Published in final edited form as:

Environ Mol Mutagen. 2014 October ; 55(8): 601–608. doi:10.1002/em.21882.

Pyrosequencing: Applicability for Studying DNA Damage-induced Mutagenesis

Irina G. Minko¹, Lauriel F. Earley^{1,2}, Kimberly E. Larlee¹, Ying-Chih Lin^{1,3}, and R. Stephen Lloyd^{1,2,*}

¹Oregon Institute of Occupational Health Sciences, Oregon Health & Science University, Portland, OR 97239, USA

²Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR 97239, USA

³Cancer Biology Program, Oregon Health & Science University, Portland, OR 97239, USA

Abstract

Site-specifically modified DNAs are routinely used in the study of DNA damage-induced mutagenesis. These analyses involve the creation of DNA vectors containing a lesion at a predetermined position, DNA replication, and detection of mutations at the target site. The final step has previously required the isolation of individual DNA clones, hybridization with radioactively-labeled probes, and verification of mutations by Sanger sequencing. In search for an alternative procedure that would allow direct quantification of sequence variants in a mixed population of DNA molecules, we evaluated the applicability of pyrosequencing to site-specific mutagenesis assays. The progeny DNAs were analyzed that originated from replication of *N*⁶-(deoxy-*D*-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-*N*-methylformamidopyrimidine (MeFapy-dG)-containing vectors in primate cells, with the lesion being positioned in the 5'-GCNGG-3' sequence context. Pyrosequencing detected ~8% G to T transversions and ~3.5% G to A transitions, a result that was in excellent agreement with frequencies previously measured by the standard procedure [Earley et al., 2013]. However, ~3.5% G to C transversions and ~2.0% deletions could not be detected by pyrosequencing. Consistent with these observations, the sensitivity of pyrosequencing for measuring the single deoxynucleotide variants differed depending on the deoxynucleotide identity, and in the given sequence contexts, was determined to be ~1-2% for A and T and ~5% for C. Pyrosequencing of other DNA isolates that were obtained following replication of MeFapy-dG-containing vectors in primate cells or *Escherichia coli*, identified several additional limitations. Collectively, our data demonstrated that pyrosequencing can be used for studying DNA damage-induced mutagenesis as an effective complementary experimental approach to current protocols.

*Correspondence to: Oregon Institute of Occupational Health Sciences, Oregon Health & Science University, L606, 3181 SW Sam Jackson Park Rd, Portland, OR 97239, USA; Tel: + 1 503 494 9957; Fax: + 1 503 494 6831; lloydst@ohsu.edu.

AUTHOR CONTRIBUTIONS Irina G. Minko and R. Stephen Lloyd designed the study. Irina G. Minko, Lauriel F. Earley, Kimberly E. Larlee, and Ying-Chih Lin conducted experiments and analyzed the data. Irina G. Minko wrote the manuscript. R. Stephen Lloyd provided a critical review of the manuscript. All authors approved the submitted version. No authors declare a conflict of interest with respect to this study.

SUPPORTING INFORMATION Additional Supporting Information containing Figures 1S-5S can be found in the online version of this manuscript.

Keywords

MeFapy-deoxyguanosine; COS7; *Escherichia coli*; DNA polymerases; translesion DNA synthesis

INTRODUCTION

To understand the molecular mechanisms that underlie DNA damage-induced mutagenesis, the use of site-specifically modified DNAs has been indispensable [Delaney and Essigmann, 2008]. In this approach, a DNA vector is constructed to contain a lesion in a pre-determined position, introduced into cells or cellular extracts and depending on the experimental system, either fully replicated or subjected to a gap-filling DNA synthesis across the lesion site (Fig. 1, Panel A). Subsequently, the frequencies and spectra of mutations at the target site are assessed. The latter analysis typically involves transformation of *Escherichia coli* to obtain the individual DNA clones, hybridization of these clonally expanded DNAs with different radioactively-labeled oligodeoxynucleotide probes, and verification of mutations by Sanger sequencing (Fig. 1, Panel B-1) [Carty et al., 1996; Earley et al., 2013; Kalam et al., 2006; Kanuri et al., 2002; Moriya, 1993; Shibutani et al., 2001; Yang et al., 2009]. Thus, this final step in the traditional experimental scheme is labor-intensive, costly, and environmentally unfriendly, and the statistical power of the generated data is limited by the number of clones that are practically possible to analyze. Alternative approaches, such as REAP (Restriction Endonuclease And Postlabeling) [Delaney and Essigmann, 2008] and REAMS (Restriction Endonuclease And Mass Spectrometry) [Hong et al., 2007] assays or next-generation sequencing [Yuan et al., 2011], have been utilized, but these methods remain expensive and/or require specialized instrumentation.

