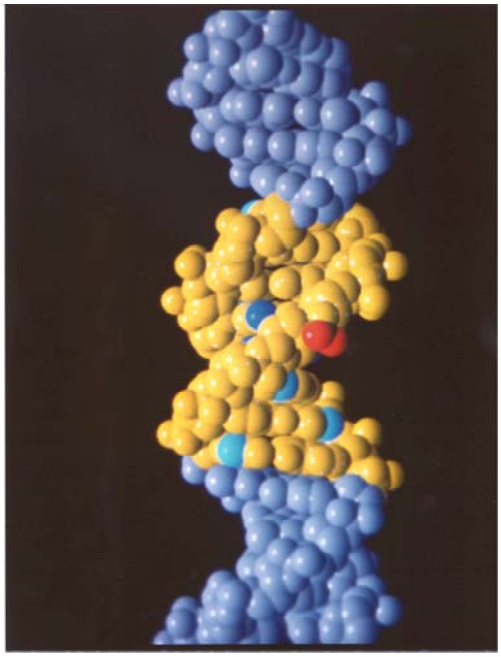
Methylation Analysis by Bisulfite Sequencing: Chemistry, Products and Protocols from Applied Biosystems



Bisulfite Sequencing

I Introduction

Methylation of C as 5mC in CpG dinucleotides in the promoter region of a gene has been associated with transcriptional silencing and plays a central role in epigenetics [1-6]. Analysis of this type of methylation in gDNA can be achieved directly by the use of methylation-sensitive restriction enzymes, or after acid-catalyzed conversion of gDNA with bisulfite that is selective for C compared to 5mC [7], This methodology was discovered by Hayatsu [8] in 1970.

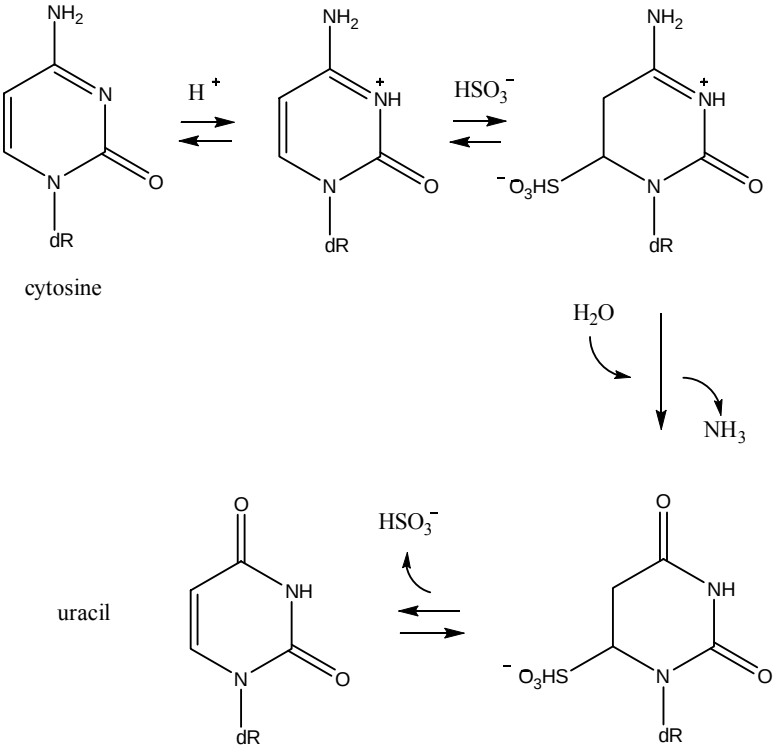


Figure 1. Chemical scheme for the conversion of cytosine to uracil. Cytosine reacts with bisulfite, but 5-Me-cytsoine does not react.

As shown in Figure 1, stepwise reaction of bisulfite with protonated C leads to deamination to form uracil (U) via reversible formation of an intermediate sulfonated adduct. This highly selective deamination of C to U, without significant conversion of 5mC to T, is presumably due, at least in part, to greater steric interference between bisulfite and CH_3 vs. H during cis-addition to the 5,6-double bond, as illustrated in Figure 2.

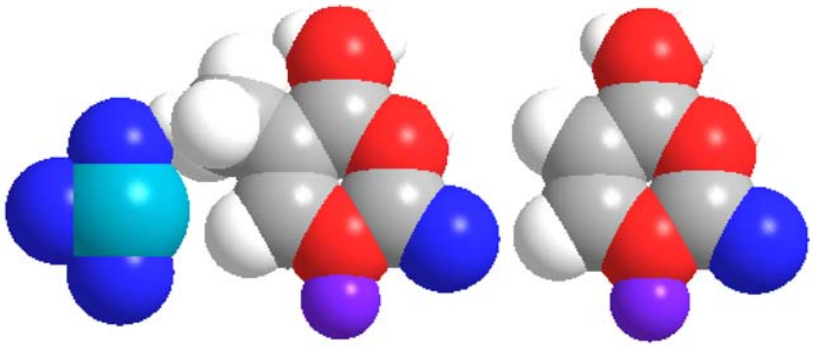


Figure 2. Acid-catalyzed cis-addition of bisulfite (HSO_3^-) to the 5,6-double bond of cytosine is favored relative to 5-methylcytosine due in part to greater steric repulsion between bisulfite and CH3 vs. H, as illustrated by space-filling 3-D structures.

Regardless of mechanistic details, the change in DNA sequence upon complete conversion by bisulfite allows methylation analysis by DNA sequencing and other DNA detection methods. To obtain sufficient analyte after bisulfite conversion, PCR is generally performed, which leads to amplicons wherein Ts replace Us (i.e., former Cs) and Cs replace 5mC. On a statistical basis, only $1/4^{th}$ of all possible CpN dinucleotides are present as a CpG motif in gDNA, and are therefore eligible for the action of methyltransferases. Consequently, bisulfite treatment eliminates all Cs that are not present as 5mCs by replacement with Ts, and thus creates a nearly C-less sequence of mostly 3-base-DNA having predominantly A, G, and (~50%) T. This reduction in sequence complexity, relative to normal 4-base-DNA, can lead to lower specificity of hybridization with probes or primers [9, 10]. It follows that PCR- and other probe-based methods for detection of bisulfite-converted, 3-base-DNA loci share a common constraint regarding specificity.

Examples of problematic PCR and sequencing traces are presented herein to illustrate diagnosis of problematic bisulfite-sequencing data. Some aspects of the problems described are not unique to bisulfite-treated gDNA, and are encountered in other types of DNA analyses. Although capillary electrophoresis offers many options for the detection of methylation in gDNA (figure 3), this troubleshooting guide will focus on improvements to bisulfite-sequencing, which most methylation researchers depend on at some point during their investigations, and on fragment analysis which offers a simple alternative to sequencing.

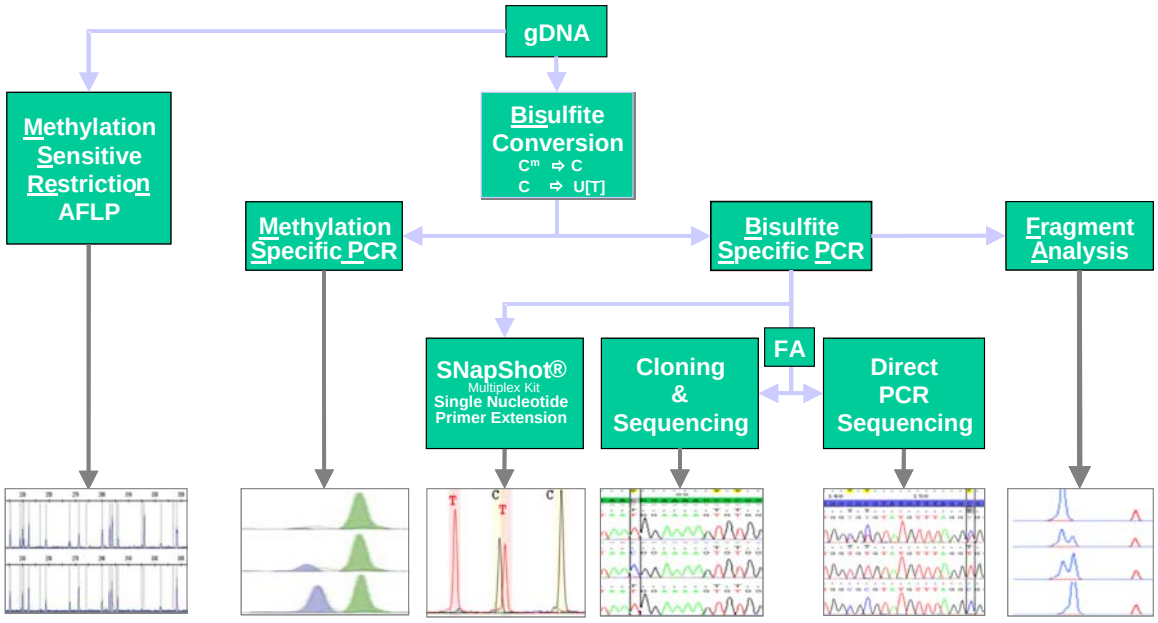


Figure 3. Workflow options for the Detection of methylation using a Capillary Electrophoresis (CE) System.

II. Workflow

A. Summary

Irrespective of the CE analysis method used, the upstream workflow for the preparation of the DNA is the same, as demonstrated in figure 4. Briefly, DNA is extracted and bisulfite converted. The bisulfite converted gDNA serves as a template in PCR using region specific primers followed by analysis either by CE fragment separation or

sequencing. Each of these steps will be dealt with in detail, identifying critical issues in each step and how they may affect your results.

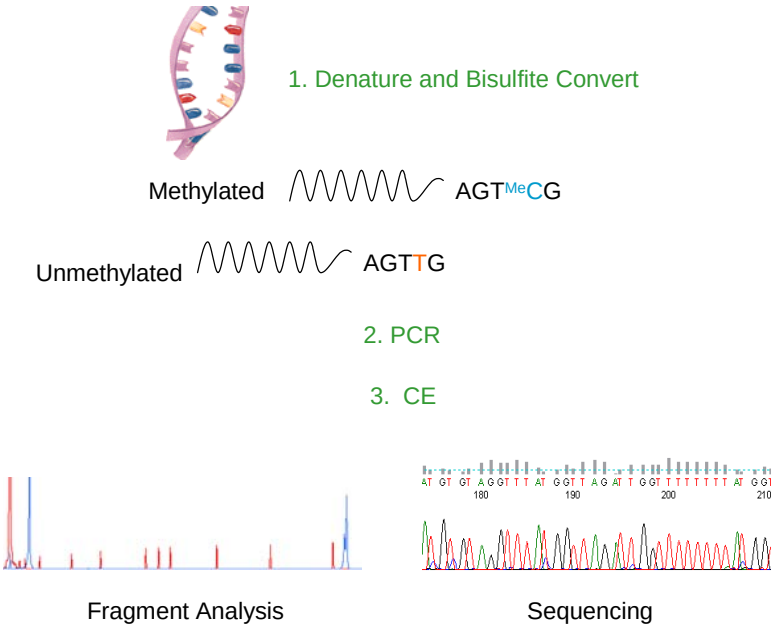


Figure 4. Methylation determination is readily achieved using AB's methylSEQr™ Bisulfite Conversion Kit and CE instrumentation for DNA analysis.

B. DNA Extraction

A published study [11] describes the importance of the purity of gDNA for the success of complete bisulfite conversion (see also figure 19). A gDNA sample was bisulfite converted both with and without a prior proteinase K incubation. The sample without the proteinase K step had both random and large section of non-converted sequence. Incomplete bisulfite conversion also occurs when too much gDNA is being bisulfite converted in a single well, or if the gDNA is not fully dentatured prior to the bisulfite treatment.

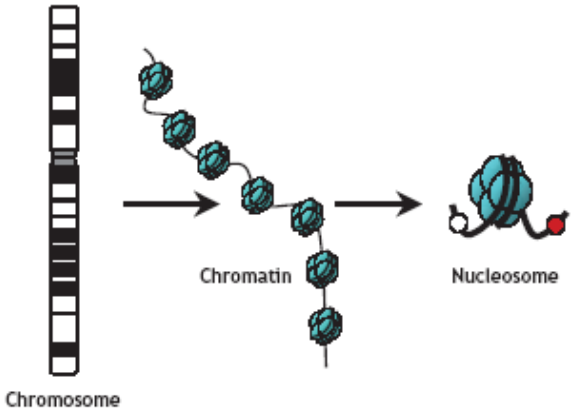


Figure 5. Genomic DNA is tightly associated with positively charged proteins called Histones. The duplex, coiled gDNA is alternately wound around a histone octamer to form a unit called a nucleosome, and separated from other nucleosomes by unbound stretches of DNA. Chromatin is a string of nucleosomes which is further compacted to form the chromosomes. Chromosomes are therefore about 50% protein in content. The protein must be removed prior to bisulfite conversion.

Formalin fixed paraffin embedded (FFPE) tissues are particularly challenging for bisulfite conversion as the tissue fixing process causes modifications to the DNA, including DNA-protein crosslinks. DNA extracted from FFPE tissue that includes a vigorous protease digestion step ensures successful removal of the bound protein from the DNA. The RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE tissue from Ambion employs a thorough on-filter protease digestion. The gDNA from FFPE sources using RecoverALL yields a bisulfite-converted product suitable for PCR, providing the overall quality of the gDNA was not seriously damaged during preservation.

Figure 6. The RecoverAll™ Total Nucleic Acid Isolation Kit for FFPE tissue provide gDNA suitable for bisulfite conversion using the methylSEQr™ Bisulfite Conversion Kit.

C. Bisulfite Conversion

Achieving essentially complete and highly selective bisulfite conversion of non-methylated Cs among 3 billion bases in the human genome, without significant interfering side-reactions or extensive cleavage of gDNA is a remarkable achievement. Treatment of denatured gDNA with relatively high concentrations (~3-9 M) [12] of bisulfite in freshly prepared solutions with added antioxidant and carefully adjusted pH at elevated temperature for extended times is required. An often cited (e.g.,[13-15]) study has reported [16] that as much as 95% of gDNA is degraded during these types of forcing reaction conditions for bisulfite conversion. However, our recently published [17] finding of excellent recoveries of converted DNA using centrifugal filtration for post-bisulfite purification indicates that degradation of DNA during the conversion process may not be as problematic as is generally thought based on earlier findings [16].

Microcon® Centrifugal Filter Units

Figure 7. A key feature of the Applied Biosystems methylSEQr™ kit is the use of centrifugal filtration for the isolation and purification of gDNA after treatment with bisulfite. This unique method reliably provides a high recovery of shelf-stable bisulfite-converted gDNA.

We believe that, under appropriate reaction conditions, gDNA is not extensively degraded but may instead be lost during recovery using techniques originally optimized for unmodified duplex DNA. The bisulfite-converted gDNA intermediate is single-stranded, due to loss of complementarity caused by the replacement of Cs with Us, and is more negatively charged due to an additional sulfonate group bonded to every U (cf. Figure 1). The sulfonated U moieties are desulfonated while still contained in the filtration device, thus eliminating the possibility of biased fractionation and/or other means of loss of material during isolation of converted gDNA. This centrifugal filtration protocol [17] thoroughly washes away all of the bisulfite, and allows complete dissolution of the intermediate sulfonated gDNA in 0.1 M NaOH for desulfonation. In our experience, the final bisulfite-converted gDNA obtained in this manner can be stored for long periods in a refrigerator, in contrast to other reported protocols that recommend immediate use or frozen storage. Bisulfite-converted gDNA purified by the centrifugal filtration protocol [17] and stored at 4°C has yielded consistent results over many months (2+ years).

