Scientific article

Techniques to overcome bottlenecks in epigenetics research

Gerald Schock and Thorsten Traeger

QIAGEN GmbH, Qiagen Straße 1, 40724 Hilden, Germany

Abstract: This article discusses the obstacles in bisulfite conversion of DNA for epigenetics research applications and uncovers possible solutions to overcome them. Furthermore, various methods for methylation analysis such as high-resolution melting (HRM®), methylation-specific PCR (MSP), MethyLight PCR, bisulfite sequencing, and Pyrosequencing® are described, and their respective uses and limitations are evaluated.

Introduction

DNA methylation, the process of adding a methyl group to cytosines, is one of the key epigenetic mechanisms in mammals. It is closely associated with histone modification, and it alters the chromatin structure to enable control and modulation of gene expression. DNA methylation plays a crucial role in mammalian development, genomic imprinting, and in genomic stability. Because its deregulation can both cause or result from abnormal cell regulation in cancer, studies in DNA methylation are expected to lead to the development of effective biomarkers and drugs against cancer or other malignant diseases.

Assessing DNA methylation typically comprises three major steps — DNA isolation, bisulfite conversion, and analysis of the bisulfite converted DNA. Although all of these steps are important, bisulfite conversion of DNA is probably the most crucial step for obtaining reliable and accurate methylation data.

Contents

DNA methylation — an important epigenetic mechanism

The genetic information for building proteins and the regulation of this process is not only encoded by the DNA sequence itself. Epigenetic mechanisms alter gene expression without a change in the actual DNA sequence. DNA methylation and histone modification are the most prominent epigenetic mechanisms, and these act together and with other regulatory proteins to remodel chromatin into a transcriptionally active (euchromatin) or inactive (heterochromatin) form, resulting in heritable alterations in gene expression profiles. CpG methylation occurs at position 5 of the pyrimidine ring of cytosines in DNA. In mammals, transfer of the methyl group (-CH3) is carried out by DNA methyltransferases, and is observed at cytosine residues in CpG dinucleotides — sites in the DNA in which a cytosine nucleotide is positioned directly 5' of a guanine nucleotide. This process typically triggers histone deacetylation and chromatin condensation, leading to gene silencing. CpG methylation is not uniformly distributed throughout the entire genome, but is instead concentrated in genes that regulate transcription, growth, metabolism, differentiation, and oncogenesis. Hence, differentially methylated cytosines create methylation patterns that are specific for different tissue types, differentiation status, and disease states.

DNA methylation in disease progression

DNA methylation is a well-balanced process in healthy cells. But in some diseases, in particular cancers, aberrant CpG methylation changes may occur, and both genomewide hypomethylation and gene-specific hypermethylation in promoters of tumor suppressor genes (1) can be observed. Many studies have shown that tumor emergence and growth are associated with changes in DNA methylation patterns, such as the case of prostate cancer development (2). Therefore, one of the most important application areas for epigenetic research is in the diagnosis and treatment of cancer — using DNA methylation patterns to detect cancer at very early stages, to classify tumors, and to predict and monitor response to drug treatments. Some demethylation agents have already been approved for therapeutic use. The hypomethylating agents 5-azacytidine and decitabine are used in the treatment of myelodysplastic syndromes, and several other compounds are currently in different stages of development. The central requirement for the realization of these applications is the ability to accurately and reliably measure CpG methylation, especially if only subtle changes in DNA methylation patterns need to be assessed.

Bisulfite conversion — the fundamental step in DNA methylation analysis

Most analysis methods cannot directly differentiate between methylated and unmethylated cytosines, so a technique to easily distinguish between the two residues and to provide this information in a simple-to-interpret format is necessary. The sodium bisulfite conversion reaction, in which sodium bisulfite converts normal cytosine residues in DNA to uracil residues, while leaving the 5-methylcytosine residues unchanged, has become central to DNA methylation analyses. During PCR amplification of bisulfite converted DNA, uracil residues in the template are replaced by thymines. The bisulfite conversion therefore introduces specific changes in the DNA sequence that reflect the methylation status of individual cytosine residues — unmethylated cytosines are substituted by thymines, whereas methylated cytosines remain cytosines (Figure 1). Subsequent analysis of the bisulfite converted DNA by techniques such as PCR, high resolution melting (HRM) analysis, Sanger sequencing, or Pyrosequencing, can easily provide extremely high resolution information about the methylation status of a DNA segment.