Pyrosequencing is a technique that is becoming increasingly popular for quantification of sequence variants in a mixed population of DNA molecules, in particular, to measure CpG methylation, ploidy levels, single nucleotide polymorphic variants, and somatic mutations in heterogeneous tumor tissues [Ahmadian et al., 2006; Bihl et al., 2012a; Bihl et al., 2012b; Fakhrai-Rad et al., 2002; Ogino et al., 2005; Shigaki et al., 2013; Tsiatis et al., 2010; Young et al., 2013]. Pyrosequencing is based on an indirect detection of pyrophosphate, the product of DNA polymerization [Ahmadian et al., 2006; Harrington et al., 2013]. In this method, dNTPs are sequentially added to the primer-template DNA substrate and, if incorporated, pyrophosphate is released and used by sulfurylase for ATP production (Fig. 1, Panel B-2). Subsequently, the ATP serves as a substrate for luciferase to generate light, and the intensity of the emitted light is measured. Unincorporated dNTPs and ATP are degraded by apyrase prior to the addition of the next dNTP. The sensitivity of pyrosequencing for measuring a single deoxynucleotide variant has been reported to be at least 5% or higher [Fakhrai-Rad et al., 2002; Ogino et al., 2005; Tsiatis et al., 2010; Young et al., 2013]. We proposed that this rapid technique could be used for analyses of mixed populations of DNA molecules originating from the replication of lesion-containing DNA, thus facilitating investigations on DNA damage-induced mutagenesis.

Recently, we have reported on mutagenic properties of *N*⁶-(deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-*N*-methylformamidopyrimidine

(MeFapy-dG) in mammalian cells (African Green Monkey kidney cells, COS7) [Earley et al., 2013]. This lesion can be formed following DNA exposure to electrophilic methylating agents, including various chemotherapeutic drugs and environmental toxicants [Gates et al., 2004; Tudek, 2003], and has been detected in rats that were treated with *N,N*-dimethylnitrosamine, *N*-methyl-*N*-nitrosourea, or 1,2-dimethylhydrazine [Beranek et al., 1983; Kadlubar et al., 1984]. Our data revealed high frequencies and diverse spectra of mutations induced by replication past MeFapy-dG in COS7 cells, with both parameters being significantly affected by the local sequence context. Replication of DNA containing this lesion in the 5'-GC(MeFapy-dG)GG-3' sequence context resulted in all possible base substitutions, which were detected at an overall frequency of ~15.1%. In addition, single-nucleotide deletions of a dG were observed at ~1.7% frequency. Due to this complex spectrum, the 5'-GCNGG-3' sequence context was chosen to address the applicability of pyrosequencing for studying DNA damage-induced mutagenesis.

In the current study, we investigated sensitivity of pyrosequencing in the 5'-GCNGG-3' sequence context and compared the data on mutagenic properties of MeFapy-dG that were generated by the traditional, hybridization-based technique and by pyrosequencing.

MATERIALS AND METHODS

Preparation of Vectors Containing Site-specific MeFapy-dG

The 12-mer oligodeoxynucleotides containing a site-specific MeFapy-dG lesion were synthesized as reported previously [Christov et al., 2008]. The corresponding non-damaged oligodeoxynucleotides were obtained from Integrated DNA Technologies. Insertion of the control and MeFapy-dG-containing oligodeoxynucleotides into single-stranded vectors using EcoRV site was performed as described in our earlier study [Earley et al., 2013].

Replication of MeFapy-dG-containing Vectors in COS7 Cells and Determination of the Mutation Spectra and Frequencies Using Hybridization-based Procedure

Transfection of control or lesion-containing vectors into COS7, isolation of the progeny DNAs, and transformation of *E. coli* DH5 α cells (MAX Efficiency DH5 α Competent Cells, Invitrogen) to obtain the individual DNA clones were performed according to the published procedures [Kanuri et al., 2002]. The analyses of DNA clones by a combination of hybridization and Sanger sequencing were done as described previously [Earley et al., 2013].

Replication of MeFapy-dG-containing Vectors in *E. coli* Cells

Control or MeFapy-dG-containing vectors were introduced into *E. coli* cells by electroporation. The *E. coli* strains were either wild type ZK126 (W3110 *lacU169ma-2*) or SF2018 (ZK126 *polB::Spc^r dinB::Kan^r umuDC::Cam^r*), a triple mutant deficient in DNA damage-inducible (SOS) DNA polymerases, pol II, pol IV, and pol V (courtesy of Dr. Steven Finkel, University of Southern California). Following electroporation, cells were incubated for 1 hr at 37 °C in 1 mL of LB medium, and to evaluate the efficiency of transformation, a total of 100 μ L of the culture was plated onto LB-agar containing 100 μ g/ml ampicillin. The remaining culture was transferred into 5 mL of LB medium and

grown overnight at 37 °C in the presence of 100 µg/ml ampicillin. The progeny vectors were isolated from the overnight culture using QIAprep® Spin Miniprep kit (Qiagen) and incubated with EcoRV endonuclease to cleave the vectors containing no insert sequences. The individual DNA clones were obtained by transforming *E. coli* DH5α cells and analyzed for mutations by a combination of hybridization and Sanger sequencing according to the published procedures [Kanuri et al., 2002; Earley et al., 2013].

