Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing™

Jörg Tost¹, Jenny Dunker², and Ivo Glynne Gut¹

¹Centre National de Génotypage, Evry, France, and

²Pyrosequencing AB, Uppsala, Sweden

BioTechniques 35:152-156 (July 2003)

DNA methylation acts as a switch that controls gene expression, giving rise to distinct patterns in cells by the interaction of genetic and environmental influences (1). In mammals, it occurs almost exclusively at the 5-position of the pyrimidine ring of cytosines in the context of the dinucleotide CpG. Changes in patterns of DNA methylation are important contributors to the pathogenesis of various human diseases (2), especially cancer where methylation exhibits tumor-specific patterns (3). A wide range of techniques has been presented in the past few years by which qualitative and/or quantitative information about either single methylation variable positions (MVPs) or overall genomic methylation levels can be obtained (4). We recently presented a quantitative assay for epigenotyping based on primer extension and matrix-assisted laser desorption (MALDI) mass spectrometric detection (5). Because of the high density of MVPs in CpG islands, it is often difficult or impossible to place oligonucleotides for primer extension reactions. A possible solution is the use of primers with degenerate bases, but the annealing behavior gets increasingly difficult with the number of degenerate bases incorporated, thereby complicating the calibration and analysis of MVPs (5). Pyrosequencing™ technology (Pyrosequencing AB, Uppsala, Sweden) is a sequencing-by-synthesis method that relies on the luminometric detection of pyrophosphate release upon nucleotide incorporation via an enzymatic cascade (6). It has been applied to genotyping (7) and haplotyping (8) of single nucleotide polymorphisms (SNPs) and microbial typing (9). An important feature

of the technology is the close correlation between the peak heights in a Pyrogram™ (Pyrosequencing AB), representing the sequence raw data, and the amount of nucleotides incorporated by the DNA polymerase during the Pyrosequencing reaction that allows for precise allele frequency determination for a SNP in pooled DNA samples (10–12). Very recently, an approach to the analysis of the degree of methylation on a single MVP after bisulfite treatment was shown using mixtures of PCR products for calibration (13). Bisulfite treatment converts unmethylated cytosines to uracils, while methylated cytosines remain unchanged under the appropriate reaction conditions. Therefore, after PCR, methylation sites can be treated as C/T SNPs with an allele frequency spectrum spanning the entire range (0%–100%). Here we demonstrate the Pyrosequencing technology to analyze and precisely quantify the degree of DNA methylation. Up to six MVPs were examined simultaneously in a single Pyrosequencing reaction. The method is amenable to the analysis of bisulfite-treated DNA derived from paraffin-embedded tissue samples, highly reproducible, and accurate if calibration is carried out properly (to detect any biased PCR amplification).

A fragment of the CpG island encompassing the transcription start of the glutathione-S-transferase π (GSTP1) was chosen for biological reasons. Unmethylated in normal prostate tissue, MVPs in this gene segment are highly methylated in tumors (14). A 140-bp fragment was amplified that contained 17 MVPs, of which 15 were analyzed with four sequencing primers

(Table 1). Sequencing primers were designed with the help of online SNP primer design software (Pyrosequencing AB) where some of the MVPs were replaced by one of the possible bases to limit the number of calculations. Designed sequencing primers were manually checked for mispriming sites. In addition, negative controls were run on the PSQ™ 96MA instrument (Pyrosequencing AB) to check for the occurrence of background generated from the sequencing primers, biotinylated primers, or the template. Sequencing primers showing background were redesigned. PCR primers (MWG, Ebersberg, Germany) were generally designed to be fully complementary to the deaminated strand and did not cover any MVPs. For amplification of the gene segment spanning the transcription start of GSTP1, 10 pmol each forward (5'-GTGATTAGTATTGG-3') and reverse (5'-biotin-AACTCTAA-ACCCCATC-3') primer and 10 ng bisulfite-converted DNA or 10 fmol purified PCR product, respectively, as template were used to yield a 144-bp product. The reaction conditions were 60 mM Tris-SO₄, pH 8.9, 18 mM (NH₄)₂SO₄, 1 mM MgSO₄, 200 μ M dNTPs, and 3 U Platinum® *Taq* DNA polymerase High Fidelity (Invitrogen, Carlsbad, CA, USA) in a 50- μ L volume. Mixtures were denatured for 4 min at 95°C and then thermal-cycled for 30 s at 95°C, 45 s at 50°C, and 20 s at 72°C, repeating the cycle 50 times to ensure complete exhaustion of the primers. A final extension step at 72°C for 4 min terminated the program. Thermal cycling procedures were carried out in an Eppendorf® Mastercycler® 96 Gradient (Brinkmann Instruments, Hamburg, Germany). Preparation of mixtures of PCR products has been described in detail elsewhere (5). ssDNA template was prepared from 20–25 μ L biotinylated PCR product using streptavidin Sepharose® HP beads (Amersham Biosciences, Uppsala, Sweden) following the PSQ 96 sample preparation guide using multichannel pipets. Fifteen picomoles of the respective sequencing primer were added for analysis. Sequencing was performed on a PSQ 96MA system with the SNP Reagent Kit (Pyrosequencing AB) according to the manufacturer's instruc-