Figure 1. Bisulfite conversion of DNA. Only unmethylated cytosines are converted to uracil and subsequently to thymine during PCR. Methylated cytosines are unaffected by the bisulfite reaction. Methylated and unmethylated cytosines can therefore be detected by comparing bisulfite converted DNA to original untreated genomic DNA.

Bisulfite conversion is the most critical step in DNA methylation analysis, because the reaction efficiency has a huge impact on the reliability of the downstream analysis methods. If conversion is not complete, the subsequent analysis will

incorrectly interpret the unconverted unmethylated cytosines as methylated cytosines, producing false-positive findings for methylation. A further complication is that only cytosines in single-stranded DNA are susceptible to bisulfite conversion, so efficiently maintaining denaturation of the DNA throughout the conversion reaction is crucial for obtaining complete conversion. This could easily be achieved by increasing the reaction temperature, but in combination with the other harsh conditions necessary for complete conversion, such as long incubation times and low pH, the incubated DNA would be heavily degraded, significantly lowering the yield and sensitivity of the subsequent analysis.

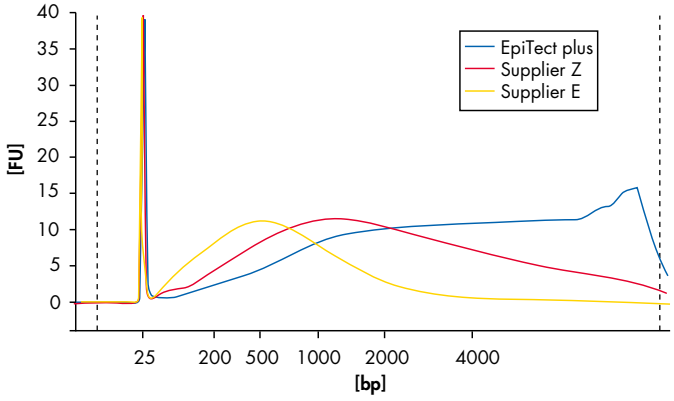

Figure 2. Analysis of DNA fragment sizes after bisulfite conversion. DNA protect mechanisms are critical for preventing DNA fragmentation during the bisulfite conversion reaction. One microgram of genomic DNA that has undergone bisulfite conversion with the EpiTect® Plus Bisulfite Kit shows the lowest level of fragmentation with a size distribution from several 100 bp up to over 6 kb, compared to other methods with peaks at 500 bp and 1 kb, respectively.

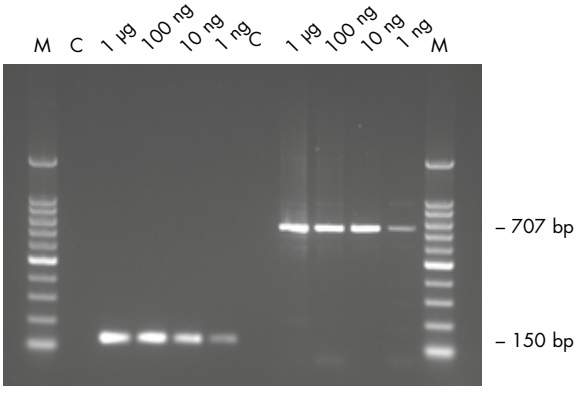

Figure 3. Amplification of large PCR products from minimal amounts of template DNA. Human genomic DNA was purified from blood using the QIAamp® DNA Blood Mini Kit, and various amounts (1ng – 1µg) were converted using the EpiTect Bisulfite Kit. PCR was performed using HotStarTaq® Master Mix Kit and two sets of primers designed to amplify converted DNA. A 5 µl aliquot of each PCR was loaded onto a 1.3 % agarose gel. As little as 1 ng DNA is sufficient for conversion using the EpiTect Bisulfite Kit. C: untreated genomic DNA (negative control). M: marker.