Preparation of the Standard Samples Containing Mixed Populations of DNA Vectors

The vectors containing 5'-GCNNG-3' sequences, where N was either A, C, G, or T, or the deoxynucleotide was absent, were obtained following replication of the MeFapy-dG-adducted vectors in COS7 cells. The identities of sequence variations were determined by Sanger sequencing. The vectors were clonally expanded in *E. coli* DH5α cells and purified by alkaline extraction [Sambrook et al., 1989]. DNA concentrations were measured in triplicate with PicoGreen reagent (Invitrogen) according to the manufacturer protocol. Fluorescence measurements were performed on Infinite® 200 plate reader (Tecan). DNA vectors containing either A, C, T, or no deoxynucleotide at the variant site were individually added to the G-containing vectors to constitute 1, 5, or 10% of the total DNA. Alternatively, the A-, C-, and T-containing vectors were added to the G-containing vectors in a combination to obtain a DNA sample containing 10% of each minor variant. The 10% sample was sequentially diluted with the G-containing vectors to obtain the 5, 2, and 1% samples. To assure the accuracy of sample preparations, the concentrations of the minor species were verified in a subset of samples by differential hybridization. Specifically, the 5% samples containing either A, C, or T variants were used to transform *E. coli* DH5α cells, and 200 individual clones per sample were analyzed as described previously [Kanuri et al., 2002; Earley et al., 2013]. Consistent with the expected numbers, the A, C, and T-containing variants were found to be represented by 9, 11, and 9 clones, respectively.

Analyses of Sequence Variants by Pyrosequencing

The primers for PCR amplification of the region surrounding the variant site and the sequencing primers were designed using PyroMark Assay Design 2.0 Software (Qiagen). The PCR primers were 5'-biotin-GCCTCTGAGCTATTCCAGAAGTA-3' and 5'-GTCATAGCCGAATAGCCTCTC-3'; the PCR amplicon length was 258 bp. DNA was amplified using PyroMark PCR Master Mix (Qiagen) as recommended by the manufacturer's basic protocol. The reactions were performed in 25 µl and contained 200 nM primers and 1.5 mM MgCl₂. With the standard samples of the sequence variants, 2 ng of the template DNA were used. With the progeny vectors obtained from COS7 or *E. coli* cells, 2% of the total isolated DNA was used as a template for PCR amplification. The PCR conditions included the initial heating step (15 min at 95 °C) to activate HotStarTaq DNA polymerase, 35 cycles of denaturation (30 s at 95 °C), annealing (30 s at 60 °C), and extension (30 s at 72 °C), followed by final extension (10 min at 72 °C). The PCR products were analyzed using PyroMark Q24 instrument (Qiagen) in the Protein and Nucleic Acid Facility at Stanford University (Stanford, CA). The sequencing primers 5'-GAGGGCCCCTGCAAG-3', 5'-CTGCAAGCGATGGAC-3', and 5'-TGCAAGCGATGGACC-3' were designed to initiate DNA synthesis from either -10, -2, or -1 position relative to the target site, respectively. The pyrosequencing assay was

customized for the given reference sequences, as opposed to de novo sequencing. PyroMark Q24 Software (Qiagen) was used to calculate the optimal order of dispensation of dNTPs and analyze the data. The PyroMark Q24 algorithm for measuring the single deoxynucleotide variants includes the following major steps. (a) Different theoretical outcomes for the sequence to be analyzed are determined. (b) The single-peak level is calculated based on the peak heights at all injections, including the sum of peaks at the polymorphic site. (c) The theoretical histograms are adjusted to the same peak levels as the experimental pyrogram. (d) The percentages of DNA variants are determined by comparing the raw data with the different theoretical outcomes. Similar algorithm is also used for measuring the single deoxynucleotide deletions/insertions, but the peak height at the deletion site is not included in calculation of the single-peak level. Statistical analyses were performed using KaleidaGraph 4.1 software (Synergy Software).

RESULTS

Sensitivity of Pyrosequencing in the 5'-GCNGG-3' Sequence Context

Since the sensitivity of pyrosequencing for measuring a single deoxynucleotide variant is known to be affected by the local sequence context [Ogino et al., 2005], the initial set of experiments was conducted to evaluate the sensitivity of pyrosequencing for a sequence of interest, 5'-GCNGG-3'. DNA vectors where N was either A, C, G, or T, or the deoxynucleotide was absent at that position, were obtained as described in Materials and Methods, and these DNAs were used to prepare standard samples with predetermined relative concentrations of sequence variants. The major DNA species in all the samples were vectors containing a G at the variant site. The mixed DNA preparations were used as templates for PCR amplification, and the PCR products were analyzed by pyrosequencing.

Detection of the A, C, and T minor variants individually present in the mixed DNA samples—First, DNA samples were analyzed in which the minor species were represented by either A-, C-, or T-containing molecules, modeling for the simplex patterns of the base-substitution mutagenesis. The A, C, and T variants were individually added to the G variant to be present at 1, 5, or 10% of the total DNA. Three sets of PCR products were independently generated and each was assayed with -10, -2, and -1 primer. As expected, the predominant species in all the cases were found to be G-containing molecules. However, sensitivity of pyrosequencing for measuring the minor species varied. The highest sensitivity was observed for a T at the target site, such that frequencies as low as 1% of this variant could be detected (Table 1 & Fig. 1S). The method was less sensitive for detecting an A, although using -10 primer, a statistically significant difference was observed between the control samples containing 100% of the G variant and the mixed DNA samples that contained 1% of the A variant (Table 1 & Fig. 2S). Measuring the C-containing molecules was most problematic in the given sequence context, probably because of the presence of another C next to the target site; these sequences could only be reliably detected when they constituted 5 or 10% of the total DNA (Table 1 & Fig. 3S).