Figure 8. The methylSEQr ™ Bisulfite Conversion Kit provides consistently high yields of shelf stable bisulfite-converted gDNA.

Selecting a Region of Interest for Methylation Analysis

Methylation is a dynamic process that varies with health, age, diet, environment, and is probably heritable [18-20]. Eukaryotes have both de novo methyltransferases and maintenance methyltransferaseases [21]. Demethylation may be passive, due to inefficient or lack of maintenance methylation during cell division, or actively mediated by enzymatic removal by processes that have been recently elucidated [22]. Methylation in cancer is believed to be a progressive event, whereby the percentage of a region that is methylated increases as the disease progresses [23]. One model proposes that methylation is initiated near the transcription start site, and spreads out over the entire region as loss of gene expression becomes more pronounced and the cancer progresses. CpG motifs are underrepresented in the human genome; however, a significant percentage of genes have regions of relatively high CpG-density in the promoter known as CpG islands. Readers are encouraged to consult a recent publication by Berg et al. [24] for an excellent update on the sequence-based definition and distribution of CpG islands in the human genome that provides new and important perspectives on this fundamental topic.

The correlation between loss of expression and methylation is most likely to be observed the closer the methylation occurs to the transcription start site. When selecting a gDNA region for methylation studies, a typical workflow would include a gene expression

study, the identification of genes which are downregulated, and an NCBI search to locate the promoter sequence and transcription start site of the gene. The selected genomic sequence is then used to aid in the selection of primers.

E. Designing Primers

While good primer design is critical for successful PCR in any analysis, designing primers for methylated DNA has many additional confounding factors to consider. The sequences obtained from methylated gDNA and unmethylated gDNA are fundamentally different after bisulfite conversion. The sequence from methylated gDNA will still have Cs at CpGs, while that obtained from unmethylated gDNA will have no Cs. Primers for bisulfite-converted gDNA can be designed to anneal to a sequence specific for methylated gDNA, or unmethylated gDNA, or designed to a region without CpGs so that PCR amplification is not dependent on methylation status.

Primers directed at CpG flanking sequences in bisulfite-converted gDNA have only 3 bases and are T-rich. This oftentimes leads to primers >30 bases in length for obtaining typical PCR T_m-values of ~60 °C. These relatively long 3-base T-rich primers are apparently more prone to mismatch hybridization since multiple amplification products are frequently obtained. The reverse primer is A-rich and thus provides favorable conditions for primer dimer formation with the T-rich forward primer. Once the first cycle of amplification wrongly amplifies a mismatched sequence, subsequent PCR cycles are correctly matched, so PCR proceeds as efficiently as it would for the correctly matched sequence. Smaller amplicons from either primer dimer or mismatched secondary amplicons have greater amplification efficiency, and can out-compete amplification of the amplicon of interest. The presence of undesired amplicons usually diminishes the PCR efficiency of the intended amplicon.

When selecting primers, a SNP database (dbSNP, SNP500, SNPbrowser) should be consulted to avoid designing primers over a SNP. Additionally, regions that are methylated, especially when investigating cancer-derived samples, are reportedly often

highly mutated, leading to an inability to PCR amplify with primer annealing sites based on non-mutated sequence.

Methyl Primer Express® Software is a free online primer design tool specifically for methylation studies which assits in designing primers in both methylated and unmethylated bisulfite modified DNA. Users simply cut and paste in the selected genomic sequence, the software then performs an *in-silico* bisulfite conversion (C's are converted to T's), and aids in the selection of primers. Methyl Primer Express software is available for free download at: (http://marketing.appliedbiosystems.com/mk/get/GAAS_CLINICAL_METHYLATED?_A=77005&_D=50613&_V=0#)

a) **Methylation-specific PCR (MSP)**

First reported in 1996 by Herman et al [25], methylation specific PCR (MSP) has become a widely used technique that is good for a "yes or no" indication concerning methylation of CpG sites. Two primer sets are designed to anneal to a region containing CpG motifs. The methylated primer set assumes the CpG's are fully methylated, thus the primer will have all 4 bases in the sequence. The unmethylated primer set anneals to gDNA that is not methylated in the (same) primer binding site, and therefore will have T in place of C in the primer sequence. Design guidelines generally have a C or T near the 3' end of the two primer sets, respectively, where any mismatches will be discriminated against by the polymerase. It is important to test the primer sets with a control gDNA of known methylation status along with the gDNA of unknown methylation status. The properly designed methylated primer set will only amplify the control methylated gDNA, and not unmethylated gDNA, and the unmethylated primer set will be positive for only unmethylated gDNA.

b) **Bisulfite Sequencing.**

Primers designed outside of a CpG region of interest will, in principle, amplify the target regardless of the methylation state of the internal sequence. Bisulfite sequencing provides

an inherently more accurate assessment of the methylation state of a sample compared to PCR primers (or probes) that select for presupposed fully methylated or fully unmethylated complementary sequences, such as MSP. The percentage of methylation, or methylation patterns are believed to be dynamic and changing during the life of a cell. Pilot genome-wide bisulfite sequencing studies [26] have demonstrated that methylation is approximately bimodal, i.e., while sequences were mostly either methylated or unmethylated throughout the region of interest there are a relatively low level of sequences with intermediate methylation states. Consequently, the nature of the methylation information being sought (i.e., exact at every CpG in a clonally pure allele vs. approximate average over all alleles) will determine whether primers should be designed to anneal in CpG-free regions or not.

Primer annealing sites are relatively restricted due to limited availability of regions that span a CpG island yet exclude CpGs in the primer annealing site. Selection of these primers can be further constrained by the relatively large number of closely spaced CpG motifs, which lead to primer sites too short to achieve a desirable T_m. An additional constraint is the presence of relatively long (>9 bases) poly(T) sequences in bisulfite-converted template that results in non-specific primer annealing, or poor amplification due to polymerase slippage.

Figure 9. Bar graph representing a bisulfite converted region of gDNA showing non-CpG primer annealing sites flanking a CpG –rich (pink hash marks) region (MethylPrimerExpress Software). Primer selection for bisulfite-PCR is often limited due to the multiple reasons described in the text. The arrows represents the forward primer

(red), and the reverse (orange) with the arrow closest to the bar the first choice, based on the user's criteria defined on a previous page of the software program.

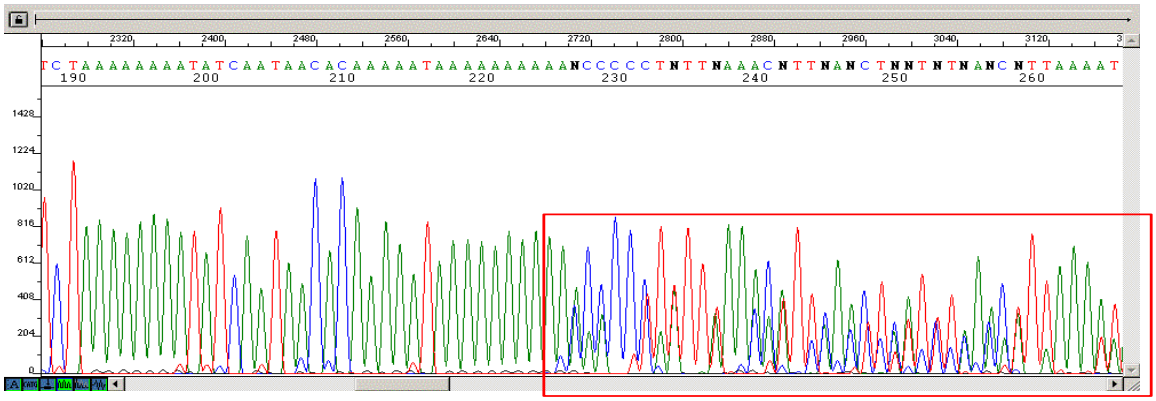


Figure 10. PCR slippage resulting in loss of callable antisense-strand sequencing (boxed sequence). The sequence from bases 188-216 is clearly resolved. However, following the homopolymer stretch of As from bases 217-226, there is both insertion and deletion of As due to polymerase slippage during PCR amplification that precedes sequencing. In CE-based fragment analysis, this PCR product mixture would be detected as a distribution of N+1 and N-1 peaks centered around the correct-size amplicon.

Primer selection is therefore often limited to a single locus with limited ability to lengthen or shorten the primer. If there are no suitable primers, the antisense strand derived from the reverse complement of the original input gDNA sequence can be used. The complementary nature of the sense and antisense strands is eliminated following bisulfite conversion. The same number of CpGs (and methylation state) will be present in both the sense and antisense strands due to the symmetry of the CpG motif, and action of methyltransferase, although the sequence of the bisulfite-converted sense and antisense strands will differ considerably.

F. PCR Conditions

a) Polymerase selection

After bisulfite conversion of gDNA, double-stranded nucleic acid is transformed into single stranded template, and is comprised of 5 different bases: A, G, T, U and 5mC. Methylated promoter regions will still be 5mCpG-rich, and likely have single-stranded

secondary structure. 5mC reportedly enhances the T_m of 5mC-G base pairing relative to C-G by 1.2 °C [27]. C-G base-pair-rich regions are known to be difficult to PCR amplify. The polymerase must be capable of reading U and 5mC during the first round of synthesis of the reverse-complement strand, much like "first strand" synthesis in reverse transcription PCR. A hot start polymerase in conjunction with relatively high temperature should be used to avoid mismatch amplification. High-fidelity polymerases from archaebacteria such as Vent or pfu DNA polymerase do not accommodate U-containing template [28].

It should be noted that PCR master mixes containing uracil DNA glycosylase (UNG) should not be used. UNG cleaves U-containing DNA and will degrade U-containing template produced by bisulfite conversion of gDNA. Also, archaebacterial DNA polymerases are strongly inhibited by the presence of small amounts of uracil-containing DNA [29]

b) PCR Bias

Amplification efficiencies for bisulfite-converted templates obtained from methylated samples and unmethylated samples are known to be different [30]. Regions of interest are usually CpG islands so that a large number of C-G base pairs are present in an amplicon from a methylated sample, and correspond to T-G base pairs in a related unmethylated sample. The degree of amplification bias relative to a corresponding region rich in T-G base pairs varies with the characteristics of a specific amplicon. However, the template derived from the unmethylated strand is frequently amplified more efficiently than the template derived from the methylated strand, and will therefore dominate in a mixed sample, as shown in Figure 11.

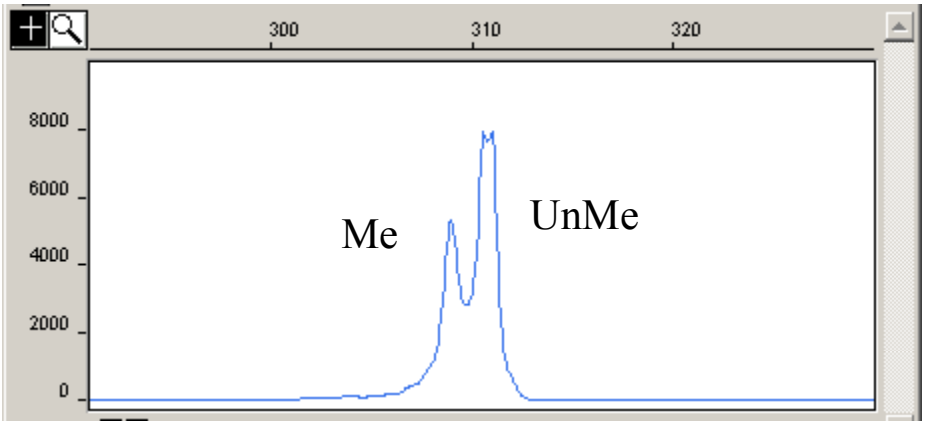


Figure 11. CE-based fragment analysis of the estrogen receptor (ER) gene in a 50/50 mixed methylated (Me)/unmethylated (UnMe) sample; size-standard not shown. Amplicons were generated using a dye-labeled (FAM) forward primer such that the methylated-derived amplicon, which has 26 CpGs, migrates faster than the amplicon from unmethylated gDNA, which instead has 26 TpGs. Note the bias in favor of the amplicon from unmethylated gDNA. The PCR conditions described in the protocols section significantly reduce the bias. Sequencing results for this mixture of amplicons are shown in Figure 17.

PCR Recommendations

Data presented in preceding examples were generated using bisulfite-sequencing primers with T_m ~55 °C and tailed with M13 forward and reverse sequences [31, 32]. The well-known 18-base-21M13 sequence, or other suitable sequence, provides a universal sequencing primer binding site. These universally tailed regions of locus-specific primers have all 4 bases present, and after a first round of PCR amplification, the longer primer binding site provides for higher T_m (and higher specificity). During thermal cycling, the annealing temperature can be raised after the first (few) cycle(s). This higher temperature PCR provides enhanced selectivity as evidenced by less background in resulting sequencing traces and fragment analysis. Reduced formation of primer dimers and/or secondary amplicons was observed in preliminary experiments using "touchdown" PCR [32] in combination with raising the annealing temperature as described above.

PCR reaction conditions generally include glycerol, a denaturant which reduces PCR bias during amplification of templates derived from bisulfite conversion of fully methylated and fully unmethylated sample. Betaine [33] also reduced such bias, but did so with less

consistency, relative to glycerol, among several gene targets that were investigated. A 451-bp amplicon [34], which previously exhibited significant bias in favor of the amplicon from unmethylated gDNA template, resulted in PCR amplicons of equal intensity from a 50/50 mixed sample using the PCR conditions provided in the protocols section. (See the "Fragment Analysis" section for resolving amplicons from methylated and unmethylated DNA). These results were obtained using M13 tailed primers, and increasing the annealing temperature after the first 5 cycles of PCR. A 3 minute extension time in the PCR thermal cycling may reduce PCR bias by permitting the polymerase to read through CG-rich regions. The 60 minutes at 72 degrees at the end of cycling permits the non-templated "A" addition-activity of the polymerase [35-37], which is important prior to fragment analysis, but not necessary if the amplicons are to be sequenced only.

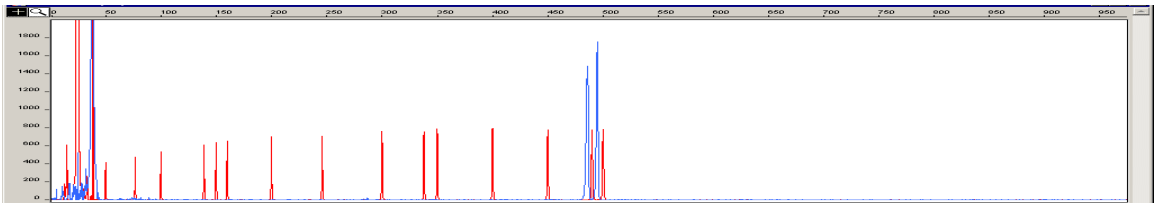


Figure 12. CE-based fragment analysis of amplicons (blue) [size standard (red)] of the RASSF gene using the presently reported protocol (protocols section) shows unbiased amplification (in contrast to our findings using an earlier protocol that does not include extra glycerol and BSA during the PCR [34]). The protocol did not eliminate bias for all regions investigated. See text under "Fragment Analysis" for resolving amplicons from methylated and unmethylated DNA.