In recent years, a method to overcome these difficulties has been developed, using a DNA protect mechanism to prevent DNA fragmentation, while still providing the effective DNA denaturation necessary for complete cytosine conversion. Such a DNA protect mechanism is available in QIAGEN's EpiTect line of solutions for bisulfite conversion (Figure 2). Preventing fragmentation also makes subsequent amplification and analysis of large PCR fragments possible (Figure 3) and enables sensitive methylation analysis of limited and precious sample materials derived from formalin-fixed paraffin embedded (FFPE) tissue samples. The DNA in these samples in particular can already be highly degraded from the fixation process, and any further degradation during purification and conversion should be avoided to ensure that enough DNA remains intact for successful analysis with the required sensitivity.

In addition to reducing DNA degradation, the chemistry used in the DNA protect mechanism also leads to high bisulfite conversion efficiencies. Bisulfite sequencing proves conversion efficiencies of 99.4 – 99.8% in reactions using this technology. It also enables higher efficiencies to be obtained by prolonging or repeating the conversion step; otherwise the extended conversion produces heavily fragmented DNA. This has been confirmed in a study by Meissner et al. (3), who reported on the generation and analysis of genome-scale DNA methylation profiles at nucleotide resolution in mammalian cells. For these studies, they successfully applied the DNA protect mechanism for the bisulfite conversion step.

Increasing yield and PCR sensitivity through direct bisulfite conversion without prior DNA isolation

The usefulness of bisulfite converted DNA in downstream applications depends on two factors — DNA quantity and DNA quality. High-quality DNA with a low degree of fragmentation is crucial for PCR success, as any strand breakage within the PCR amplification target sequence leads to PCR failure. Therefore, using larger fragments of bisulfite converted DNA as the template enhances PCR sensitivity and decreases amount of template needed.

The quantity of bisulfite converted DNA obtained is primarily influenced by the amount of starting DNA used in the bisulfite

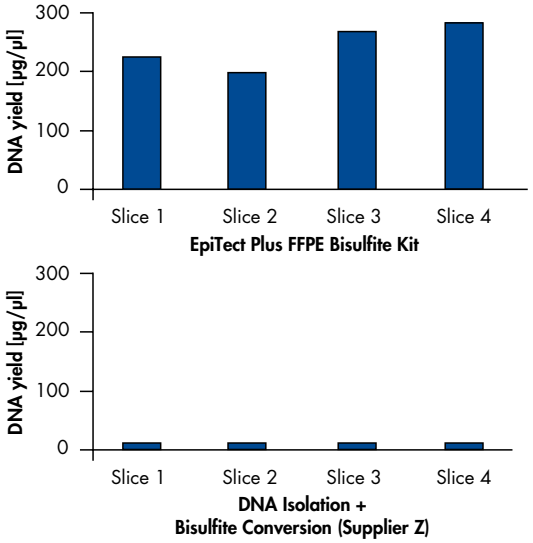

Figure 4. Quantification of bisulfite converted DNA. DNA from rat liver FFPE tissue (10 µM) was purified and bisulfite converted. DNA yield in the eluates was quantified by optical density measurement. The EpiTect Plus FFPE protocol, which combines lysis of FFPE tissue and bisulfite conversion, yields on average 25 times more DNA than a protocol requiring two separate steps for DNA isolation and bisulfite conversion.

