin,P.K. (1991) *Carbohydr. Res.*, **216**, 129–139.
19. Luo,J., Bergstrom,D.E. and Barany,F. (1996) *Nucleic Acids Res.*, **24**, 3071–3078.
20. Iocono,J.A., Gildea,B. and McLaughlin,L.W. (1990) *Tetrahedron Lett.*, **31**, 175–178.
21. Day,J., Bergstrom,D., Hammer,R. and Barany,F. (1999) *Nucleic Acids Res.*, **27**, 1810–1818.
22. Day,J., Hammer,R., Bergstrom,D. and Barany,F. (1999) *Nucleic Acids Res.*, **27**, 1819–1827.
23. Tong,J., Cao,W. and Barany,F. (1999) *Nucleic Acids Res.*, **27**, 788–794.
24. Aitman,T.J., Hearne,C.M., McAleer,M.A. and Todd,J.A. (1991) *Mamm. Genome*, **1**, 206–210.
25. Ellegren,H. (1993) *Animal Genet.*, **24**, 367–372.
26. Powell,W., Morgante,M., Andre,C., McNicol,J.W., Machray,G.C., Doyle,J.J., Tingey,S.V. and Rafalski,J.A. (1995) *Curr. Biol.*, **5**, 1023–1029.
27. Provan,J., Corbett,G., Waugh,R., McNicol,J.W., Morgante,M. and Powell,W. (1996) *Proc. R. Soc. Lond. Ser. B Biol. Sci.*, **263**, 1275–1281.
28. Provan,J., Corbett,G., McNicol,J.W. and Powell,W. (1997) *Genome*, **40**, 104–110.
29. Schalling,M., Hudson,T.J., Buetow,K.H. and Housman,D.E. (1993) *Nature Genet.*, **4**, 135–139.
30. Gerry,N., Witowski,N., Barany,G. and Barany,F. (1999) *J. Mol. Biol.*, **292**, 251–262.