The PCR anomalies such as amplification bias due to variable CG-content, slippage, primer dimer, competing amplification of secondary amplicons due to lower primer specificity, and incomplete bisulfite conversion militate against quantitative accuracy. One region may be more difficult to analyze relative to a nearby region that has fewer CpG's, or avoids homoploymer sequences. Accurate determination of methylation status may require analyses of multiple sites within a given region.

A summary of recommendations to reduce PCR bias when amplifying templates of mixed methylated states is provided below:

- Hot start (AmpliTaq Gold® DNA Polymerase)
- Tailed primers with all 4 bases in design
- PCR denaturant, such as glycerol
- Annealing temperature approx 2-5 degrees above calculated Tm (gene specific portion)
- Increased annealing temperature after the first few PCR cycles
- Touchdown PCR
- Increase extension time/temperature during PCR
- Decrease primer concentration (to reduce primer-dimer)

Elimination of PCR bias is reportedly [38] dependent on the amount of genomic bisulfite-converted DNA used in the PCR. Our recommended protocol using 3 ng of bisulfite converted gDNA per 5-uL reaction gives a concentration in excess of the minimum of 10 ng per 25-uL reaction reported [38] for reproducible amplification prior to pyrosequencing. In addition to further optimization of PCR conditions, researchers at Applied Biosystems are investigating alternative protocols for bisulfite sequencing that will eliminate PCR bias concerns.

.

III Capillary Electrophoresis Analysis.

a. Fragment Analysis

A newly developed and very simple workflow for methylation analysis after bisulfite conversion involves PCR using bisulfite-sequencing primers and CE analysis of the PCR amplicon (Figures 11,12 and 16) [34]. Formation of a correct-sized amplicon serves as proof of the presence of the intended target sequence.

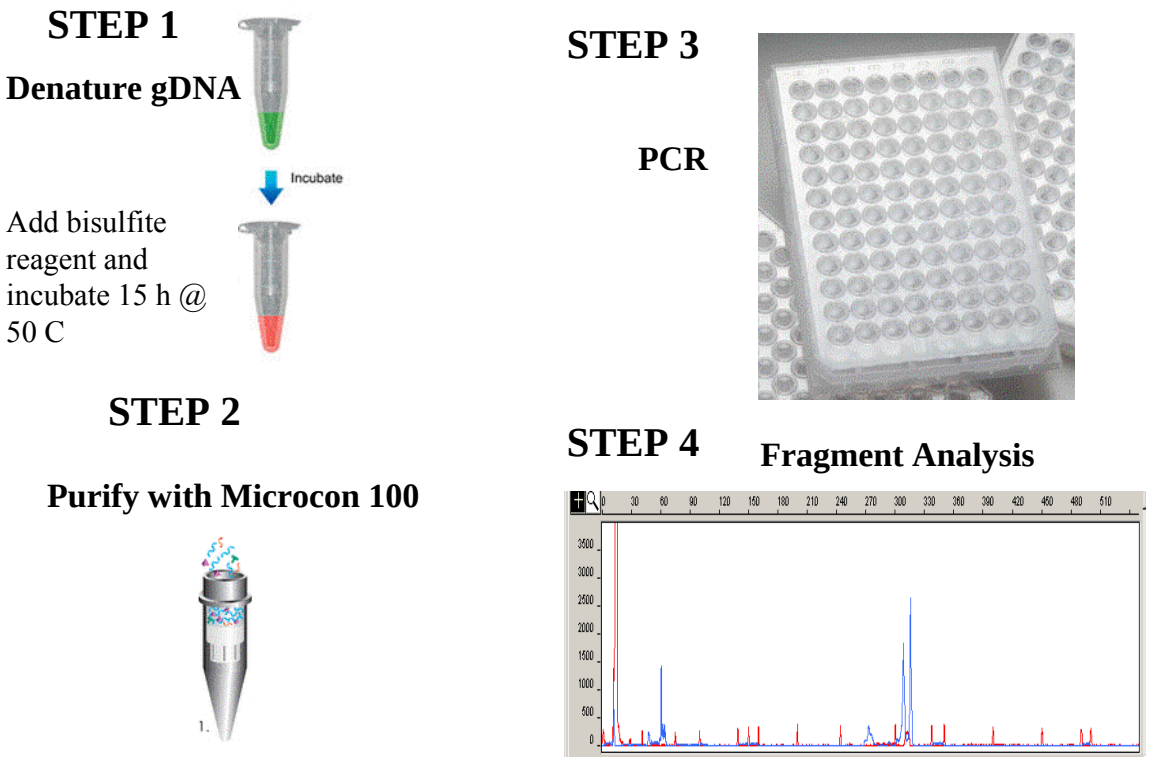


Figure 13. Workflow for methylation-dependent fragment separation. Step 1, bisulfite conversion of Cs to Us, except for methylated CpG's. Step 2, PCR amplification using a primer set designed in a non-CpG region to generate amplicons regardless of the methylation status. The forward primer has a fluorescent dye-label (FAM™). Step 3, CE separation of the faster migrating C-rich strand from slower migrating T-rich strand derived from methylated (Me) and unmethylated (UnMe) DNA, respectively. The same amplicon is also analyzed by direct sequencing.

When there are a relatively large number of CpGs in an amplicon (i.e., ~1 CpG dinucleotide per ~10-12 bases), CE can lead to separation of amplicons derived from fully methylated and fully unmethylated gDNA regions. Moreover, the amplicon used for this easy fragment-based method of methylation analysis can also be directly sequenced (Figure 17).

Amplicons from bisulfite-converted methylated and unmethylated gDNA are co-amplified with the same CpG-free primers using a fluorescent dye label (e.g., FAM) on one of the primers. A fluorescent label on the forward primer during PCR provides detection of a C or T at CpG sites, whereas the variable positions are G or A with a dye on the reverse strand. The amplicon with Cs (FAM™ dye on forward primer) will

migrate faster than the corresponding amplicon wherein all the Cs are Ts. The cumulative effect of a lower mass/charge for a C vs. a T was discovered to be large enough in CpG-rich sequences to permit near baseline resolution in POP-4™ CE polymer at 60 °C. Separation of amplicons from fully methylated or unmethylated gDNA was greater in POP-6™ or POP-7™ CE polymer. A mathematical algorithm [34] that predicts the extent of separation can be applied to pre-select amplicons that are candidates for methylation ratio determination by this novel and easy method of fragment analysis by CE.

Coefficient	Length (N) vs. size observed	Composition (A,G,T,C) vs. size observed
k	4.247	4.010
a	-	0.819
g	-	1.180
t	-	0.916
c	-	0.812
n	0.914	-
Standard deviation σ (nt.)	± 1.1355	± 0.936
R^2	0.988	0.993

$$size = k + a\,A + g\,G + t\,T + c\,C$$

Figure 14. The algorithm above can be used to predict amplicons that will separate based on methylation status for a fully methylated and fully unmethylated region. A simple rule of thumb from the algorithm is that there is an apparent 0.1 nt difference per C vs T substitution in any given DNA sequence.

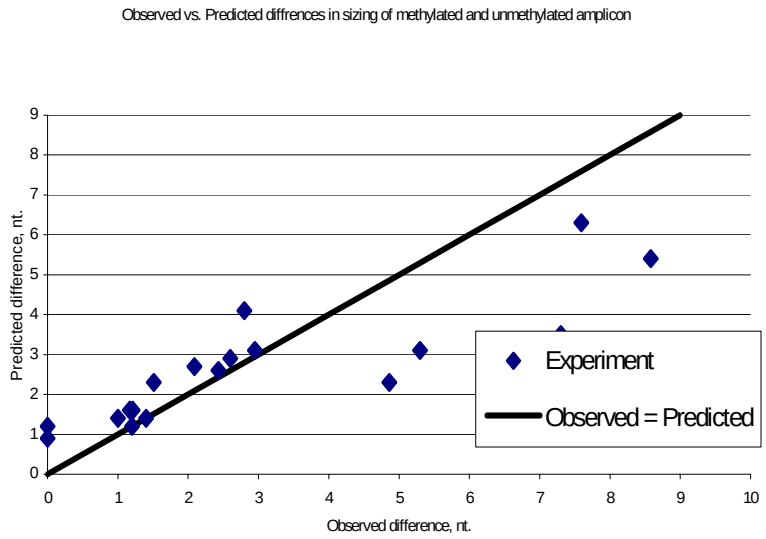


Figure 15. The predicted size (shown as a line) and the experimental size (shown as data points) is displayed above for amplicons of varying size and C vs. T content, generated from previously published data [34]. Amplicons less than ~300 bp can be predicted more accurately than the larger amplicons; the larger amplicons have competing secondary structure not fully removed under the denaturing conditions of the CE experiment.

Limitations of this CE method for fragment analysis to determine methylation ratios include:

1- Analysis of amplicons with >9 Ts. Homopolymeric runs of Ts lead to polymerase-related "slippage" phenomena that are seen as multiple sizes or a single broad signal, depending on CE resolution, analogous to what is known from PCR as short tandem repeat sequences [39].

2- Amplicons need to be in a CpG rich region.

3- Amplicons must be greater than 200 bp.

4- Analysis of regions that have highly variable methylation states at individual CpGs

An important additional benefit of fragment analysis is that it will reveal the presence of primer dimer [40] and/or secondary amplicons that do not match the anticipated size. Real-time PCR methods using SYBR® green dye detection or a method that selects for only a targeted sequence with a hybridization probe, as in MethyLight [41], do not detect such competing side reactions. Real-time PCR side reactions contribute to the measured SYBR green dye-derived Ct values. Post-PCR dissociation curves sometimes discriminate for the presence of secondary amplicons, but cannot serve to correct Ct values.

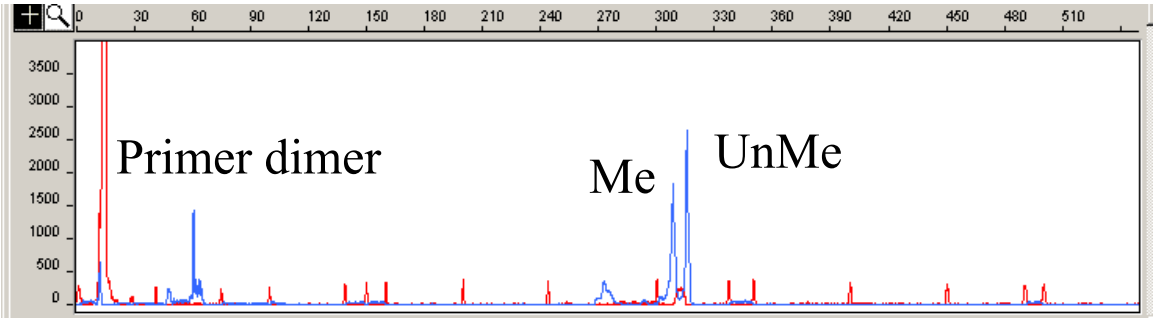


Figure 16. CE fragment-size trace of ~300-bp FAM-labeled amplicons (blue) derived from the p15 gene locus in human gDNA following bisulfite-conversion (size standard in red). Amplicons of the same length but different base composition were derived from methylated (Me) and unmethylated (UnMe) gDNA using the same primer set. Due to sequence composition differences (C vs. T) the p15 amplicon from methylated gDNA migrates faster than the amplicon from unmethylated gDNA. Primer dimer is seen at ~60 bp. The variable presence of primer dimer competes with the yield of the targeted sequence during PCR and can lower the apparent quantitation. CE permits detection of all fragments from a PCR reaction which can provide valuable clues when troubleshooting quantitation and sequence analysis.

b. Bisulfite Sequencing

Bisulfite sequencing is the “gold standard” for the analysis of methylation and is used for both discovery and routine analysis. Essentially all DNA methylation researchers depend

either directly or indirectly on bisulfite-sequencing data obtained for their region(s) of interest. Sequencing primers with or without universal tails are generally designed to anneal to non-CpG regions flanking each region of interest, and thus amplify bisulfite-converted gDNA regardless of methylation status. Consequently, if one of the primers is fluorescently labeled at the 5' end, the resultant amplicon can be analyzed by the CE-based fragment analysis method described above and also used for sequencing.

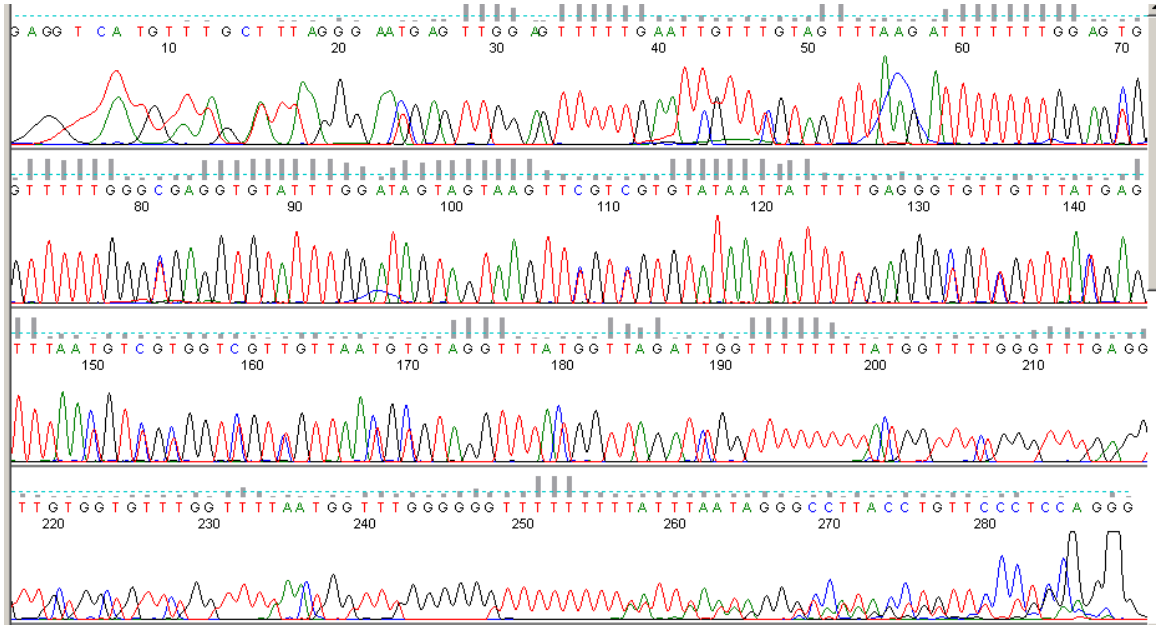


Figure 17. CE sequencing results for ER amplicons analyzed by fragment analysis as described in figure 11. A mixed C/T signal is seen at all 5mCpG/CpG sites due to the PCR template being a 50/50 mixture derived from methylated and unmethylated gDNA. Cs not adjacent to Gs at the end of the sequence are present in the M13-primer tail. Full-length FAM-labeled amplicons from fully methylated and unmethylated gDNA are seen at the end of the sequence run as two relatively large G signals.