Notably, the percentage values that were obtained increased as the distance between the end of the sequencing primer and the target site decreased, and the optimal primer positioning

differed depending on the identity of the minor species. T and A variants could be measured more accurately using -10 primer, while values obtained with -2 or -1 primer, were higher than expected. In contrast, measurement of the C variant appeared to be more accurate and reliable when DNA synthesis was initiated from the -2 site. The precision of pyrosequencing for the deoxynucleotide variants was generally high, since with few exceptions, the standard deviations did not exceed 20% of the respective mean values. Consistent with this high precision, the difference between samples that contained 5 or 10% of same sequence variant was highly significant ($p < 0.005$) in all the cases. For the T and A variants, highly significant differences were also observed between samples that contained 1 or 5% of the variant.

Detection of the A, C, and T minor variants concomitantly present in the mixed DNA samples—The population of DNA molecules to be analyzed in site-specific mutagenesis assays (Fig. 1) is often represented by a variety of species and not a simple combination of the wild-type sequence and a single variant. In order to simulate such a situation, a series of DNA samples was prepared that contained a combination of all deoxynucleotide variants. As in the initial experiments, the major DNA species were vectors containing a G at the variant site. The A-, C-, and T-containing vectors were equally present at 1, 2, 5, or 10% each. The pyrosequencing was performed using the -2 primer. The data (Table 2 and Fig. 4S) were generally consistent with the analyses of samples, in which the minor DNA species were present individually (Table 1). For all three variants, 1% frequencies appeared to be below the detection limit, while highly significant differences ($p < 0.005$) were observed between the 5% samples and the baseline levels. The A and T variants could also be detected in the 2% sample.

Detection of deletions—Next, DNA samples were analyzed that contained 1, 5, or 10% of molecules with the target deoxynucleotide being absent. When the assay was performed using the -10 primer, no deletion sequences were detected in any of these samples. The result reports contained a warning about the presence of a homopolymer at the variable position. Using the -2 or -1 primer, the data were inconsistent and warnings were frequently received regarding a high pattern deviation in the variable position. The most concerning results were high background levels (~5 to 26%) and significant deviations of these values that were observed in many independent trials. Thus, measurement of low to moderate levels of deletions appeared to be impossible in the given sequence context.

Validation of Applicability of Pyrosequencing-based Approach for Studying DNA Damage-induced Mutagenesis

Pyrosequencing of the progeny DNAs obtained from replication of the MeFapy-dG-containing vectors in COS-7 cells: the 5'-GCNGG-3' sequence context—Prior analyses using hybridization-based approach demonstrated that replication of DNAs containing MeFapy-dG in the 5'-GC(MeFapy-dG)GG-3' sequence resulted in various base substitutions that were found at an overall frequency of ~15.1% (Table 3, column 2) and ~1.7% of deletions [Earley et al., 2013]. To generate these data, 172 individual DNA clones had been tested that were collected following three independent transfections of COS7 cells with lesion-containing vectors. In the current study, the progeny

DNAs obtained from these experiments were analyzed by pyrosequencing to assess frequencies of base substitutions (Fig. 1, Panel B-2).

Initially, DNAs isolated in Experiment 1 were tested in triplicate, using the -10, -2, and -1 primers (Table 3, columns 3-5, & Fig. 5S). Consistent with the data on sensitivity of pyrosequencing (Table 1 & 2), both G to T transversions and G to A transitions were detected. As with the mixed DNA samples, the data obtained for these sequences were precise with standard deviations being always less than 10% of the respective mean values. Also, the accuracy of measurement appeared to be higher when polymerization was initiated from the -10 site. Although the percentage of G to T transversions detected using the -10 primer was nearly same as was found in the initial hybridization-based analyses, the percentage of G to A transitions was slightly higher than expected. This discrepancy could be explained by possible variability in the relative content of A-containing molecules in different DNA isolates. To address such a possibility, the progeny DNAs obtained in Experiments 2 and 3 were tested by pyrosequencing using the -10 primer (Table 3, columns 6 and 7), and the mean values of the three experiments were calculated (Table 3, column 8). For both G to T and G to A substitutions, an excellent correspondence was found between the data that were generated by the traditional and pyrosequencing-based procedures. However, the molecules with G to C transversions could not be detected with any of the primers.

Pyrosequencing of the progeny DNAs obtained from replication of the MeFapy-dG-containing vectors in COS-7 cells: additional sequence contexts

—Mutagenic properties of MeFapy-dG in COS7 cells have been previously investigated in eight other sequence contexts, in addition to the 5'-GCNCG-3' context [Earley et al., 2013]. The data revealed that the most common type of mutations were G to T transversions that occurred at frequencies of ~4.3 to 13.5%. Other base substitutions were also found, but usually at much lower frequencies. The levels of various types of frame-shift mutations ranged from 0 to up to ~ 8.2%. Here, a subset of the previously obtained DNA isolates was tested by pyrosequencing to obtain an initial assessment regarding the general applicability of this technique to the site-specific mutagenesis assays. PCR products were generated using progeny DNAs that were originated from the 5'-GANGG-3', 5'-GGNGG-3', 5'-GTNCG-3', 5'-GTNAG-3', 5'-GTNCG-3', and 5'-GTNTG-3' lesion-containing vectors as well as the corresponding G-containing control plasmids. Pyrosequencing was performed in triplicates using the -2 primer, and the percentages of the deoxynucleotide variants in DNA isolates were calculated following subtraction of the background levels.