Table 1. Assays for the Analysis of 15 MVPs in the GSTP1 Gene

Gene and CpG	Sequencing Primer	Sequence to Analyze	Dispensation Order
GSTP1 M -7, -6, -5 (F)	5'-GTGATTAGTATAGGG-3'	5'-GT/CGGAGT/CGGGGT/CGGG-3'	5'-AGTCTGAGTCAGGTCTG-3'
GSTP1 M -4, -3, -2 (F)	5'-GGATTATTTTATAAGGT-3'	5'-TT/CGGAGGTT/CGT/CGAGG-3'	5'-ATCTGAGTCAGTCTG-3'
GSTP1 M 1, 2, 3 (F)	5'-GTGAGGTTTTTGTGGA-3'	5'-GTTTT/CGTT/CGTT/CGTAG-3'	5'-CGTCAGTCAGTCAGT-3'
GSTP1 M 5, 6, 7, 8, 9, 10 (F)	5'-AGTTTTYGTATTAGTGAGTA-3'	5'-T/CGT/CGT/CGGTTT/CGT/CGTTTT/CGGGGA-3'	5'-ATCAGTCAGTCAGT-CAGTCAGTCTG-3'
GSTP1 M -5, -6, -7 (R)	5'-CCTTATAAAAATAATCCC-3'	5'-G/ACCCCG/ACTCCG/ACCCCT-3'	5'-TGATCCGAGCTCGAGC-3'
GSTP1 M -2, -3, -4 (R)	5'-CTCCAACAAAACCT-3'	5'-CG/ACG/AACCTCCG/AAACC-3'	5'-TCGATCGAGCTCAGA-3'
GSTP1 M 3, 2, 1 (R)	5'-ACTCACTAATAACAAAAC-3'	5'-ACG/AACG/AACG/AAAAC-3'	5'-GACGATCGATCGA-3'
GSTP1 M 10, 9, 8, 7, 6, 5 (R)	5'-TCTAAACCCCATCCCC-3'	5'-G/AAAAACG/AAACCG/ACG/ACG/ATAC-3'	5'-TAGATCGATCAGATC-GATCGATCGAGT-3'
GSTP1 M-4	5'-GGGATTATTTTATAAGG _{pt} T _{CT} T-3'	GOOD assay for epigenotyping	
GSTP1 M10	5'-AACTCTAAACCCCATC _{pt} C _{pt} C _{CT} T-3'	GOOD assay for epigenotyping	

Sequencing primers (F, forward; R, reverse), sequence to analyze, and order of nucleotides dispensed by the PSQ 96MA instrument used for the analysis of CpG positions in the GSTP1 fragment by the Pyrosequencing technology and mass spectrometry (pt, phosphorothioate; CT, charge tag). The nomenclature of the positions refers to the transcription start. The sequencing primer GSTP1 M-7 contains a mismatch to reduce background signals due to mispriming to the template in the unmethylated state.