The amplicon can be either directly sequenced, or is cloned and then sequenced. Bisulfite sequencing of clonal regions of interest provides unambiguous data on methylation patterns. Bisulfite sequencing also provides information regarding possible incomplete

bisulfite conversion as well as detection of mutations or single nucleotide polymorphisms (SNPs).

GTGGGCGGAGGGACTGGGGTTCTTCTCCCGACACCACCTTTCCGCCACCACCTCCAAGTCCTG
AGAATGTCTCACTGGACGACGAGTTGCTCTTTGGTTGGGACAGGTGAAGGGAGGAGCGCGGT
TCTTTCTGAGGCCAAGGAAGAAACGGGTACCTACCTTGTCGCTTCCCATGGGGGGAGGGAGG
CTGATGATGAGTG

Figure 18. Example of a section of a CG-rich region which will have long stretches of T's after bisulfite conversion. There is lost specificity in primer design leading potentially to more than one amplification product and the poly T stretches will cause polymerase slippage during PCR amplification. Amplification regardless of methylation status requires primers that do not anneal to a CpG site, which limits primer design choices.

Sequencing reveals the extent of bisulfite conversion. If incomplete, this will lead to the appearance of Cs that are not adjacent to the 5' side of Gs. For this reason, sequencing provides its own "internal reference standard" for completeness of bisulfite conversion. Sequencing primers are designed to anneal to the bisulfite-converted sequence and thus select for fully bisulfite-converted gDNA template. Some regions, however, within an amplicon may be difficult to bisulfite-convert to completion, as shown by the example below.

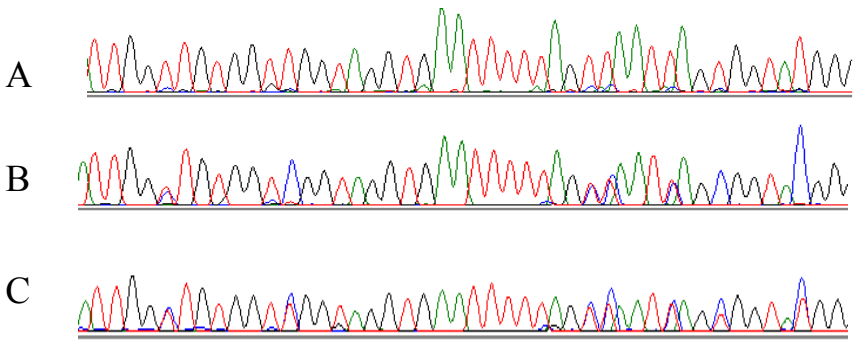


Figure 19. The ability to drive the bisulfite conversion to completion may depend on the purity of the gDNA [11]. gDNA from three different sources (and different methylation states) are shown here: A. DNA isolated from Leukocytes B. DNA isolated from RKO cell line C. A mix of gDNA from Coriell and a universal methylated gDNA from Serologicals. Sample A is a non-methylated sample and shows complete bisulfite conversion (no Cs, not even CpGs) , whereas samples B and C have both CpGs and incompletely bisulfite converted C/T mix signals at non CpGs.

Direct sequencing of a PCR amplicon derived from bisulfite-converted gDNA may lead to observation of superimposed signals due to contamination by secondary, co-amplified sequence(s) and/or primer dimer sequence(s). The presence of these types of shorter amplicons can contribute to off-scale signals at the beginning of the sequence trace, and depletion of the sequencing reactants, which causes rapid drop-off of signal intensity for the remaining sequence.

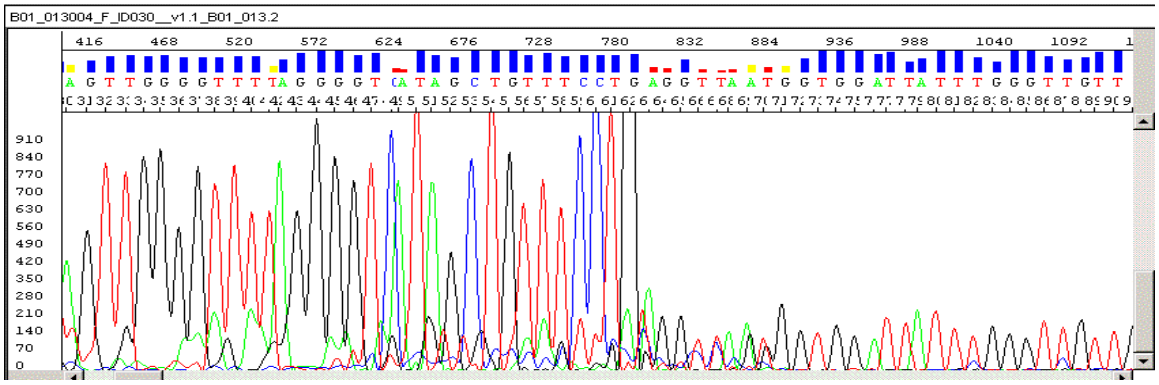


Figure 20. Primer-dimer signals will oftentimes dominate the initial sequence trace, as seen here for bases 30-62, followed by much lower intensity signals due to the amplicon of interest.

Consequently, sequence-callable extension products may not be obtained even if the amplicon of interest is present. Such problems can be addressed by use of hot start PCR methods and/or application of "nested PCR" [42] protocols prior to sequencing to enrich for the amplicon of interest. Nested PCR, which requires a second set of internal PCR primers, is widely used in other applications to select for the desired amplicon.

As for all sequencing reactions, an estimate of the amount of the PCR amplicon added to the sequencing reaction is needed to prevent off-scale signals. For subsequent clean-up of the sequencing reaction using the recently introduced BigDye® XTerminator™ Purification Kit, ½ to 1/5th the amount of amplicon commonly used in a sequencing reaction is recommended to prevent off-scale signals during analysis.

M13-primer-tailed amplicons obtained from a bisulfite-PCR reaction are sequenced using the standard BigDye® Terminator v1.1 kit protocol. BigDye® Terminator v1.1 kit is recommended because it provides better signal resolution at the beginning of the sequence, although BigDye® Terminator v3.1 kit can be used. The sequence of the entire amplicon can be seen when using M13 tailed primers, and the universal tail simplifies the sequencing workflow when processing many samples. BigDye® Terminator v1.1 kit and a 2-temperature thermal cycling profile is used in place of a 3-temperature profile to maintain the level of signal strength.

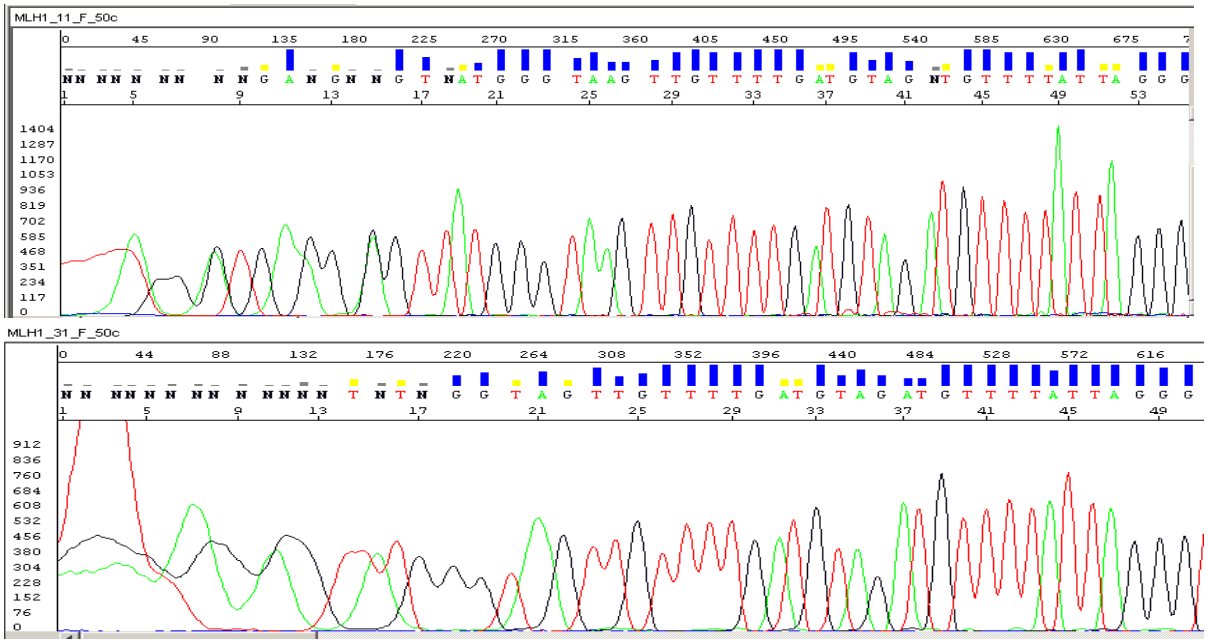


Figure 21. Comparison of the same amplicon (unmethylated MLH1) sequenced with BigDye® Terminator v1.1 kit (top) and with BigDye® Terminator v3.1 kit (bottom). V1.1 kit chemistry provides better resolution at the beginning of the sequence run. However, when using M13-tailed, the lack of resolution at the beginning of a sequence

run does not interfere with sequence analysis of the amplicon, so that either chemistry is suitable for bisulfite sequencing analysis.

Characteristics of the Bisulfite-Sequencing Trace

Direct sequencing will detect mixed bases at CpG sites when obtained from samples with mixed methylation states. Control studies where the amplicon from bisulfite-converted methylated and unmethylated gDNA is mixed in known ratios indicated that accurate mixed-base signals were obtained if the minor component is ≥15%. Below 15% the minor-component signal may either be not detected or have exaggerated intensity in the analyzed data.

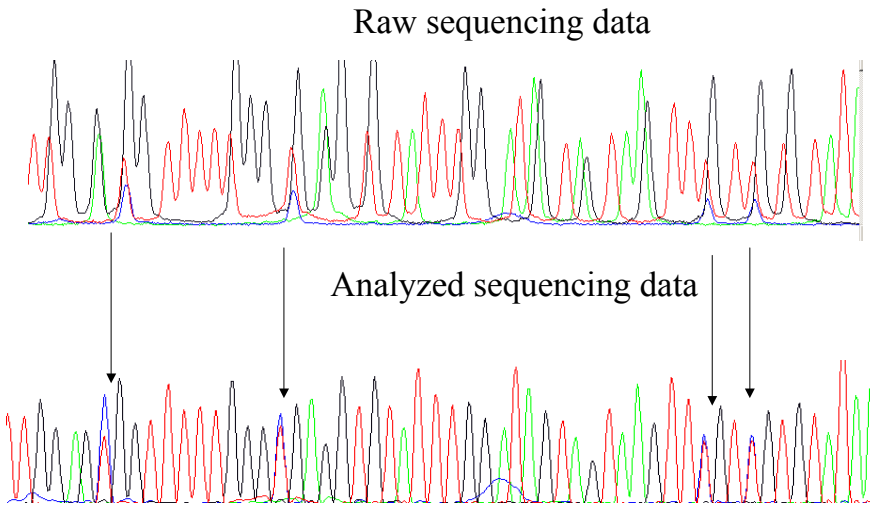


Figure 22. In bisulfite sequencing of amplicons derived from mixed methylated and unmethylated gDNA, unmethylated Cs are absent due to conversion to Ts via Us throughout the amplicon, while methylated Cs are detected as mixed-base signals of C (blue) and T (red). A section of a sequencing trace is presented here comparing the raw signal (top) to the analyzed (i.e., software-processed) trace (bottom). The software normalizes the data so that each of the four bases has approximately equal signal strength, which therefore artificially distorts the actual C/T ratio that is better (but not exactly, due to fluorescence differences) represented by the raw data.

Older Applied Biosystems sequencing platforms (i.e. 3700 system) have the less current analysis software and may analyze by over-normalizing the less represented (C) base. Applied Biosystems KB™ Basecaller Software adjusts for sequences with skewed base

composition in such a way that the underrepresented color is not artificially overly exaggerated.

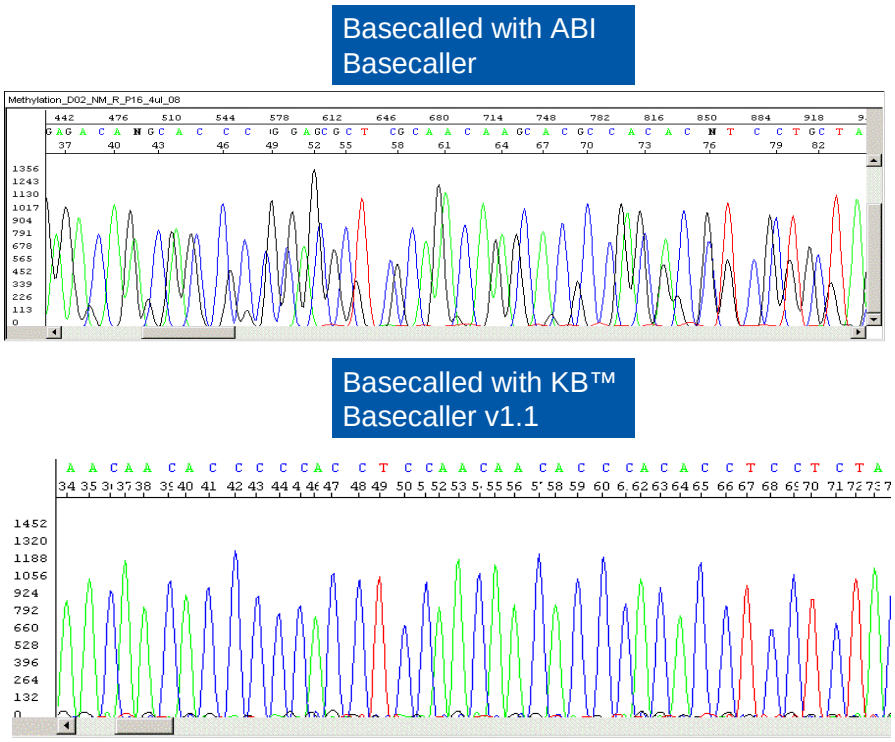
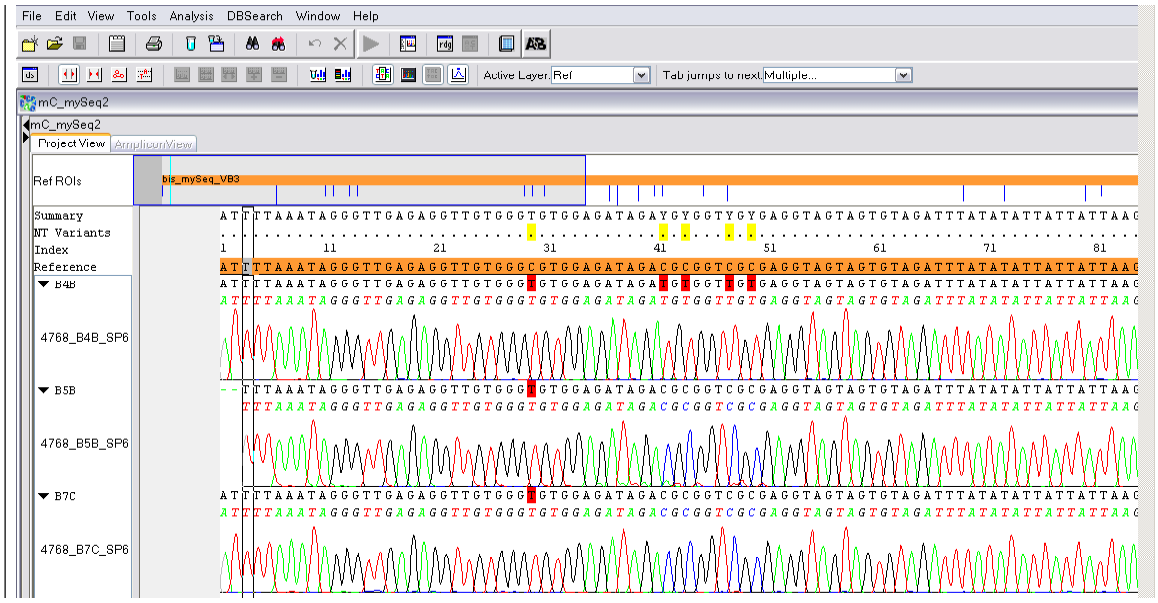


Figure 23. The same solution of a sequencing reaction of a bisulfite-converted sequence of a fully unmethylated sample was analyzed by both the AB basecaller and the KB™ basecaller. The older Applied Biosystems software exaggerated the missing color, but analysis with KB™ basecaller did not overly normalize the missing base (G in the reverse strand sequencing shown)

Software-related factors also account for the fact that amplicon sequences produced using universally tailed primers, or cloned sequences having sequence-content from the cloning vector, usually lead to more level sequence traces without background signal problems. The FAM dye on the full length amplicon-template from PCR is seen at the end of a sequence trace as a large, black colored “G” signal (Figure 17). The FAM dye adds the

missing black colored “G” when sequencing the reverse strand, which fortuitously aids normalization of the signal strength for actual G bases.