For sequences 5'-GANGG-3', 5'-GGNGG-3', 5'-GTNAG-3', and 5'-GTNCG-3', the major base substitutions, G to T transversions, were detected at the frequencies of 6.3 ± 0.8 , 8.0 ± 1.1 , 7.4 ± 1.5 , and $22.2 \pm 0.6\%$, respectively. The corresponding values obtained by hybridization-based analyses were 4.3, 5.2, 11.8, and 13.5% [Earley et al., 2013]. In addition, pyrosequencing could detect G to A transitions (1.3 ± 1.0 versus 1.7%) and G to C transversions (1.7 ± 0.2 versus 0.9%) in the 5'-GTNAG-3' sequence and G to A transitions (8.4 ± 0.8 versus 5.6%) in the 5'-GTNCG-3' sequence. All other base substitutions were not detectable by pyrosequencing, although based on the hybridization data, these mutations were present at low, ~0.5 to 2.1% frequencies. Thus, this set of DNA isolates could be

analyzed by pyrosequencing. Further analyses are required to substantiate these initial observations, including determination of sensitivity and accuracy for the listed sequences.

The attempts to analyze the DNA isolate that was originated from the 5'-GTNNGG-3' sequence failed. Such a negative outcome could be explained by the presence of a sizable fraction of DNA species containing mutations outside of the interrogated site. In this particular sequence context, MeFapy-dG caused ~5.7% base substitutions and ~6.1% deletions, with the major type being an unusual deletion of a T that occurred immediately downstream of the lesion site [Earley et al., 2013]. The pyrosequencing data for the 5'-GTNTG-3' sequence were even more concerning. There were no warnings in the result report, but the level of G to T transversions was significantly underestimated ($2.2 \pm 0.9\%$ versus 9.7% determined by hybridization). This misleading result could also be attributable to the complex nature of the sample; according to the hybridization-based analyses, there were ~2.9% DNA molecules that contained G, T, or tandem TG deletions [Earley et al., 2013].

Mutagenic Properties of MeFapy-dG in *E. coli*

The mutagenic outcome of replication past MeFapy-dG in bacteria has not been previously examined. In this study, DNA vectors containing MeFapy-dG in the 5'-GC(MeFapy-dG)GG-3' sequence were introduced into *E. coli* cells that were either wild type (ZK126) or a mutant, deficient in all three DNA damage-inducible (SOS) DNA polymerases (SF2018, polIII⁻/polIV⁻/polV⁻). Following replication, mixed populations of the progeny DNAs were isolated and analyzed by pyrosequencing for the presence of base substitutions. No mutated sequences were detected in these DNA isolates obtained from three independent bacterial transformations (data are not shown). Collectively with the data on sensitivity of pyrosequencing in the given sequence contexts (Table 1 & 2), this result suggested that in *E. coli*, the overall frequency of MeFapy-dG-induced base substitutions most likely did not exceed 9% (less than 2% of G to T transversions, 2% of G to A transitions, and 5% of G to C transversions), and was definitely lower than in COS7 cells.

In order to validate the conclusion on a low mutagenic potential of the MeFapy-dG adduct in *E. coli*, 73 clones that were obtained following three independent transformations of the wild-type cells were analyzed by a combination of hybridization and Sanger sequencing. Consistent with the pyrosequencing data, no mutations at the target site were found. However, three C to A transversions were detected at the site immediately downstream of the lesion.

DISCUSSION

The goal of this work was to evaluate the applicability of pyrosequencing for studying DNA damage-induced mutagenesis. It was found that sensitivity of pyrosequencing for measuring deoxynucleotide variants in a mixed population of DNA molecules ranged from as low as 1 to 5%, suggesting that this method can be used to measure moderate to high levels of base substitutions. The precision of measurements was generally high. The data generated by pyrosequencing for the MeFapy-dG-induced G to T and G to A substitutions in COS7 cells in the 5'-GCNNGG-3' sequence context corresponded well to both, pyrosequencing of the

standard mixed DNA samples and the results of prior studies using the traditional, hybridization-based procedure [Earley et al., 2013]. Thus, with certain limitations addressed below, pyrosequencing can be applied to the analyses of DNAs from site-specific mutagenesis assays.

The sensitivity of pyrosequencing for measuring deoxynucleotide variants in the 5'-GCNGG-3' sequence context depended on the identity of the deoxynucleotide variant and the positioning of sequencing primer relative to the target site. The accuracy of measurement was also affected by the deoxynucleotide identity and the primer positioning. Based on these observations, we concluded that each particular sequence intended to be analyzed by pyrosequencing, has to be characterized with regard to the sensitivity and accuracy, and ideally, the standard DNA samples should be included in each pyrosequencing run.

One significant limitation of pyrosequencing is that this method cannot be universally applied for assessing frame-shift mutations. In this study, we could not reliably measure deletions in the 5'-GC[G]GG-3' sequence context when these species constituted 1, 5, or 10% of the total DNA. It is also important to consider that bypass of a DNA lesion by the cellular replication machinery rarely results in deletions/insertions exclusively, and that the concomitant presence of base substitutions in the progeny DNA could further complicate the analyses.