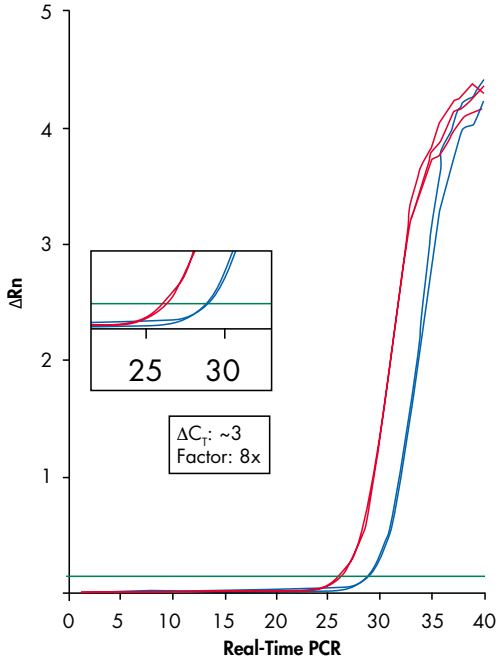

Figure 5. Real-time PCR of bisulfite converted DNA. DNA from rat liver FFPE tissue was bisulfite converted as described in Figure 4, and the same amounts were used for real-time PCR of a 280 bp fragment. The ΔC_T of approximately 3 corresponds to an 8-fold higher PCR sensitivity with DNA converted with the EpiTect Plus Bisulfite procedure.

conversion reaction. This can be increased by simply using more efficient DNA isolation methods to obtain higher yields of starting DNA. Instead, the most efficient strategy is to use all of the DNA released from cell or tissue lysis directly in the bisulfite conversion reaction. QIAGEN® has therefore developed dedicated procedures for lysing cells, whole blood, fresh-frozen tissue, and even FFPE samples, that also enable bisulfite conversion directly in the resulting lysates, eliminating DNA loss and increasing DNA yield by a factor of 25 (Figure 4). Quantitative PCR Analysis of DNA produced by this integrated lysis/conversion process results in lower C_T values than those from DNA produced by the two-step process and without a DNA protect mechanism (Figure 5). This ΔC_T of approximately 3 corresponds to an 8-fold higher PCR sensitivity. Therefore, combining a direct bisulfite conversion protocol with DNA protect technology showed a PCR sensitivity which was 200-fold higher compared to the standard protocol without DNA protection.

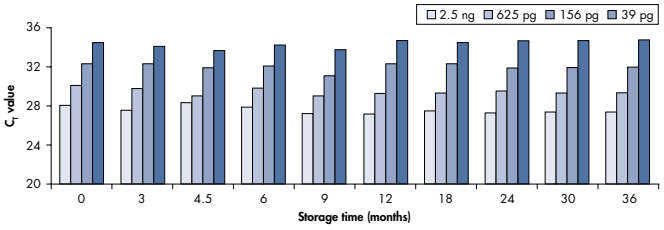

Figure 6. Stability of bisulfite converted DNA upon storage. A real-time PCR assay using the EpiTect MethyLight PCR Kit and primers specific for converted DNA was performed on various amounts of DNA converted using the EpiTect Bisulfite Kit before and after DNA storage in elution buffer at –20°C for 3, 4, 5, 6, 9, 12, 18, 24, 30, and 36 months. Even low concentrations of DNA converted using the EpiTect Bisulfite Kit show no loss of DNA quality upon storage.

Stability of bisulfite treatment — a challenge for storing converted DNA for downstream analysis

Bisulfite converted DNA can undergo significant degradation upon storage, lowering the sensitivity of DNA methylation assays and making this DNA unsuitable for further analysis after years, months, or even after weeks. Researchers recommend against long-term storage of bisulfite converted DNA — even if it is frozen — and they generally try to completely use samples immediately after bisulfite treatment (4). The main cause of DNA degradation after bisulfite treatment is the presence of remaining reactants from the chemical conversion reaction. Combining a DNA protect mechanism and an efficient DNA purification step overcomes this problem and produces bisulfite converted DNA that is stable for at least 36 months (Figure 6).