SeqScape 2.5.0 Genotyping Report

Applied Biosystems

Ref1

Page 2

Specimen	C / 29 / bis_mySEQ	C / 41 / bis_mySEQ	C / 43 / bis_mySEQ	C / 47 / bis_mySEQ	C / 49 / bis_mySEQ	C / 54 / bis_mySEQ	C / 90 / bis_mySEQ	C / 92 / bis_mySEQ	C / 95 / bis_mySEQ	C / 111 / bis_mySEQ	C / 113 / bis_mySEQ	C / 118 / bis_mySEQ	C / 122 / bis_mySEQ	C / 124 / bis_mySEQ	C / 134 / bis_mySEQ	C / 140 / bis_mySEQ	T / 196 / bis_mySEQ	C / 205 / bis_mySEQ	C / 224 / bis_mySEQ	C / 228 / bis_mySEQ
B4B	T (37)	T (37)	T (36)	T (36)	T (35)	T (36)	T (37)	T (37)	T (36)	T (37)	T (37)	T (36)	T (37)	T (37)	T (37)	T (37)	T (33)	T (37)	T (35)	T (32)
B5B	T (37)	C (29)	C (33)	C (32)	C (35)	T (37)	C (33)	C (32)	C (37)	T (41)	T (36)	T (32)	C (29)	C (36)	C (36)	T (37)	T (35)	T (36)	T (36)	C (41)
B7C	T (37)	C (36)	C (29)	C (37)	C (41)	T (37)	C (35)	C (36)	C (41)	C (37)	T (36)	C (36)	C (36)	C (33)	C (37)	C (36)	C (35)	C (36)	C (37)	C (41)

Figure 24. SeqScape® software Analysis can be used to present the CpG methylation status of specific Cs in several clones, aligned as shown above. Cloning and sequencing permits analysis of methylation patterns and can be used for quantification. Both PCR and cloning bias may cause misrepresentation of the actual methylation states. However, the cloned amplicon will have pure signals, i.e. no mixed bases or interference from secondary amplicons.

Misalignment of signals is a problem encountered when sequencing longer bisulfite-converted amplicons (>300 bp). There are two sources for this misalignment, which is seen more frequently in bisulfite sequencing than in sequencing in general. PCR slippage is common, and is typically seen when there are >9 sequential Ts.

Sequencing is unambiguous up to the point of slippage, and is sometimes callable after the misalignment occurs. A second cause of misaligned signals that appears more gradually is related to the large number of C vs. T (or G vs. A) sites in the bisulfite-converted amplicon for samples that have a mixed methylation status. At each mixed-base site, the difference in CE migration increases for an amplicon derived from unmethylated gDNA vs. methylated gDNA; consequently, longer sequencing fragments no longer co-migrate.

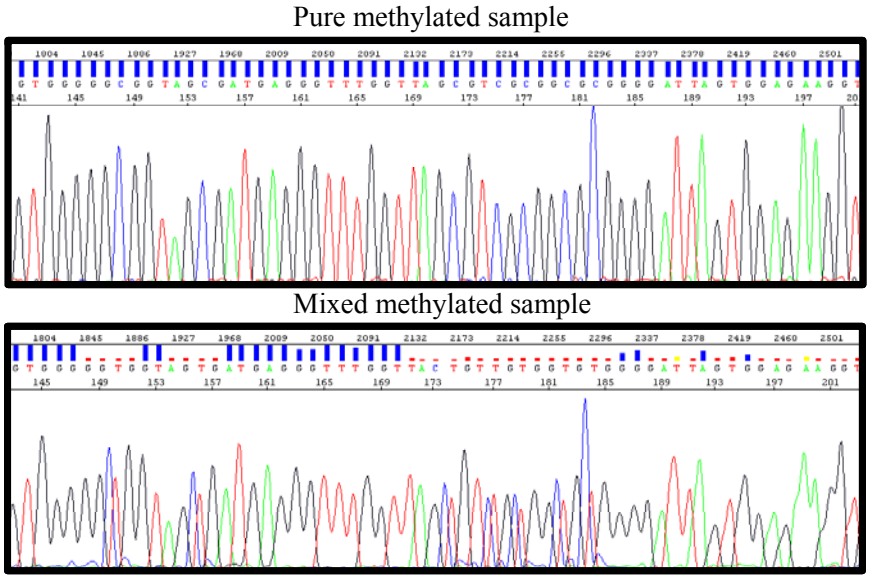


Figure 25. Bisulfite sequencing traces for an amplicon derived from a pure gDNA template (top) and a template that is a 50/50 mixture of methylated and unmethylated states (bottom). The sequencing products derived from the methylated and unmethylated sample do not co-migrate because of multiple C vs. T differences in the sequence (see text).

Migration differences is seen in samples of mixed methylation states and has been previously reported [26]. Signals gradually broaden, and then split, and for longer amplicons, will appear as an N-1 sequence superimposed on the signals of the more predominant amplicon (Figures 25 and 32). Sequencing in both the forward and reverse

direction permits sequence reads up to the point of the misalignment(s). Shortened PCR amplicons that avoid homopolymer sequences are generally purer templates for sequencing. PCR bias is reduced for shorter amplicons, and the base composition, proportionately, is better represented by all 4 bases due to the tailed sequence representing a greater percentage of the amplicon length.

Based on the foregoing discussion, troubleshooting can be complicated due to multiple factors that may underlie poor sequencing results. Each step in the bisulfite sequencing workflow (bisulfite conversion, PCR and sequencing) may contribute to poor sequencing data. Troubleshooting typically requires optimization of more than just one parameter.

The following is a list of recommendations for direct sequencing of amplicons derived from bisulfite-converted gDNA:

- Quantification of PCR amplicons prior to sequencing
- Use of M13-tailed primers
- Use of full-strength BigDye® Terminator v1.1 Ready Reaction mix
- 2-temperature cycle-sequencing
- Use of XTerminator solution clean-up of sequencing reactions
- Analysis with KB™ basecaller software

Cloning and Sequencing

Regardless of the technique used, quantitative PCR applied to bisulfite-converted gDNA is potentially unreliable due to loss of specificity of primers and probes. Background signals are often unavoidable for oligonucleotides designed to anneal to bisulfite-converted non-CpG or unmethylated sequences. Primers and probes designed to anneal to bisulfite-converted CpG-rich regions, and which anneal over several of such positions, can have 4 bases, if methylated, and thus exhibit greater specificity, so that quantification of the methylated sample has a higher probability of being more accurate vs.

unmethylated sample. There are several other limitations to accuracy of methods intended to provide quantification of methylation:

- Purity of the sample extracted from biological sources
- Completeness of bisulfite conversion
- PCR bias, which is a variable for each sample, and each analysis

Additionally, for cloning and sequencing:

- Cloning bias
- Limited number of clones sequenced

The accuracy of quantification by cloning and sequencing is limited by potential bias in the PCR amplification process as well as the cloning process. Relative to direct sequencing of amplicons from bisulfite-converted gDNA templates, sequencing of clones provides much "cleaner" sequencing due to a single amplicon insert per clone; there are no secondary sequences, no PCR slippage, no mixed bases, misaligned sequences due to mobility differences are eliminated, and all 4 bases are represented for signal normalization due to sequence content from the cloning vector. The improved appearance of the sequencing data permits semi-quantitation and determination of variable methylation patterns. However, the workflow is time-consuming, and often relatively few clones are sequenced—well below the number needed for statistical accuracy. PCR (and cloning bias) may still contribute to distortion of the data when attempting methylation quantification. Lack of variability seen in the clones indicates a low copy number of template going into the PCR, or PCR bias during amplification. Any sequence analysis method used, including pyrosequencing [38], will be subject to all the sample handling biases introduced during the PCR and cloning steps.

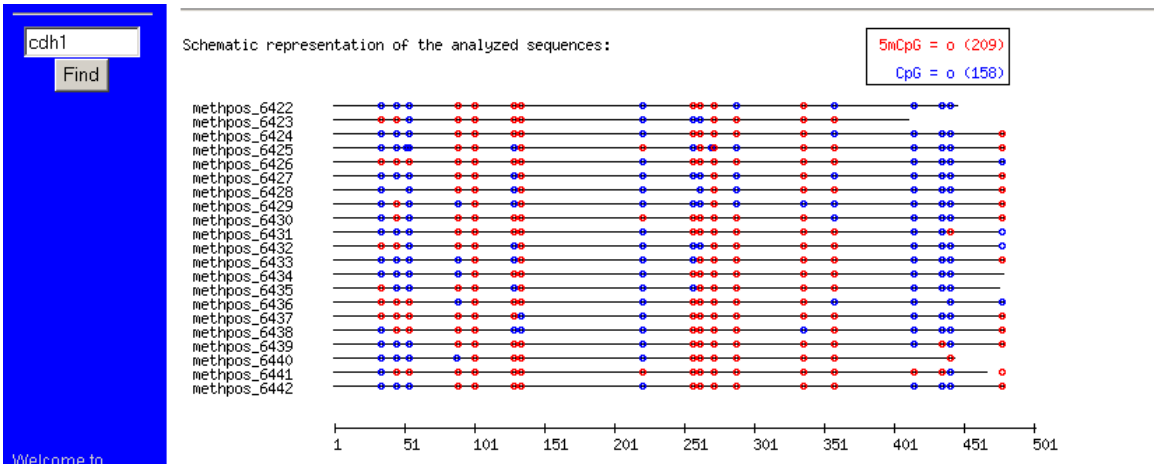


Figure 26. Bisulfite cloning and sequencing permits assessment of methylation at individual sites. The bar graph represents the relative position and methylation states of the individual CpGs (red = mCpG, blue = CpG) in a sequence. Analyzing many clones provides methylation patterns in a sample. (Data was copy and pasted from the MethDB website: http://www.methdb.net/)

Bisulfite Sequencing Clean-Up

Complete removal of fluorescently labeled dideoxy terminators, dNTPs and salts is needed prior to CE analysis of the sequencing extension products. Bisulfite sequencing reactions can be purified using Centri-Sep™ columns (96-well plate) or ethanol precipitation. Both these purification methods, which require careful attention to detail by laboratory personnel, encounter some loss of sample during processing. Gel purification (Centri-Sep) requires careful pipeting into the center of the gel-bed. A new product and protocol from Applied Biosystems, BigDye® XTerminator™ Purification Kit, offer a simpler alternative that reduces sample loss. After the sequencing reaction is carried out, a slurry of absorptive, biphasic material is added directly to each sequencing well, which are then sealed and vortexed for 30 minutes. After centrifugation to settle particulates to the bottom of the wells, the plate is directly used with a Model 3730 or 3130 instrument for sequencing. A run module must first be downloaded to adjust the Z-axis of the autosampler to permit direct electrokinetic injection from the top of each well. The resulting sequence is free of so-called “dye-blobs” and provides much stronger signal

than alternative purification protocols, due to significant desalting of the sample. Purification using BigDye® XTerminator™ Purification Kits is both time-saving and very cost-effective.

IV Summary of Workflow for Methylation analysis

- Bisulfite conversion with methylSEQr™ Bisulfite Conversion Kit (P/N 4379580)
- User selected gene, sequence 500-bp +/- transcription start site
- Methyl Primer Express® software for selection of 55 ºC bisulfite (PCR) sequencing primers
- PCR using M13-tailed primers, (and FAM-dye on the Forward primer), as per the conditions described below
- Optional: fragment analysis of the PCR amplicon to obtain the ratio of amplicons derived from bisulfite-converted methylated and unmethylated gDNA.
- ExoSap-IT® removal of PCR primers and dNTP
- Direct sequencing of amplicon using BigDye® Terminator v1.1 kit, conditions described below
- Sequencing clean-up with BigDye® XTerminator™ Purification Kit
- Sequencing using Model 3730 or 3130 systems and analysis with KB™ Basecaller software

V First Time User

After bisulfite conversion with the methylSEQr kit, PCR analysis is often the first readout to evaluate the success of the bisulfite conversion. For a first experiment, select gDNA of known purity and quantity, not exceeding 300 ng for a 150 uL bisulfite conversion reaction. Follow instructions carefully for the denaturation step and addition of the bisulfite reagent to the freshly denatured gDNA. The most common sources of poor bisulfite conversion are from insufficient denaturation due to excess gDNA concentration or poor sample purity and possible renaturation of the freshly denatured gDNA. After ~15 hour (overnight) at 50^{o}C, the bisulfite is removed by centrifugal filtration. In situ

desulfonation (while the gDNA is contained in the filtration devise) ensures complete desulfonation without loss of DNA. The purified, stable solution of bisulfite-converted gDNA is immediately ready as a template for PCR. Select tried and true primer sets that produce an amplicon of a known size and sequence from literature reports or extracted from online resources (http://www.methdb.net). Recommended PCR conditions and a primer set used in testing of the bisulfite kit are provided in the Protocols Section. One of the X-chromosomes is silenced in female genomic DNA by methylation. A primer set that amplifies an X-chromosome gene can serve as a 50/50 methylated/non-methylated "control". Our reported investigations using fragment analysis [34] on the FMR1 gene region in a normal female control (non Fragile X individuals) provides an example of bisulfite conversion, purification and PCR amplification of the methylated and unmethylated strands of the X chromosomes. Due to the large number of Cs (methylated strand) and Ts (unmethylated strand) in the sequence of the amplicon from methylated or non-methylated X-chromosome template, the amplicons separate during capillary electrophoresis. (See section describing fragment analysis).