A related concern is that the presence of frame-shift variants in a mixed population of DNA molecules could interfere with the analyses for base substitutions. In the 5'-GCNGG-3' sequence, we could reliably measure the MeFapy-dG-induced G to T transversions and G to A transitions, despite the presence of ~1.7% deletions in this DNA sample [Earley et al., 2013]. The data on the MeFapy-dG-induced base substitutions in the 5'-GGNGG-3' context were also reasonably close to expected, even though this sequence was specifically prone to the frame-shift mutations; a G was deleted in ~7.2% progeny DNAs [Earley et al., 2013]. Thus, low to moderate levels of deletions at the variable position seemed not to have a major effect on the measurements of the single deoxynucleotide variants. We anticipate however that at a certain threshold level, the presence of such deletions will affect the outcome of analyses by interfering with both determination of the single-peak level and calculation of the percentages of variants. Interestingly, when deletions were present outside of the interrogated site, pyrosequencing either failed (the 5'-GTNGG-3' sequence) or produced the data that were inconsistent with the hybridization-based analyses (the 5'-GTNTG-3' sequence). The latter observations suggested that the use of pyrosequencing for studying DNA damage-induced mutagenesis would be limited to the analyses of samples with relatively simple spectra of mutations. In practice, this means that mutagenic properties of the particular DNA adduct should be initially characterized by hybridization (or an alternative well-validated method), whereas pyrosequencing could be used for additional experimental repeats to expedite the study.

The importance of the initial characterization of the mutation spectra by an alternative method was demonstrated in the current study. We attempted to utilize the pyrosequencing as a rapid method for evaluation of a DNA lesion with respect to its ability to cause moderate to high levels of base substitutions. Specifically, targeted mutations were assessed

that were induced by MeFapy-dG in *E. coli* cells. Although subsequent analyses using the traditional procedure confirmed the general conclusion on a low mutagenic potential of this adduct in *E. coli*, unexpected base substitutions were found at the site, immediately downstream of the lesion.

Despite the limitations, pyrosequencing appeared to be a useful screening tool to study various factors that could modulate DNA damage-induced mutagenesis. As an example, we utilized this method to compare mutagenic outcome of replication of the MeFapy-dG-containing vectors in COS7 cells that were either control or depleted of DNA polymerase η . No difference in frequencies of G to T or G to A mutations was found between these DNA isolates. The analyses of the progeny DNAs in this pilot experiment required only few hours, and the cost of interrogation of the entire population of DNA molecules was comparable to the cost of Sanger sequencing, traditionally used in hybridization-based approach to verify mutations in the individual clones. Thus, we envision that applications of the pyrosequencing technique will be amenable to investigating factors that could influence DNA damage-induced mutagenesis, such as proteins involved in replication, DNA repair or cell cycle control, as well as components of the intracellular environment and exogenous stressors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Carmelo J. Rizzo and Plamen P. Christov (Vanderbilt University, Nashville, TN) for synthesis of site-specifically modified MeFapy-dG-containing oligodeoxynucleotides, Steven E. Finkel (University of Southern California, Los Angeles, CA) for bacterial strains, Masaaki Moriya (State University of New York, Stony Brook, NY) for pMS2 vector, Shripa Patel, Director of the Protein and Nucleic Acid Facility at the Stanford University (Stanford, CA) for help with data interpretation, Michael R. Lasarev (Oregon Health & Science University, Portland, OR) for consulting on statistical analyses, and Amanda K. McCullough and Harini Sampath (Oregon Health & Science University, Portland, OR) for the critical reading of the manuscript.

FUNDING This work was supported by National Institutes of Health [grant numbers P01 ES05355 and P01 CA160032 to R.S.L.] and Summer Student Research Program at Oregon Institute of Occupational Health Sciences [to K.E.L].

REFERENCES

- Ahmadian A, Ehn M, Hober S. Pyrosequencing: history, biochemistry and future. *Clin Chim Acta*. 2006; 363:83–94. [PubMed: 16165119]
- Beranek DT, Weis CC, Evans FE, Chetsanga CJ, Kadlubar FF. Identification of N5-methyl-N5-formyl-2,5,6-triamino-4-hydroxypyrimidine as a major adduct in rat liver DNA after treatment with the carcinogens, N,N-dimethylnitrosamine or 1,2-dimethylhydrazine. *Biochem Biophys Res Commun*. 1983; 110:625–31. [PubMed: 6838542]
- Bihl MP, Foerster A, Lugli A, Zlobec I. Characterization of CDKN2A(p16) methylation and impact in colorectal cancer: systematic analysis using pyrosequencing. *J Transl Med*. 2012a; 10:173–182. [PubMed: 22925370]
- Bihl MP, Hoeller S, Andreozzi MC, Foerster A, Ruffe A, Tornillo L, Terracciano L. KRAS mutation testing in colorectal cancer: comparison of the results obtained using 3 different methods for the analysis of codons G12 and G13. *Diagn Mol Pathol*. 2012b; 21:14–23. [PubMed: 22306671]