Overcoming limitations due to small sample amounts

The amount of bisulfite converted DNA obtained is often too small to permit extensive DNA methylation analyses, and the conversion of additional sample material — if it is even available — may lead to irreproducible experimental conditions. This can be overcome by introducing an amplification step to produce more DNA for further analysis. This amplification must be carried out after the bisulfite conversion, because genomic DNA amplified before bisulfite treatment loses its epigenetic modifications during the amplification process.

DNA degradation during bisulfite conversion makes it difficult to apply methods such as multiple displacement amplification (MDA), which because of its reliable and unbiased amplification results, has become the gold standard for amplifying genomic DNA. This method of whole genome amplification (WGA) employs the Phi29 DNA polymerase, which exhibits a unique 3'–5' exonuclease proofreading activity, maintaining high fidelity throughout the replication process.

Recently however, the MDA method has also become available for whole bisulfitome amplification (WBA). The term bisulfitome refers to the entire bisulfite converted genome and is analogous to all transcripts in the transcriptome and genes in the genome. An amplification rate of up to 50-fold can be achieved with bisulfite converted DNA, and experiments have shown that the DNA amplified with this new technology produce reliable results in downstream real-time PCR and end-point PCR analyses. The WBA method does require use of a DNA protect mechanism during the bisulfite conversion to ensure the minimum fragment size required for the amplification is maintained. However, DNA derived from FFPE tissue samples may still be unsuitable for amplification due to high levels of degradation caused by the fixation process.

Screening for unknown methylation CpG positions — high-resolution melting analysis (HRM)

High-resolution melting analysis is a new technique that has raised enormous scientific interest for its potential as a valuable research tool. HRM analysis is an extension of previous DNA dissociation (or "melting") analyses, and it is used to characterize DNA samples according to their dissociation behavior as they transition from double-stranded DNA (dsDNA) to single-stranded DNA (ssDNA) with increasing temperature (Figure 7). It is similar to classical melting curve analysis, but it can provide much more information for a wider range of applications. Samples can be discriminated according to sequence, length, GC content, or strand complementarities, down to single-base pair alterations. Reliable and sensitive HRM results require the capture of many data points per run and a high well-to-well precision. Two scientific developments were necessary to enable the HRM technique. The first was the introduction of intercalating dyes that do not inhibit PCR at the concentrations needed to fully saturate the target DNA complexes (5). The second was the development of instruments that can accurately monitor the changes in fluorescence (4). Due to its special design, QIAGEN's new Rotor-Gene® Q real-time cycler combines the highest thermal precision with high-speed data capture, making it highly suited for HRM applications.

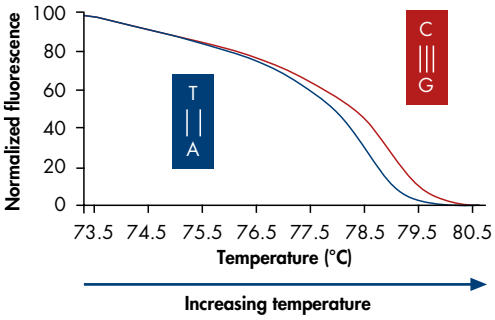

Figure 7. Principle of HRM technology. HRM analysis is based on the dissociation behavior of dsDNA due to increasing temperature. Melting of dsDNA depends on GC-content. AT-rich regions melt faster.

The central basis for measuring DNA methylation with HRM is that changes are introduced into DNA during the bisulfite conversion step, depending on the degree of methylation, and these changes can be measured by examining the melting behavior of the DNA. When used with bisulfite converted DNA standards, HRM can also be used to estimate the proportion of methylation within a sample (4). Because HRM analysis requires short PCR products, which can easily be generated from a large number of different samples, it is a fast and inexpensive tool suitable for high-throughput screening. HRM is easier and more cost-effective than probe-based assays and, unlike conventional methods, it is a closed-tube system that prevents carry-over contamination with PCR products (6).