tion without further optimization. For quantitation, four PCR products/sample were analyzed to account for differences in the PCR performance. The degree of methylation was calculated as allele frequency [methylation % = peak height methylated/(peak height methylated + peak height unmethylated)*100] using the allele quantification

functionality of the PSQ 96MA software. Calibration curves were recorded for the 15 positions using five mixtures of PCR products (0%, 25%, 50%, 75%, and 100% methylation) prepared from cloned completely methylated and unmethylated human DNA. Highly linear calibration curves were recorded with a correlation factor greater than or equal

to 0.99, a slope of approximately 1, and a standard deviation of approximately 1% for all positions (Figure 1A, position 10). However, if calibration was performed after mixing the two cloned DNA templates before PCR amplification, then calibration did not show a linear trend. A polynomial fit of the third order gave correlation coefficients greater than 0.99 and standard deviations around 1% for each sample and each position (Figure 1B, position 10). Figure 1C shows the resulting Pyrograms for positions 5–10. This example stresses several strengths of the Pyrosequencing assay. Six MVPs could be analyzed simultaneously, thereby significantly decreasing the cost and effort per position. Position 10 is in a homopolymeric stretch that results in a 4- to 5-fold incorporation of T, but calibration was as accurate as for the other positions. Calibration curves only differed slightly between the 15 positions analyzed. As the Pyrosequencing technology has been shown to be a quantitative detection platform giving absolute results, the shape of the calibration curves depends exclusively on the PCR amplification. In this case, the shape is due to a PCR

Figure 1. Calibration of Pyrosequencing methylation analysis. (A) Calibration curve for the MVP 10 of the GSTP1 gene segment using five mixtures of PCR products (0%, 25%, 50%, 75%, and 100% methylation) prepared from cloned completely methylated and unmethylated human DNA (mean of four replicates). Calibration curve (B) of 12 PCR-amplified mixtures with known methylation degree (0%, 10%, 20%...50%, to 100%, with an additional point at 95% methylation) and selected Pyrograms for the sequencing primer analyzing positions 5–10 (C).

Table 2. Results of Analysis of Nine Prostate Cancer Samples by the Pyrosequencing Technology

Sample	Pyrosequencing															Mass Spectrometry		
	CpG-7	CpG-6	CpG-5	CpG-4	CpG-3	CpG-2	CpG-1	CpG-2	CpG-3	CpG-4	CpG-5	CpG-6	CpG-7	CpG-8	CpG-9	CpG-10	CpG-4	CpG-10
1	69.3	68.6	72.9	63.4	72.9	61.8	66.4	70.8	62.5	60.2	61.2	63.4	66.5	66.0	68.7	61.2	63.2	63.2
2	70.3	68.8	73.5	63.3	65.5	63.5	65.4	68.9	57.3	57.5	60.5	61.0	65.8	63.7	67.2	60.3	62.4	62.4
3	65.3	65.0	68.7	67.0	70.1	62.6	68.7	70.7	58.8	55.8	56.2	57.0	61.0	60.1	63.2	64.2	63.9	63.9
4	62.4	60.5	68.6	63.4	64.0	62.1	67.9	65.8	55.4	53.8	55.7	56.4	58.8	57.5	59.7	65.8	57.4	57.4
5	58.7	55.9	71.8	82.2	81.2	80.1	82.7	93.8	90.2	69.5	67.4	69.5	64.3	65.5	68.2	84.7	65.7	65.7
6	56.3	58.8	57.8	62.0	69.6	58.0	63.8	66.8	60.8	57.4	54.5	56.7	59.8	55.0	61.0	62.0	58.3	58.3
7	63.5	62.2	69.4	79.6	85.1	75.7	79.5	90.0	86.6	77.3	73.4	71.0	72.9	70.0	77.7	77.4	74.3	74.3
8	68.2	64.8	70.0	77.0	88.4	73.5	80.1	90.7	86.1	67.6	64.1	65.1	63.4	62.1	71.1	71.4	71.4	71.4
9	39.2	36.5	33.3	48.0	58.5	41.6	47.4	66.5	55.5	42.9	40.2	39.4	42.5	45.3	47.0	46.8	43.7	43.7

Analysis of the degree of methylation (in %) at 15 CpG positions within the GSTP1 PCR product in paraffin-embedded cancerous prostate samples by the Pyrosequencing technology and verification of two positions by the GOOD assay for epigenotyping.

bias towards the unmethylated allele caused by a sequence difference of 12% between the completely methylated and unmethylated allele. The difference in the number of hydrogen bonds (methylated = C-rich, unmethylated = T-rich) strongly influences the melting and annealing properties during PCR amplification. This underlines the necessity for control for preferential amplification during PCR or appropriate calibration accounting for this phenomenon.