M13 tailed FMR1 primers used in PCR:

FMR1fwdM13
GTGTAAAACGACGGCCAGTTGAGTGTATTTTTGTAGAAATGGG

FMR1RevM13
GCAGGAAACAGCTATGACCTCTCTCTTCAAATAACCTAAAAAC M

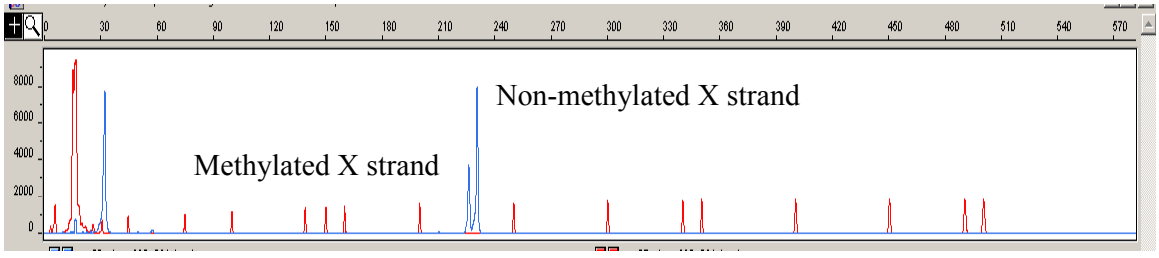


Figure 27. The success of the methylation analysis process (purified gDNA, methylSEQr bisulfite conversion, PCR with a verified primer set and analysis by capillary electrophoresis) is displayed in the electropherogram above. Targeting an X-chromosome gene provides an automatic internal 50/50 control for methylation/non-methylation. PCR bias will often preferentially amplify the non-methylated strand over the methylated strand, as seen in the example above.

Additional troubleshooting examples

There are multiple sources for the appearance of poor sequencing data following bisulfite conversion. It is important to systematically determine which one, or ones, are responsible for the poor data.

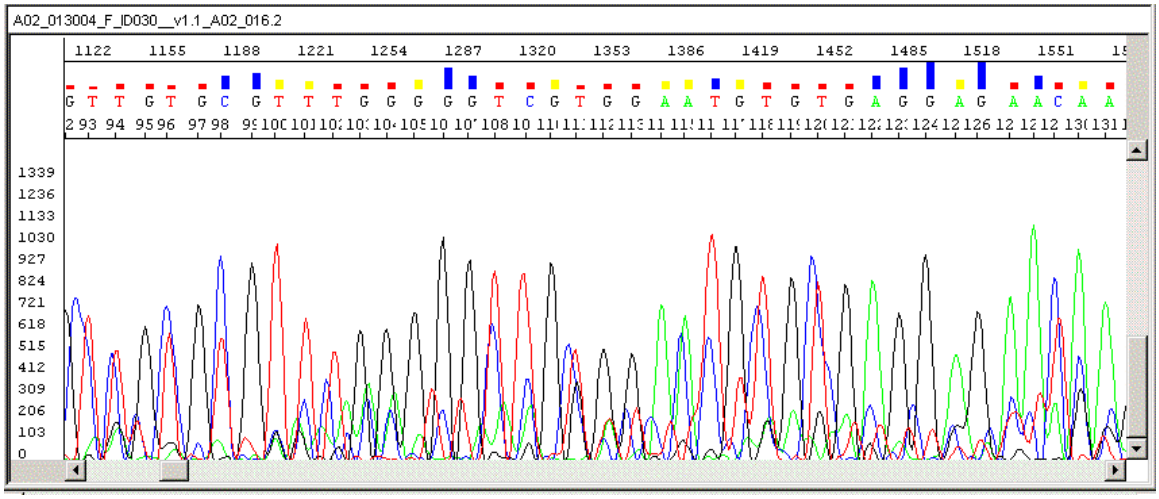


Figure 28. A typical first attempt at bisulfite sequencing often results in data as shown above. There may be multiple reasons for this poor sequencing result.

Is the amplicon pure?

Fragment analysis (using CE if a FAM-labeled primer is used) will reveal the purity of the amplicon

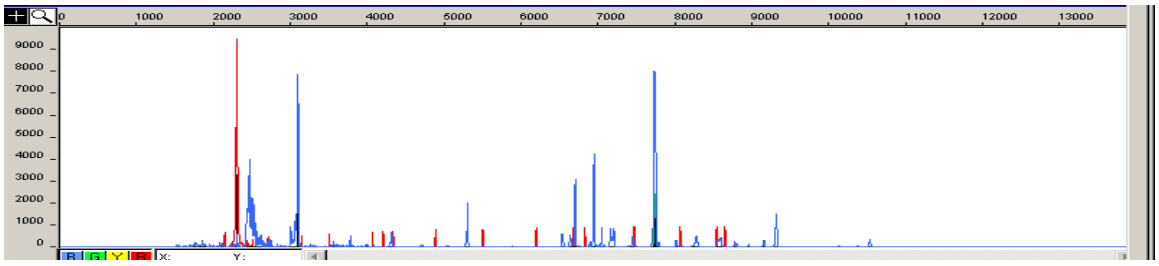


Figure 29. There are multiple PCR products present due to mis-matched amplification and primer dimer in this sample. The sequencing of this sample below shows that in addition to multiple products, incomplete bisulfite conversion was also present. PCR of other regions of the same bisulfite-converted gDNA sample amplified a single product.

Incomplete bisulfite conversion may affect only some regions, and not others. To improve bisulfite sequencing, a purer genomic DNA sample prior to bisulfite PCR may help, and/or selection of a different amplicon, which requires a redesign of the primers.

Was bisulfite-conversion complete?

```
TATGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
||||:||||+++||||:|:|++::||||||:::||:|:||||||||:
TATGTTGGGCGCGGTGGTTTACGTTTGTAATTTTAGTATTTTGGGAGGT
```

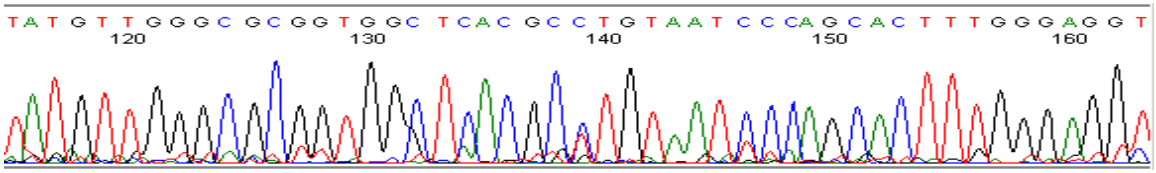
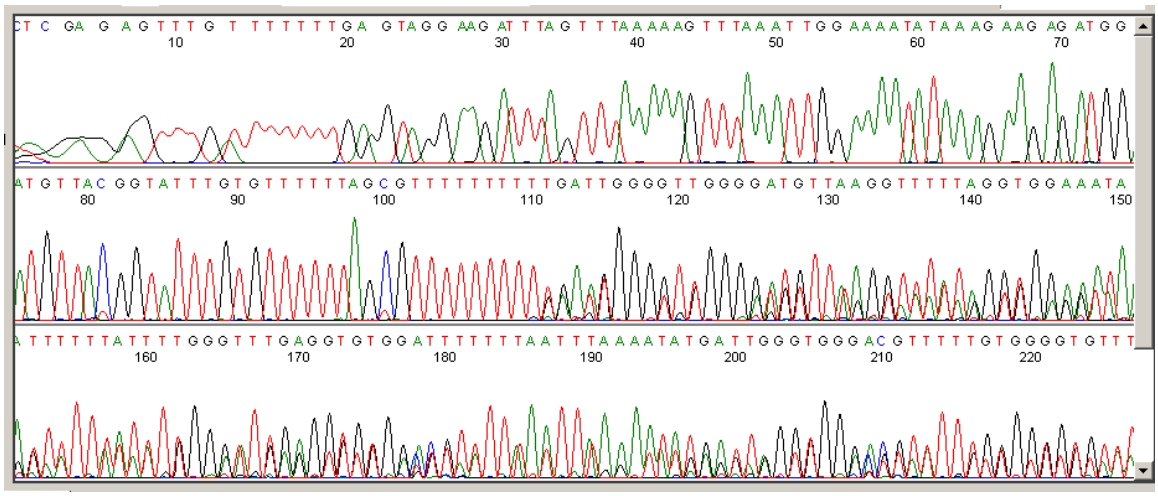


Figure 30. The presence of Cs (blue) not adjacent to Gs (black) is diagnostic for incomplete bisulfite conversion (i.e., non-CpG's) and therefore are not due to methylation of Cs. A C at a non-CpG position therefore serves as an internal control for complete bisulfite conversion. Incomplete bisulfite conversion may be due to the lack of purity of the gDNA, too much gDNA used in the bisulfite conversion, or inadequate denaturation of gDNA prior to the bisulfite conversion (see section on the bisulfite conversion). Primers designed to anneal to regions that are inherently difficult to bisulfite convert (e.g., due to strong secondary structure or DNA sequence still associated with protein) can provide incomplete, inaccurate, or misleading data.
The DNA sequence (SRBC AF408198), showing both the genomic input sequence (top) and the bisulfite converted sequence (bottom) was obtained from the free online resource: http://www.urogene.org/methprimer/

PCR slippage



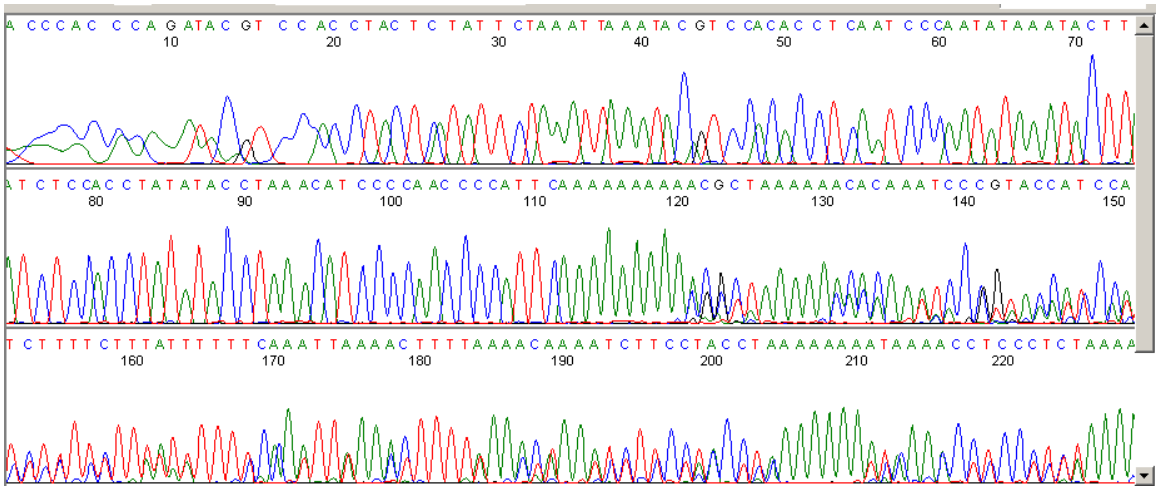


Figure 31 A&B. A 250 bp amplicon with a homopolymer stretch exceeding >9 bases resulting in slippage during PCR amplification. In the forward sequencing trace, the onset of slippage occurs after 10 Ts seen between 100-110 nt and in the reverse (bottom) sequence, slippage occurs after the corresponding 10 As between 110-120. By sequencing in both directions the sequence (and methylation status) of the entire sample can determined.

The bisulfite conversion results in DNA sequence that is 50% T (or A) resulting in a greatly increased probability of homopolymer tracts of T (forward) or A (reverse).

Misaligned sequence due to variable methylation states

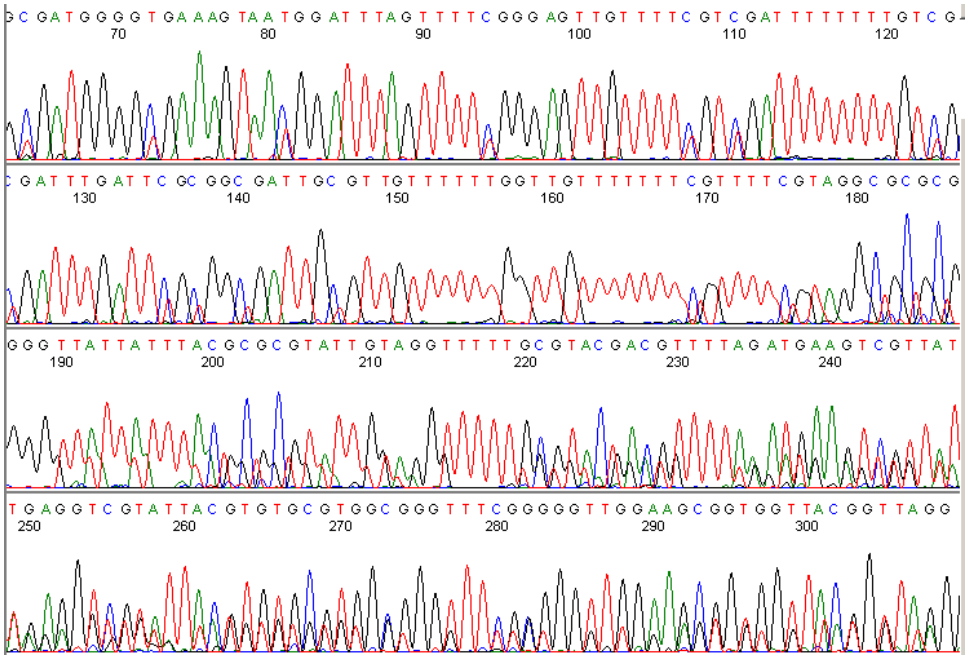


Figure 32. Sequencing of an amplicon from PCR of a 50/50 mix of a fully methylated bisulfite-converted gDNA and a fully unmethylated sample, leading to misaligned signals. PCR with bisulfite-sequencing primers (designed to amplify regardless of methylation) simultaneously amplifies all methylation states. Initially the signals from fully methylated and fully unmethylated align, but the longer extension fragments have cumulatively more C vs. T differences, and begin to migrate at different rates. Peaks are observed to mis-align, then broaden and eventually split, appearing as an N + 1 sequence.