Methylation-specific PCR (MSP) — an easy way to determine CpG methylation

For detecting CpG methylation following bisulfite conversion by PCR, primer pairs are designed to be either methylation-specific or non-methylation-specific. Methylation-specific primers complement sequences containing unconverted 5-methylcytosines while, in contrast, non-methylation-specific primers complement sequences containing uracils from conversion of unmethylated cytosines. The methylation status of a CpG site is determined by the ability of each primer to achieve amplification. However, designing primers for MSP is often difficult, because the sequence difference between bisulfite converted DNA from methylated and unmethylated sequences is very small. In addition, each primer must cover at least two or three CpG sites, further complicating primer design and making MSP unsuitable for analyzing single CpG sites. This limitation can be overcome by using novel genetically engineered polymerases, such as HotStarTaq d-Tect included in QIAGEN's EpiTect MSP Kit. This polymerase only extends primers that perfectly match the target DNA at the 3' primer end (the starting point of primer extension), enabling analysis of single CpG sites. This is in contrast to wild-type DNA polymerases that often ignore 3' mismatches and extend mis-primed primers, resulting in nonspecific, false-positive methylation results. Control experiments to demonstrate specific amplification of either methylated or unmethylated sequences after bisulfite conversion are highly recommended for every MSP system that has been designed. Although MSP can be highly sensitive, it does not allow quantification of the methylation status.

Quantifying CpG methylation by probe-based real-time PCR — MethyLight PCR

MethyLight PCR is a method for quantifying CpG methylation by quantitative real-time PCR with dual-labeled probes. Different formats of MethyLight assays can be used depending on the experimental purpose (7). For quantifying both methylated and unmethylated DNA in a single combined assay, a set of amplification primers and two different fluorescent labeled probes are required. The primers should complement methylation-insensitive sites. One probe is specific for a region of methylated DNA, and the other is specific for unmethylated DNA. Because the probes are linked to different fluorophores,