We analyzed nine samples taken from different regions of a paraffin-embedded cancerous prostate tissue biopsy, which was extracted and bisulfite-treated as previously described (5). All samples could be successfully analyzed (Table 2). Figure 2 shows an example of the analysis of 15 MVPs with the four different sequencing primers of one sample. A high degree of co-methylation was found, and samples ranged in their degree of methylation between 30% and 90% in terms of the position of sampling and the distance to the center of the tumor. Standard deviation was again approximately 1% if four PCR products of the same sample were analyzed to account for differences in PCR

performance. To verify our results, we also used primers in reverse direction, sequencing a single-stranded PCR product amplified using a biotinylated forward primer. Again, all positions could be analyzed successfully. However, the accuracy was slightly lower than for the forward sequencing primers, as the methylation degree on MVPs is now analyzed as A/G SNP (standard deviation 2%–3%). A possible explanation is the use of α -S-dATP instead of dATP in the Pyrosequencing reaction, resulting in slightly higher A signal intensities in the Pyrogram compared to G, T, or C. If the calibration curves were recorded at the same time as the analysis of the biological samples was performed, the same accuracy as with forward sequencing primers could be obtained. This approach may offer an alternative approach if sequencing primers in the forward direction give rise to background signals that may complicate the accurate analysis of MVPs.

For further verification, two MVPs (M-4 and M10) were analyzed by the GOOD assay for epigenotyping (5) and by the Pyrosequencing technology. The two positions were chosen for epigenotyping with mass spectrometric detection, as the primers for primer extension reaction (Table 1) did not contain any degenerate bases that might complicate the annealing behavior. For direct comparison of biological samples between the two technologies, 50- μ L PCRs were split and analyzed by both methods. Calibration curves for the

Figure 2. Analysis of the GSTP1 fragment of bisulfite-treated genomic DNA extracted from a paraffin-embedded prostate biopsy. Fifteen of the 17 MVPs in the fragment derived from sample 6 in Table 2 were analyzed with four sequencing primers, GSTP1 M-7 (A), M-4 (B), M1 (C), and M5 (D).

GOOD assay were similar to those obtained by the Pyrosequencing technology. Results obtained with the two technologies differed at most by 5% (Table 2). The mean standard deviation for mass spectrometric results was around 2%, while it was around 1% for analysis with the PSQ 96MA. Our experiments demonstrate for the first time the simultaneous quantitative analysis of several variable positions in a single Pyrosequencing reaction. The simultaneous analysis of up to six positions provides a viable alternative to multiplexing using several sequencing primers, which is difficult because of the low sequence diversity of bisulfite-treated DNA. Sequencing primers should only be positioned in sequence stretches containing no MVPs, thereby not introducing further possibilities of bias. All MVPs including those in homopolymeric sequences could be successfully analyzed in biological samples without a loss of accuracy. An important factor to stress if dealing with bisulfite-treated genomic DNA is to check for preferential amplification of an allele during PCR. Calibration using mixtures of PCR products amplified from completely methylated and unmethylated DNA may not be necessary, as the Pyrosequencing technology—in the absence of a PCR bias—gives absolute quantitative results. However, the sequence difference after bisulfite treatment of methylated and unmethylated DNA may give rise to PCR bias (15), which is impossible to predict. The approach of recording calibration curves for the PCR fragments of interest allows for accurate quantification and compensates for any effects influencing the calibration results.

The application of Pyrosequencing technology for the analysis of DNA methylation is quite straightforward, and no further optimization of the technology was necessary. A major advantage compared to primer extension methods for allele discrimination is the possibility to analyze practically all MVPs in a fragment of interest with few sequencing primers. Usually, not all MVPs are accessible to primer extension reactions because of their close proximity. Sequencing primers can be placed in regions containing no variable positions, as the use of primers

containing degenerate bases has been shown to influence the annealing behavior and may complicate accurate quantification (5). The biological samples analyzed in this study exhibited a high degree of co-methylation. Therefore, it may be sufficient to analyze only a few MVPs to yield sufficient information, which can be done by both technologies with sufficient resolution. Because of its possibility of automation, the GOOD assay has its advantages when it comes to the analysis of a large number of samples. For smaller sample sizes and first screens for differentially methylated positions, the Pyrosequencing technology may be more appropriate. Our study comes into context with current pilot projects to create a map of epigenetic patterns in different nondiseased human tissues (1) to localize these MVPs. The presented assay is a fast, easy-to-implement, and accurate assay for the analysis and quantification of DNA methylation.