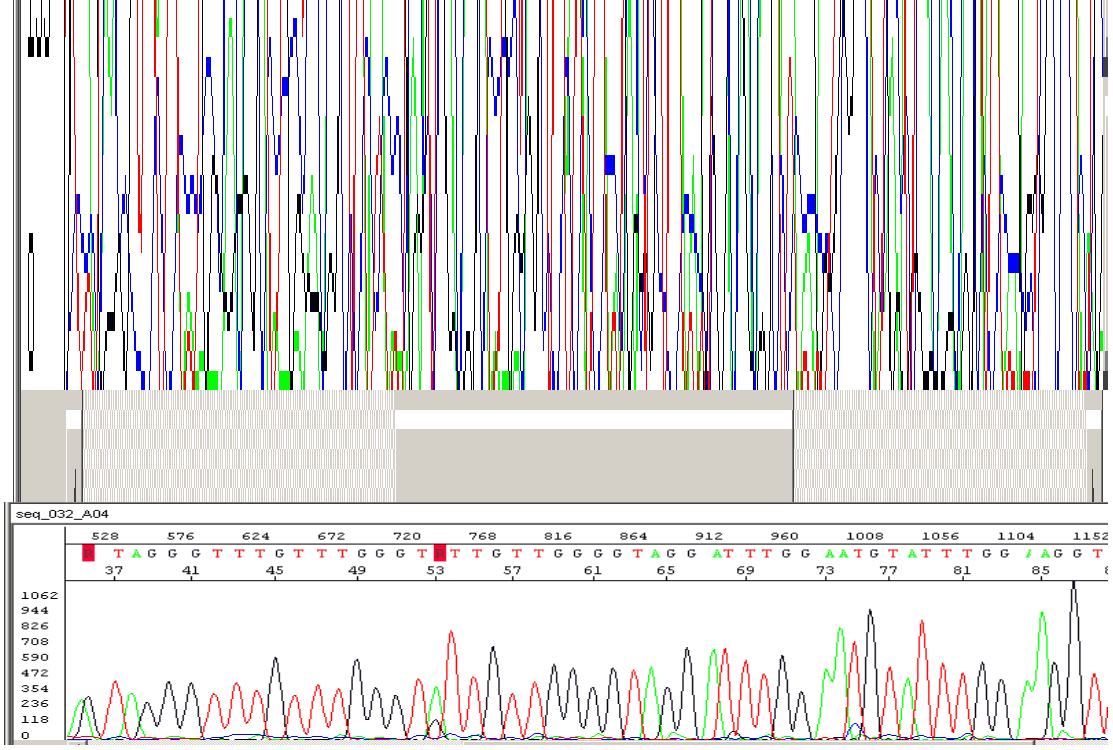


Figure 33. Offscale signals due to too much PCR amplicon-template used in the sequencing may be misinterpreted as a problematic sequence run. The same amplicon was significantly diluted and re-sequenced resulting in a correct and interpretable sequence. (FFPE tissue, rarb gene)

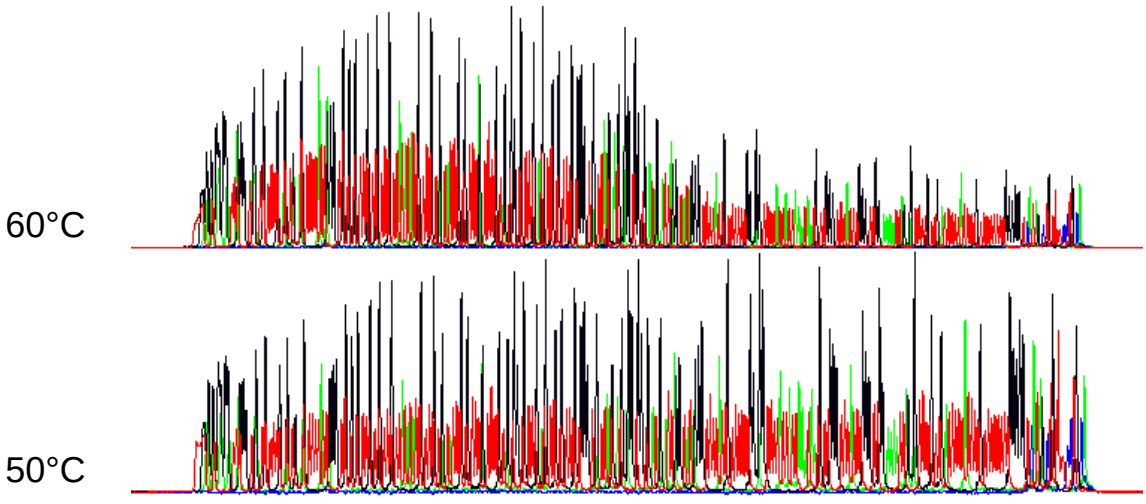


Figure 34. Four-color "raw" (i.e., unprocessed) sequencing data for a “standard” 3-temperature cycle sequencing (top), which has a 4-minute extension time at 60 °C, was replaced with 2-step cycle sequencing (bottom) that has a 4-minute extension time at 50 °C. The lower temperature provides better “signal balance” when applied to a mix of amplicons derived from bisulfite-converted methylated and unmethylated gDNA.

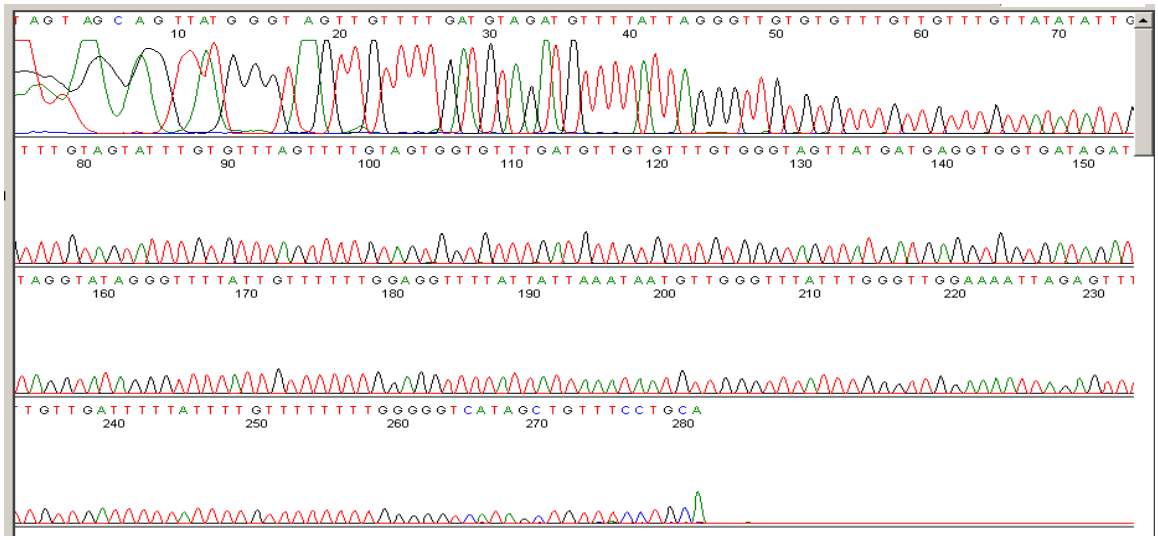


Figure 35. Loss of signal strength, presented in the raw data (60°C) in the preceding figure, results in reduced signal strength in the analyzed sequencing trace. Lost signal strength is greater for an amplicon from fully unmethylated gDNA than the corresponding fully methylated amplicon. The loss in signal can be significant enough to prevent a full sequence read. The recommended 2-step cycle sequencing program with a 50°C anneal/extension during sequencing greatly reduces the signal strength loss.

VII Protocol for Bisulfite Sequencing

Bisulfite Conversion

Bisulfite conversions were performed using the methylSEQr™ Bisulfite Conversion Kit (Applied Biosystems) according to manufacturers directions.

PCR

PCR conditions described below were optimized for bisulfite PCR using M13-tailed primers with a 55 ºC T_m for the gene specific portion of the primer. When analyzing by fragment analysis, a dye-label (FAM) is included on the 5' end of one of the primers.

-21 M13

Forward Primer

5´TGTAAAACGACGGCCAGT 3´

M13 Reverse

Primer

5′CAGGAAACAGCTATGACC 3′

Recommended "Tried and True" amplicon/primer sequences

CDH1 L34545

Fwd TGTAAAACGACGGCCAGTTTTAGTAATTTTAGGTTAGAGGGTTAT
Rev CAGGAAACAGCTATGACCTAACTACAACCAAATAAACCCC

SRBC AF408198

Fwd: TGTAAAACGACGGCCAGTTGGGGTTAATAGGTTTTTTAGTAGG
Rev: CAGGAAACAGCTATGACCAACTCCAACTATAACTCAAACAAAC

CDH1 amplicon (bisulfite converted)

1 TGTAGGTTTTATAATTTATTTAGATTTTAGTAATTTTAGGTTAGAGGGTTATCGCGTTTA
61 TGCGAGGTCGGGTGGGCGGGTCGTTAGTTTCGTTTTGGGGAGGGGTTCGCGTTGTTGATT
121 GGTTGTGGTCGGTAGGTGAATTTTTAGTTAATTAGCGGTACGGGGGGCGGTGTTTTCGGG
181 GTTTATTTGGTTGTAGTTACGTATTTTTTTTTAGTGGCGTCGGAATTGTAAAGTATTTGT

SRBC amplicon

721 AAATTTAGAGTGAGAGGGTTTGTAGGGGGTCGATTTGGGGTTAATAGGTTTTTTAGTAGG
781 TTTTCGGCGCGGGATAGCGGAAGGCGAAACGTTTTTAAGAGATTTCGTTGTTAATATTTT
841 TACGTTTTCGCGTTTTTTCGTCGTTTTAGAAGGTTAATTTCGTTTGTTTGAGTTATAGTT
901

GGAGTTGGGGAGGAGTTAGGGAAAGGAGGTTTTTGATCGTAGTGCGGTTAGTAGTTGTAG

A denaturant is included in the PCR reaction to reduce PCR bias. Prepare a solution containing 5 mg/mL BSA and 5% glycerol:

250 uL of 20 mg/mL BSA solution (Sigma B8667)

700 uL of molecule biology grade water (Sigma W4502)

50 uL of molecular biology-certified glycerol (Shelton Scientific IB15760)

Required Materials for PCR:

AmpliTaq Gold® DNA Polymerase, 10X Gold buffer and 25 mM $MgCl_2$ (Applied Biosystems P/N 4311814)
dNTP's (P/N N808-0007)
For a 5 uL reaction:

Gold 10X buffer	0.5
dNTP 2.5 mM each	0.4
$MgCl_2$ 25 mM	0.4
AmpliTaq Gold polymerase (5U/uL)	0.1
Fwd primer 5 uM	0.25
Rev Primer 5 uM	0.25
Bisulfite-gDNA template 6ng/uL	0.5
BSA-glycerol solution	0.5
Water	2.1

The following thermal cycling conditions are used with M13 tailed primers. The additional 60 minutes at 60 ºC at the end of the thermal cycling is to allow full conversion to the non-templated A-addition amplicon-product. Complete A-addition (and not partial) is important when analyzing by fragment analysis, but not necessary if the amplicon is for sequencing only.

	95 ºC /5 min (activate AmpliTaq Gold® DNA Polymerase)
5X	95 ºC /30 sec
	60 ºC /2:00 min
	72 ºC / 3:00 min
35X	95 ºC /30 sec
	65 ºC /1:00 min
	72 ºC / 3:00 min

60 ºC /60 min

4 ºC hold until storage

For faster thermalcycling, when PCR bias or quantitative representation of the methylation information is not a concern, a simple 3 step program can be used:

95 ºC/5:00 min

40X 95 ºC /30 sec
60 ºC /2:00 min
72 ºC / 45 sec

4 ºC hold until storage

Estimate of PCR Amplicon Concentration

Run 1/10th of the PCR reactions on a 2% E-Gel to determine quality of PCR products. Use Invitrogen's low molecular weight standard as size/quantitation standard. If the amplicons are FAM labeled and are analyzed by fragment analysis by capillary electrophoresis using GeneMapper® software, an estimate of the amount of amplicon can be obtained from the CE analysis. The fragment analysis protocol has been reported elsewhere [34].

ExoSAP-IT® Treatment

Required Materials:

ExoSAP-IT® (USB) (P/N 78201)

Protocol:

1. Add 2 uL of ExoSAP-IT® to each 5-uL well of PCR products. Make sure the enzymes are added and mixed well with the PCR products.

2. Cover plate and thermal cycle in a 9700 Thermal Cycler with the following cycling profile:

37°C	30 min
80°C	15 min
4°C	hold until storage

3. Remove plate from thermal cycler and spin down contents. Make sure contents are in the bottom of each well.

4. Adjust all amplicons to 1-5 ng/uL concentration based on the agarose gel results.

5. The ExoSAP-IT® treated samples can now be used for sequencing reactions.

Cycle Sequencing Protocol

Required Materials:

BigDye® Terminator v1.1 kit ((P/N 4337450) or 3.1 (P/N 4337455)

M13 Fwd or Rev primer (3.2 uM)

Protocol:

1. Add the following reagents in a well to set up your sequencing reaction. Use 1 ng or less when sequencing reactions are cleaned-up with the BigDye® XTerminator™ Purification Kit:

PCR amplicon (bisulfite treated, 1-5 ng/uL)	1 uL
BigDye® Terminator Ready Reaction Mix	8 uL
Primer (M13 Forward or Reverse, 3.2 uM))	1 uL
DH2O	qs
Total volume	20 uL

2. Seal the plate with Thermal seal and quickly spin down the contents in each well in a centrifuge. Place the plate in a 9700 Thermal Cycler. Run the following Sequencing Cycling Profile on 9700 Thermal Cycler:

96° C	1 min	
96° C	10 sec	\
		25X
50° C	4 min	/
4° C	hold until storage	

Post Cycle-Sequencing Clean-Up Protocol

The preferred protocol for removing unincorporated dye terminators and unused primer is with Applied Biosystems BigDye® XTerminator™ Purification Kit (P/N 4376486, following manufacturer instructions). A run module to accommodate injection from a larger volume with the purification resin settled on the bottom of the well must first be installed.

Alternative Post-sequencing Clean-Up Protocol with Centri-Sep® Columns

A pre-treatment of samples with hot SDS to help with removal of unincorporated dye terminators is needed when using column clean-up methods.

Required Materials:
2.2% SDS
9700 Thermal Cycler

SDS Pre-treatment Protocol:

1. Prepare 2.2% SDS in deionized water from stock solution. This SDS solution is stable at room temperature.
2. Add an appropriate amount of the 2.2% SDS solution to your sample to bring the final SDS concentration to 0.2%.

   For example: Add 2 uL of 2.2% SDS to each 20uL completed cycle sequencing reaction.
3. Seal the plate and mix thoroughly.
4. Heat the plate to 98°C for 5 min and then allow the plate to cool to ambient temperature before proceeding to the next step.

NOTE: A convenient way to perform this heating/cooling cycle is to place the plate in a thermal cycler and set it as follows:

98°C for 5 min

25°C for 10 min

5. Spin down the contents of the plate briefly.
6. Continue with spin column or 96-well plate purification.

Sequencing Cleanup Protocol

Required Materials: Princeton Separations Centri-Sep® Products (Columns, 8-Strip, or 96-well)

Please follow manufacturer's recommended protocol when performing post cycle sequencing clean-up.

VIII Methylation **KIT COMPARISONS**

A comparative study of the methylSEQr Bisulfite Conversion Kit vs. an alternative commercially available kit was performed. Identical aliquots of the same Coriell gDNA were bisulfite converted and purified following the exact manufacturers' protocol provided with the kits. A carrier RNA was used with the alternative kit to ensure the greatest possible recovery. The recovery volumes were adjusted to be equal. One uL aliquots of the final bisulfite-treated gDNA were used in 7 different PCR reactions.

Summary of results:

- The methylSEQr™ Bisulfite Conversion Kit provided overall better recovery.
- The stability of the final bisulfite converted product is much higher using the methylSEQr kit, lasting 1 year or more when refrigerated. (providing no growth occurs in the solution)
- Elapsed reaction times and centrifugation times differ, but the actual "hands-on time" was similar.