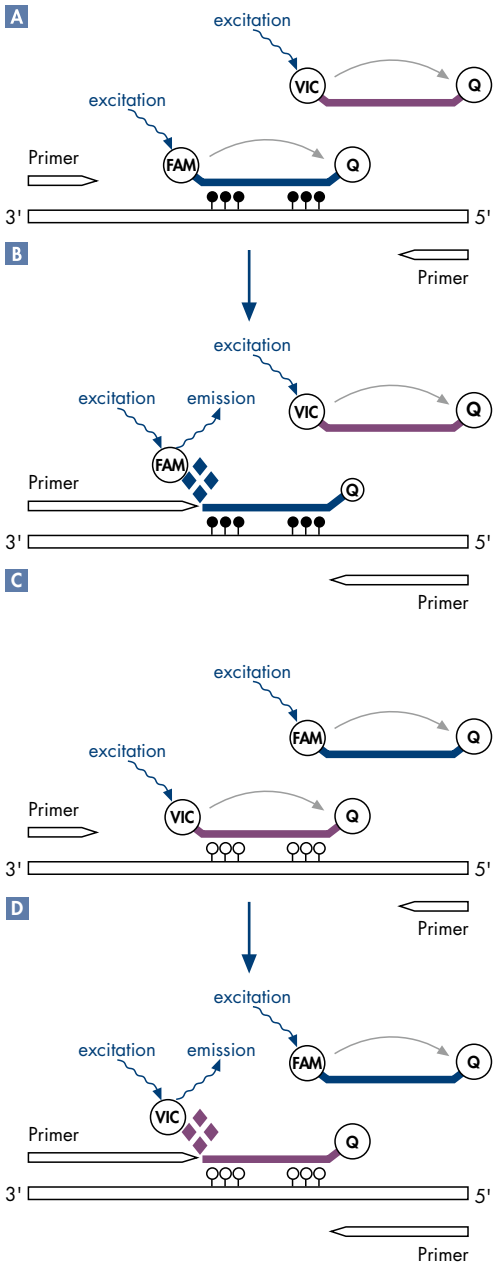

Figure 8. Principle of methylation-specific TaqMan® probes in quantitative MethyLight real-time PCR. In methylation-specific TaqMan assays, two methylation-sensitive TaqMan probes are present along with methylation-insensitive PCR primers. Depending on the methylation status of the targeted sequence, only the TaqMan probe specific for bisulfite converted methylated DNA A or the TaqMan probe specific for unmethylated DNA C can hybridize to the target sequence. Both probes are labeled with different fluorophores that are released during PCR if the probe hybridizes to the DNA B and D. The fluorescence from the released fluorophore is proportional to the amount of accumulated PCR product.

the degree of methylation in the sample can be determined by measuring the emission of both fluorophores in the same reaction in separate channels (Figure 8). Since methylated and unmethylated bisulfite converted DNA differ only in a few nucleotide positions, the reliability of MethyLight assays depends on both the specificity of the PCR primer annealing, as well as the binding of the probes. Specially-formulated systems for MethyLight PCR are available, allowing detection of differences in the methylation degree as low as 5%.

Sequencing of bisulfite converted DNA — bisulfite sequencing

The methylation status of CpG sites can also be determined by bisulfite sequencing, which simply uses routine sequencing methods, such as Sanger sequencing, on bisulfite treated genomic DNA to detect the sequence alterations resulting from bisulfite conversion. The main disadvantage of bisulfite sequencing is the difficulty in determining the methylation ratio by simultaneous sequencing of cytosines and thymines, which is required for quantification of the methylation degree of a specific CpG site. Achieving reliable and sensitive quantitative data requires cloning and sequencing a larger number of clones, a labor-intensive process that is particularly unsuitable for higher throughput. Pyrosequencing, which can simultaneously provide sequence data and sensitively measure the degree of methylation, overcomes this difficulty.

Pyrosequencing — combining sequencing with methylation quantification

The status and exact degree of methylation for single CpG sites is the highest level of data that can be obtained in DNA methylation analyses, and bisulfite Pyrosequencing makes this possible. This level of information is particularly important in cases where CpG islands with different methylation patterns correspond to different tumor types, or when the methylation status following treatment of patients with demethylating agents should be monitored in detail. Traditional bisulfite sequencing can provide sequence information on such CpG islands, and HRM can be used to measure changes in the methylation when compared to a reference sample. But these methods fall short when quantification of individual CpG sites in a sequence containing multiple CpG sites is required. Pyrosequencing provides both reliable quantification of single, consecutive CpG sites and sequence information in a reasonable time (8–10).