- Carty MP, Lawrence CW, Dixon K. Complete replication of plasmid DNA containing a single UV-induced lesion in human cell extracts. *J Biol Chem*. 1996; 271:9637–47. [PubMed: 8621639]
- Christov PP, Brown KL, Kozekov ID, Stone MP, Harris TM, Rizzo CJ. Site-specific synthesis and characterization of oligonucleotides containing an N6-(2-deoxy-D-erythro-pentofuranosyl)-2,6-diamino-3,4-dihydro-4-oxo-5-N-methyl formamidopyrimidine lesion, the ring-opened product from N7-methylation of deoxyguanosine. *Chem Res Toxicol*. 2008; 21:2324–33. [PubMed: 19053322]
- Delaney JC, Essigmann JM. Biological properties of single chemical-DNA adducts: a twenty year perspective. *Chem Res Toxicol*. 2008; 21:232–52. [PubMed: 18072751]
- Earley LF, Minko IG, Christov PP, Rizzo CJ, Lloyd RS. Mutagenic spectra arising from replication bypass of the 2,6-diamino-4-hydroxy-N5-methyl formamidopyrimidine adduct in primate cells. *Chem Res Toxicol*. 2013; 26:1108–1114. [PubMed: 23763662]
- Fakhrai-Rad H, Pourmand N, Ronaghi M. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. *Hum Mutat*. 2002; 19:479–85. [PubMed: 11968080]
- Gates KS, Nooner T, Dutta S. Biologically relevant chemical reactions of N7-alkylguanine residues in DNA. *Chem Res Toxicol*. 2004; 17:839–56. [PubMed: 15257608]
- Harrington CT, Lin EI, Olson MT, Eshleman JR. Fundamentals of pyrosequencing. *Arch Pathol Lab Med*. 2013; 137:1296–303. [PubMed: 23991743]
- Hong H, Cao H, Wang Y. Formation and genotoxicity of a guanine-cytosine intrastrand cross-link lesion in vivo. *Nucleic Acids Res*. 2007; 35:7118–27. [PubMed: 17942427]
- Kadlubar FF, Beranek DT, Weis CC, Evans FE, Cox R, Irving CC. Characterization of the purine ring-opened 7-methylguanine and its persistence in rat bladder epithelial DNA after treatment with the carcinogen N-methylnitrosourea. *Carcinogenesis*. 1984; 5:587–92. [PubMed: 6722978]
- Kalam MA, Haraguchi K, Chandani S, Loechler EL, Moriya M, Greenberg MM, Basu AK. Genetic effects of oxidative DNA damages: comparative mutagenesis of the imidazole ring-opened formamidopyrimidines (Fapy lesions) and 8-oxo-purines in simian kidney cells. *Nucleic Acids Res*. 2006; 34:2305–15. [PubMed: 16679449]
- Kanuri M, Minko IG, Nechev LV, Harris TM, Harris CM, Lloyd RS. Error prone translesion synthesis past g-hydroxypropano deoxyguanosine, the primary acrolein-derived adduct in mammalian cells. *J Biol Chem*. 2002; 277:18257–65. [PubMed: 11889127]
- Moriya M. Single-stranded shuttle phagemid for mutagenesis studies in mammalian cells: 8-oxoguanine in DNA induces targeted G.C→T.A transversions in simian kidney cells. *Proc Natl Acad Sci USA*. 1993; 90:1122–6. [PubMed: 8430083]
- Ogino S, Kawasaki T, Brahmandam M, Yan L, Cantor M, Namgyal C, Mino-Kenudson M, Lauwers GY, Loda M, Fuchs CS. Sensitive sequencing method for KRAS mutation detection by pyrosequencing. *J Mol Diagn*. 2005; 7:413–21. [PubMed: 16049314]
- Sambrook, J.; Fritsch, EF.; Maniatis, T. Molecular cloning. A laboratory manual. Second edition. Cold Spring Harbor Laboratory Press; 1989.
- Shibutani S, Suzuki N, Tan X, Johnson F, Grollman AP. Influence of flanking sequence context on the mutagenicity of acetylaminofluorene-derived DNA adducts in mammalian cells. *Biochemistry*. 2001; 40:3717–22. [PubMed: 11297440]
- Shigaki H, Baba Y, Watanabe M, Miyake K, Murata A, Iwagami S, Ishimoto T, Iwatsuki M, Yoshida N, Baba H. KRAS and BRAF mutations in 203 esophageal squamous cell carcinomas: pyrosequencing technology and literature review. *Ann Surg Oncol*. 2013; 20:485–491.
- Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM. Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *J Mol Diagn*. 2010; 12:425–32. [PubMed: 20431034]
- Tudek B. Imidazole ring-opened DNA purines and their biological significance. *J Biochem Mol Biol*. 2003; 36:12–9. [PubMed: 12542970]
- Yang IY, Hashimoto K, de Wind N, Blair IA, Moriya M. Two distinct translesion synthesis pathways across a lipid peroxidation-derived DNA adduct in mammalian cells. *J Biol Chem*. 2009; 284:191–8. [PubMed: 18981176]

- Young EC, Owens MM, Adebiyi I, Bedenham T, Butler R, Callaway J, Cranston T, Crosby C, Cree IA, Dutton L. A comparison of methods for EGFR mutation testing in non-small cell lung cancer. *Diagn Mol Pathol*. 2013; 22:190–5. others. [PubMed: 24193010]
- Yuan B, Wang J, Cao H, Sun R, Wang Y. High-throughput analysis of the mutagenic and cytotoxic properties of DNA lesions by next-generation sequencing. *Nucleic Acids Res*. 2011; 39:5945–54. [PubMed: 21470959]

Figure 1.

Site-specific approach to study DNA damage-induced mutagenesis: generation (A) and analyses of the DNA progenies by hybridization-based procedure (B-1) and pyrosequencing (B-2). To generate the template for pyrosequencing, the region of interest is amplified by PCR using a set of primers with one of these being biotinylated. The biotinylated strand is pulled down and analyzed.