Feature	methylSEQr	Alternative kit
Reaction Time	Overnight	5 hours
Technician Time*	3 hours	2 hours
Temperature	50 degree (isothermal)	50-95 (cycling)
Desulfonation	Solution phase (in situ)	On the column (in situ)
DNA recovery	Dissolved in TE (tris-EDTA)	Elution buffer
Yield**	Consistent	Variable (Occasional loss)
Stability	4 deg/indefinite (1+ yr)	-20 deg/12 weeks

* Actual hand's on time to process 10 samples. Includes a 0.5-1 hr set-up and a 1-2.5 hr purification (adding and removing wash solutions, etc.) and tube labeling.
** Based on comparative analysis of 75 ng or greater bisulfite reaction (~150 uL scale) of a Coriell gDNA.

IX Q&A

Q: How do I get rid of incomplete bisulfite conversion?

A: Several recommendations:

- Do not exceed 400 ng gDNA per bisulfite conversion
- Incubate the gDNA using the methylSEQr denaturation buffer for a longer (25 min) time as well as a slightly higher temperature (38-42^{o}C) to fully denature the gDNA

- Be sure to store the methylSEQr denaturation buffer tightly capped to prevent neutralization of the solution due to CO_2 in the atmosphere.
- Evaporation due to improper capping could result in an increased molarity of the methylSEQr denaturation buffer (which contains NaOH), adversely affecting the final pH of the bisulfite conversion reaction.

Q: How does heat denaturation of the gDNA compare to a pre-denaturation under basic condition?

A: Heat denaturation is more likely to fragment the gDNA, reducing PCR yields especially for amplicons above 500 bp. *It is safer, especially if analyzing different genes everyday, to avoid heating the gDNA in the bisulfite solution above 50^{o}C* . Harsher denaturation conditions may improve bisulfite conversion in high CG-regions, but may be an overkill for other regions.

Q: How does the use centrifugal filtration for bisulfite conversion clean-up in the methylSEQr kit compare to using a resin based purification?

A: The recovery of the bisulfite-converted gDNA is reliably higher with size exclusion because purification does not depending on the binding efficiencies of the gDNA to a resin where leaching or permanent sticking can occur. The desulfonation step is achieved in solution, and not while the DNA is bound to a resin.

Q: What are the maximum and minimum amounts of gDNA that can be safely used with the methylSEQr kit?

A. Approximately 100-300 ng of human gDNA is an optimal amount of gDNA per bisulfite conversion when scaled to the methylSEQr kit. Success with smaller quantities was demonstrated (Boyd and Zon, Anal Biochem 326 (2004) 278-280), but the small amount of gDNA that is bisulfite treated significantly limits the number of downstream analyses (PCR reactions) that can be done, with only half of the targeted amplicons providing a PCR product in the cited study. Larger amounts of gDNA during the bisulfite conversion (> 300 ng in the 150 uL total volume) will fail to completely bisulfite convert as noted above.

Q: How stable is the bisulfite converted gDNA?
A: With the methylSEQr kit, the bisulfite-converted gDNA has been demonstrated to be stable indefinitely at 4^{o}C (>1yr). The clean-up protocol thoroughly removes any impurities that may lead to degradation of the single-stranded uracil-containing bisulfite-converted product. Storage in the refrigerator avoids loss of material due to multiple freeze-thaws.

X Online resources for gDNA methylation information

There are several online resources that provide methylation information. The listed resources help with primer design, provide lists of genes silenced by methylation, and general information on the mechanism and importance of methylation in epigenics.

http://marketing.appliedbiosystems.com/mk/get/GAAS_CLINICAL_METHYLATED?_A=77005&_D=50613&_V=0
http://www.mdanderson.org/departments/methylation/
http://www.missouri.edu/~hypermet/list_of_promoters.htm
http://www.methdb.net
http://www.dnamethsoc.com/
http://www.faculty.iu-bremen.de/ajeltsch/name/index.htm

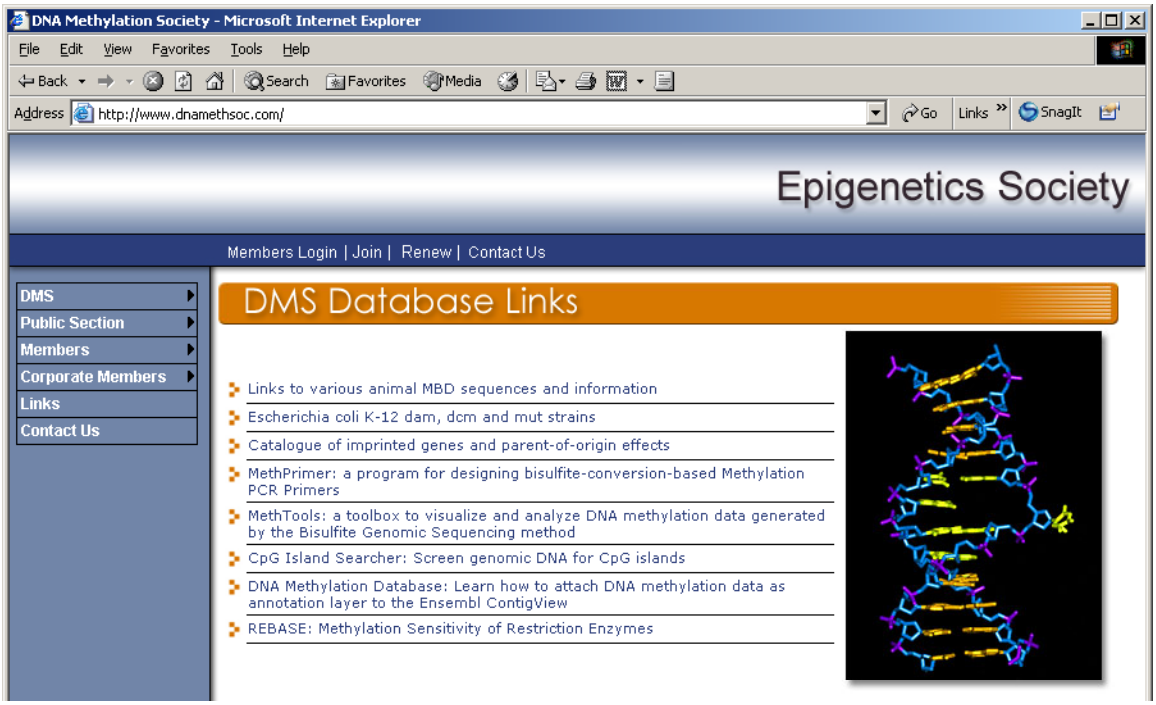


Figure 36. There are several valuable online resources for information on methylation of gDNA.

XI References

[1] P. A. Jones, and D. Takai, The role of DNA methylation in mammalian epigenetics, Science 293 (2001) 1068-1070.
[2] J. P. Issa, Methylation and prognosis: of molecular clocks and hypermethylator phenotypes, Clin Cancer Res 9 (2003) 2879-2881.
[3] K. L. Novik, I. Nimmrich, B. Genc, S. Maier, C. Piepenbrock, A. Olek, and S. Beck, Epigenomics: genome-wide study of methylation phenomena, Curr Issues Mol Biol 4 (2002) 111-128.
[4] G. A. Garinis, G. P. Patrinos, N. E. Spanakis, and P. G. Menounos, DNA hypermethylation: when tumour suppressor genes go silent, Hum Genet 111 (2002) 115-127.
[5] M. Widschwendter, and P. A. Jones, DNA methylation and breast carcinogenesis, Oncogene 21 (2002) 5462-5482.
[6] M. Widschwendter, and P. A. Jones, The potential prognostic, predictive, and therapeutic values of DNA methylation in cancer. Commentary re: J. Kwong et al., Promoter hypermethylation of multiple genes in nasopharyngeal carcinoma. Clin. Cancer Res., 8: 131-137, 2002, and H-Z. Zou et al., Detection of aberrant p16 methylation in the serum of colorectal cancer patients. Clin. Cancer Res., 8: 188-191, 2002, Clin Cancer Res 8 (2002) 17-21.
[7] E. J. Oakeley, DNA methylation analysis: a review of current methodologies, Pharmacol Ther 84 (1999) 389-400.
[8] H. Hayatsu, Y. Wataya, K. Kai, and S. Iida, Reaction of sodium bisulfite with uracil, cytosine, and their derivatives, Biochemistry 9 (1970) 2858-2865.
[9] T. Aranyi, A. Varadi, I. Simon, and G. E. Tusnady, The BiSearch web server, BMC Bioinformatics 7 (2006) 431.
[10] G. E. Tusnady, I. Simon, A. Varadi, and T. Aranyi, BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes, Nucleic Acids Res 33 (2005) e9.
[11] P. M. Warnecke, C. Stirzaker, J. Song, C. Grunau, J. R. Melki, and S. J. Clark, Identification and resolution of artifacts in bisulfite sequencing. Methods 27 (2002) 101-107.
[12] H. Hayatsu, K. Negishi, and M. Shiraishi, Accelerated bisulfite-deamination of cytosine in the genomic sequencing procedure for DNA methylation analysis, Nucleic Acids Symp Ser (Oxf) (2004) 261-262.
[13] J. Mill, S. Yazdanpanah, E. Guckel, S. Ziegler, Z. Kaminsky, and A. Petronis, Whole genome amplification of sodium bisulfite-treated DNA allows the accurate estimate of methylated cytosine density in limited DNA resources, Biotechniques 41 (2006) 603-607.
[14] D. W. Bianchi, Will Epigenetic Allelic Ratio Analysis Turn Prenatal Diagnosis of Trisomy 18 on Its EAR?, Clin Chem 52 (2006) 2182-2183.
[15] Y. K. Tong, C. Ding, R. W. Chiu, A. Gerovassili, S. S. Chim, T. Y. Leung, T. N. Leung, T. K. Lau, K. H. Nicolaides, and Y. M. Lo, Noninvasive prenatal detection of fetal trisomy 18 by epigenetic allelic ratio analysis in maternal plasma: theoretical and empirical considerations, Clin Chem 52 (2006) 2194-2202.
[16] C. Grunau, S. J. Clark, and A. Rosenthal, Bisulfite genomic sequencing: systematic investigation of critical experimental parameters, Nucleic Acids Res 29 (2001) E65-65.

[17] V. L. Boyd, and G. Zon, Bisulfite conversion of genomic DNA for methylation analysis: protocol simplification with higher recovery applicable to limited samples and increased throughput, Anal Biochem 326 (2004) 278-280.
[18] R. Goyal, R. Reinhardt, A. Jeltsch, Accuracy of DNA methylation pattern preservation by Dnmt1 methyltransferase, NAR 34 (2006) 1182-1188.
[19] J.-P. Issa, Age related epigenetic changes and the immune system, Clinical Immunology 109 (2003) 103-108.
[20] B. Richardson, DNA methylation and autoimmune disease, Clinical Immunology 109 (2003) 72-79.
[21] B. Reinhart, A. Paoloni-Giacobino, and J. R. Chaillet, Specific differentially methylated domain sequences direct the maintenance of methylation at imprinted genes, Mol Cell Biol 26 (2006) 8347-8356.
[22] C. Kress, H. Thomassin, and T. Grange, Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks, Proc Natl Acad Sci U S A 103 (2006) 11112-11117.
[23] C. Jeronimo, R. Henrique, M. O. Hoque, E. Mambo, F. R. Ribeiro, G. Varzim, J. Oliveira, M. R. Teixeira, C. Lopes, and D. Sidransky, A quantitative promoter methylation profile of prostate cancer, Clin Cancer Res 10 (2004) 8472-8478.
[24] S. Saxonov, P. Berg, and D. L. Brutlag, A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters, Proc Natl Acad Sci U S A 103 (2006) 1412-1417.
[25] J. G. Herman, J. R. Graff, S. Myohanen, B. D. Nelkin, and S. B. Baylin, Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands, Proc Natl Acad Sci U S A 93 (1996) 9821-9826.
[26] V. K. Rakyan, T. Hildmann, K. L. Novik, J. Lewin, J. Tost, A. V. Cox, T. D. Andrews, K. L. Howe, T. Otto, A. Olek, J. Fischer, I. G. Gut, K. Berlin, and S. Beck, DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project, PLoS Biol 2 (2004) e405.
[27] J. D. Hoheisel, A. G. Craig, and H. Lehrach, Effect of 5-bromo- and 5-methyldeoxycytosine on duplex stability and discrimination of the NotI octadeoxynucleotide. Quantitative measurements using thin-layer chromatography, J Biol Chem 265 (1990) 16656-16660.
[28] A. Roychowdhury, H. Illangkoon, C. L. Hendrickson, and S. A. Benner, 2'-deoxycytidines carrying amino and thiol functionality: synthesis and incorporation by Vent (exo-) polymerase, Org Lett 6 (2004) 489-492.
[29] R. S. Lasken, D. M. Schuster, and A. Rashtchian, Archaebacterial DNA polymerases tightly bind uracil-containing DNA, J Biol Chem 271 (1996) 17692-17696.
[30] P. M. Warnecke, C. Stirzaker, J. R. Melki, D. S. Millar, C. L. Paul, and S. J. Clark, Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA, Nucleic Acids Res 25 (1997) 4422-4426.
[31] W. S. Oetting, H. K. Lee, D. J. Flanders, G. L. Wiesner, T. A. Sellers, and R. A. King, Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers, Genomics 30 (1995) 450-458.
[32] R. H. Don, P. T. Cox, B. J. Wainwright, K. Baker, and J. S. Mattick, 'Touchdown' PCR to circumvent spurious priming during gene amplification, Nucleic Acids Res 19 (1991) 4008.

[33] K. O. Voss, K. P. Roos, R. L. Nonay, and N. J. Dovichi, Combating PCR bias in bisulfite-based cytosine methylation analysis. Betaine-modified cytosine deamination PCR, Anal Chem 70 (1998) 3818-3823.
[34] V. L. Boyd, K. I. Moody, A. E. Karger, K. J. Livak, G. Zon, and J. W. Burns, Methylation-dependent fragment separation: direct detection of DNA methylation by capillary electrophoresis of PCR products from bisulfite-converted genomic DNA, Anal Biochem 354 (2006) 266-273.
[35] G. Hu, DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment, DNA Cell Biol 12 (1993) 763-770.
[36] J. M. Clark, C. M. Joyce, and G. P. Beardsley, Novel blunt-end addition reactions catalyzed by DNA polymerase I of Escherichia coli, J Mol Biol 198 (1987) 123-127.
[37] J. M. Clark, Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases, Nucleic Acids Res 16 (1988) 9677-9686.
[38] J. M. Dupont, J. Tost, H. Jammes, and I. G. Gut, De novo quantitative bisulfite sequencing using the pyrosequencing technology, Anal Biochem 333 (2004) 119-127.
[39] D. Shinde, Y. Lai, F. Sun, and N. Arnheim, Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites, Nucleic Acids Res 31 (2003) 974-980.
[40] S. Mehra, and W. S. Hu, A kinetic model of quantitative real-time polymerase chain reaction, Biotechnol Bioeng 91 (2005) 848-860.
[41] C. A. Eads, K. D. Danenberg, K. Kawakami, L. B. Saltz, C. Blake, D. Shibata, P. V. Danenberg, and P. W. Laird, MethyLight: a high-throughput assay to measure DNA methylation, Nucleic Acids Res 28 (2000) E32.
[42] J. Albert, and E. M. Fenyo, Simple, sensitive, and specific detection of human immunodeficiency virus type 1 in clinical specimens by polymerase chain reaction with nested primers, J Clin Microbiol 28 (1990) 1560-1564.

Stock #107MI10-02