ACKNOWLEDGMENTS

We thank Epigenomics AG (Berlin, Germany) for the PCR products derived from clones of bisulfite-treated completely methylated respectively unmethylated human genomic DNA and assistance with bisulfite treatment. The prostate tissue sample was kindly provided by Dr. Glen Kristiansen (Charité, Berlin, Germany). In addition, thanks to Christina Johansson, Monica Pettersson, and Maria Röjmyr at Pyrosequencing AB for good and valuable discussions. This work was in part supported by a grant from the European Union (QLRT-1999-30417) and the Ministère de l'Éducation, Recherche et Technologie of the French Government (MENRT).

REFERENCES

- Novik, K.L., I. Nimmrich, B. Genc, S. Maier, C. Piepenbrock, A. Olek, and S. Beck. 2002. Epigenomics: genome-wide study of methylation phenomena. *Curr. Issues Mol. Biol.* 4:111-128.
- Robertson, K.D. and A.P. Wolffe. 2000. DNA methylation in health and disease. *Nat. Rev. Genet.* 1:11-19.
- Jones, P.A. and S.B. Baylin. 2002. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* 3:415-428.
- Laird, P.W. 2003. Early detection: the power and the promise of DNA methylation markers. *Nat. Rev. Cancer* 3:253-266.
- Tost, J., P. Schatz, M. Schuster, K. Berlin, and I.G. Gut. 2003. Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res.* 31:E50.
- Ronaghi, M., M. Uhlen, and P. Nyren. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281:363-365.
- Ahmadian, A., B. Gharizadeh, A.C. Gustafsson, F. Sterky, P. Nyren, M. Uhlen, and J. Lundeberg. 2000. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* 280:103-110.
- Odeberg, J., K. Holmberg, P. Eriksson, and M. Uhlen. 2002. Molecular haplotyping by Pyrosequencing™. *BioTechniques* 33:1104-1108.
- Ronaghi, M. and E. Elahi. 2002. Pyrosequencing for microbial typing. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 782:67-72.
- Gruber, J.D., P.B. Colligan, and J.K. Wolford. 2002. Estimation of single nucleotide polymorphism allele frequency in DNA pools by using Pyrosequencing. *Hum. Genet.* 110:395-401.
- Neve, B., P. Froguel, L. Corset, E. Vaillant, V. Vatin, and P. Boutin. 2002. Rapid SNP allele frequency determination in genomic DNA pools by Pyrosequencing™. *BioTechniques* 32:1138-1142.
- Wasson, J., G. Skolnick, L. Love-Gregory, and M.A. Permutt. 2002. Assessing allele frequencies of single nucleotide polymorphisms in DNA pools by Pyrosequencing™ technology. *BioTechniques* 32:1144-1150.
- Uhlmann, K., A. Brinckmann, M.R. Toliat, H. Ritter, and P. Nürnberg. 2002. Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* 23:4072-4079.
- Millar, D.S., K.K. Ow, C.L. Paul, P.J. Russell, P.L. Molloy, and S.J. Clark. 1999. Detailed methylation analysis of the glutathione S-transferase pi (GSTP1) gene in prostate cancer. *Oncogene* 18:1313-1324.
- Warnecke, P.M., C. Stirzaker, J.R. Melki, D.S. Millar, C.L. Paul, and S.J. Clark. 1997. Detection and measurement of PCR bias in quantitative methylation analysis of bisulfite-treated DNA. *Nucleic Acids Res.* 25:4422-4426.

Received 4 April 2003; accepted 2 May 2003.

Address correspondence to Ivo Glynne Gut, Centre National de Génotypage, Bâtiment G2, 2 rue Gaston Crémieux, CP 5721, 91057 Evry Cedex, France. e-mail: ivogut@cng.fr