TABLE 1
Detection of the Minor Deoxynucleotide Variants in the Presence of the 5'-GC $\overline{\text{G}}$ GG-3' Major Variant[/]

Minor Species	Expected, %	Experimental, % (mean $\pm$ s.d.)					
		-10 primer		-2 primer		-1 primer	
		Measured	Baseline subtracted	Measured	Baseline subtracted	Measured	Baseline subtracted
5'-GC $\overline{\text{T}}$ GG-3'	0	0.8 $\pm$ 0.1	-	0.4 $\pm$ 0.3	-	0.6 $\pm$ 0.3	-
	1	1.7 $\pm$ 0.1**	0.9 $\pm$ 0.1	1.2 $\pm$ 0.5	ND	1.8 $\pm$ 0.3*	1.2 $\pm$ 0.4
	5	6.0 $\pm$ 0.2**	5.2 $\pm$ 0.2	6.8 $\pm$ 0.7**	6.4 $\pm$ 0.8	7.6 $\pm$ 0.4**	6.6 $\pm$ 0.5
	10	11.5 $\pm$ 0.2**	10.7 $\pm$ 0.2	13.8 $\pm$ 0.5**	13.5 $\pm$ 0.6	14.1 $\pm$ 0.2**	13.5 $\pm$ 0.3
5'-G $\overline{\text{C}}$ AGG-3'	0	0.9 $\pm$ 0.3		0.2 $\pm$ 0.2		0.4 $\pm$ 0.3	
	1	1.4 $\pm$ 0.1*	0.5 $\pm$ 0.3	0.8 $\pm$ 0.4	ND	1.5 $\pm$ 0.6	ND
	5	5.4 $\pm$ 0.1**	4.5 $\pm$ 0.3	6.2 $\pm$ 0.3**	6.0 $\pm$ 0.4	6.5 $\pm$ 0.2**	6.1 $\pm$ 0.3
	10	10.7 $\pm$ 1.2**	9.8 $\pm$ 1.2	12.1 $\pm$ 0.9**	11.9 $\pm$ 0.9	13.2 $\pm$ 0.4**	12.8 $\pm$ 0.5
5'-G $\overline{\text{C}}$ CGG-3'	0	3.5 $\pm$ 0.4	-	0.1 $\pm$ 0.2	-	0.1 $\pm$ 0.1	-
	1	2.4 $\pm$ 0.4	ND	0.0 $\pm$ 0.0	ND	0.6 $\pm$ 0.5	ND
	5	5.9 $\pm$ 2.1	ND	3.8 $\pm$ 0.3**	3.7 $\pm$ 0.4	5.8 $\pm$ 1.7*	5.7 $\pm$ 1.7
	10	11.3 $\pm$ 0.8**	7.8 $\pm$ 0.9	9.4 $\pm$ 0.5**	9.3 $\pm$ 0.5	13.0 $\pm$ 0.6**	12.9 $\pm$ 0.6

[/] Student's t-test was used to compare the data with the corresponding baseline level; one star indicates that p value was < 0.05; two stars indicate that p value was < 0.005. The mean values were calculated following the pyrosequencing of three independently obtained PCR products. S.d. stands for "standard deviation". ND stands for "not detected", e.g. no statistically significant difference was observed between the data and the corresponding baseline level.

TABLE 2
Analyses of Mixed Populations of the Minor Deoxynucleotide Variants in the Presence of the 5'-GCGGG-3' Major Variant¹

Minor Species	Expected, %	Experimental, % (mean ± s.d.)
5'-GCTGG-3'	1	ND
	2	3.2±0.8**
	5	7.2±0.9**
	10	13.3±1.0**
5'-GCAGG-3'	1	ND
	2	2.9±1.0*
	5	6.0±1.2**
	10	11.2±1.1**
5'-GCCGG-3'	1	ND
	2	ND
	5	4.8±2.5**
	10	12.6±2.4**

¹ Student's t-test was used to compare the data with the corresponding baseline level; one star indicates that p value was < 0.05; two stars indicate that p value was < 0.005. The mean values were calculated following the pyrosequencing of three independently obtained PCR products. S.d. stands for "standard deviation". ND stands for "not detected", e.g. no statistically significant difference was observed between the data and the corresponding baseline level.

TABLE 3

Detection of MeFapy-dG-induced Mutations in COS7 Cells¹

Mutation	Traditional protocol, %	Pyrosequencing, %							
		Exp. 1, (mean ± s.d.)				Exp. 2		Exp. 3	
		-10 primer	-2 primer	-1 primer	-10 primer	-10 primer	-10 primer	-10 primer	-10 primer
G to T	8.1	8.3±0.2**	11.0±0.9**	10.5±0.7**	7.1	7.8	7.7±0.6		
G to A	3.5	4.7±0.4**	6.2±0.5**	6.4±0.4**	2.4	3.7	3.6±1.2		
G to C	3.5	ND	ND	ND	ND	ND	ND		ND

¹The pyrosequencing data represent the experimental values with baseline subtracted. Student's t-test was used to compare the data with the corresponding baseline level; two stars indicate that p value was < 0.005. S.d. stands for "standard deviation". ND stands for "not detected", e.g. no differences were observed between the data and the corresponding baseline level.