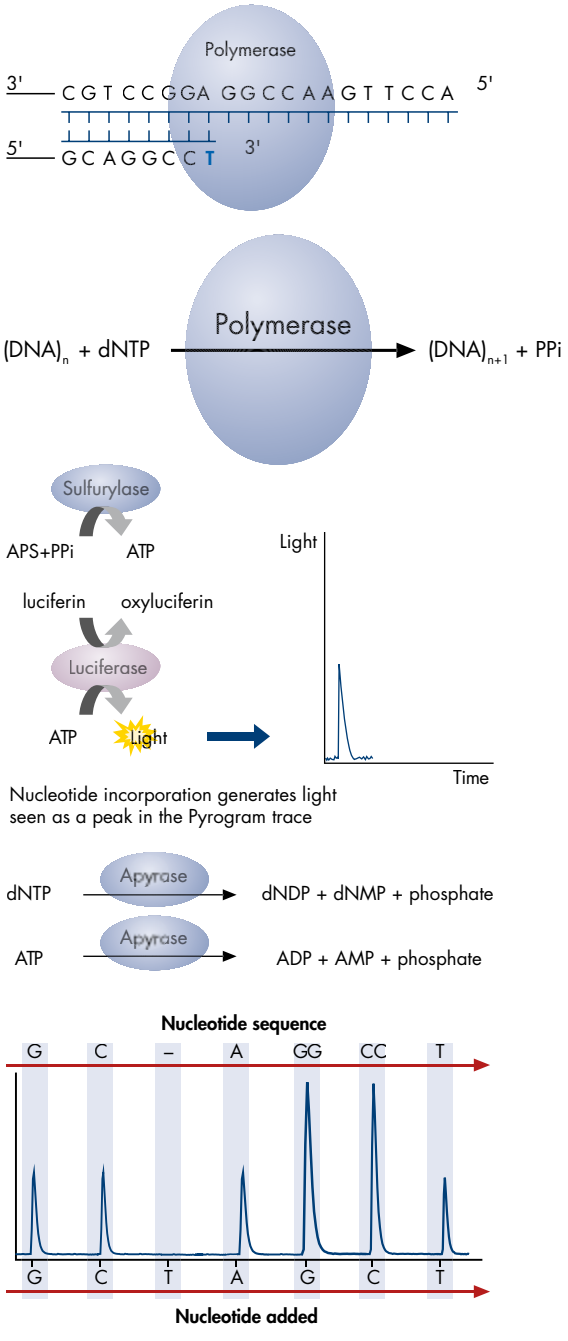

Figure 9. The Principle of Pyrosequencing. See text for details.

The principle of Pyrosequencing — sequence-based detection and quantification of individual methylation sites

The elegant chemistry of the Pyrosequencing reaction cascade, together with sensitive detection methods, is fundamental to the technology. An initial PCR amplification using a biotinylated primer introduces a biotin label on one of the DNA strands of the PCR amplicon. Following this amplification, streptavidin-coated sepharose beads are used to isolate the biotin-tagged single-stranded DNA, which is the template for the Pyrosequencing reaction. A sequencing primer is hybridized to this single-stranded template, and this complex is then incubated with the enzymes — DNA polymerase, ATP sulfurylase, luciferase, and apyrase — as well as the substrates — adenosine 5' phosphosulfate (APS) and luciferin (Figure 9).

A deoxribonucleotide triphosphate (dNTP) is added to the reaction, one by one, in a specific order. DNA polymerase catalyzes the incorporation of the dNTP into the new DNA strand, if it is complementary to the base in the template strand. Each incorporation event is accompanied by release of pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide (11). ATP sulfurylase converts PPi to ATP in the presence of adenosine 5' phosphosulfate (APS). This ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) camera and is recorded as a peak in the raw data output (Pyrogram®). The height of each peak (light signal) is proportional to the number of nucleotides incorporated. Apyrase, a nucleotide-degrading enzyme, continuously degrades unincorporated nucleotides and ATP. When degradation is complete, another nucleotide is added. Addition of dNTPs is performed sequentially. As the process continues, the complementary DNA strand is elongated, and the nucleotide sequence is determined from the signal peaks in the Pyrogram trace.

For methylation analysis, the DNA is bisulfite converted, and the ratio of C to T at individual sites can be determined quantitatively based on the amount of C and T nucleotide incorporated during the extension of the new strand. Thus, variation in the methylation status of various sites can be detected with high accuracy, and since Pyrosequencing provides real sequence data, the methylation status is presented in a sequence context (Figure 10). Pyrosequencing offers a built-in quality control for the bisulfite conversion efficiency, as it can easily be determined from each Pyrogram if and how many unmethylated cytosines are left in the converted DNA. In addition, Pyrosequencing enables extension to high-throughput sample processing.

Typically, CpG islands are first sequenced by conventional Sanger sequencing and then are further analyzed and quantified by other techniques. This time- and labor-intensive approach can be replaced by Pyrosequencing, in which several sequencing primers can be used to generate overlapping sequence information (8, 12). Combining various DNA methylation analysis methods is a common approach to overcome bottlenecks of some methods at certain stages within an epigenetics project. For example, methods for screening for methylation differences usually need to cover all genes of an entire pathway or even the entire genome, but are usually not precise enough to determine methylation at single CpG resolution, or are simply too cost-intensive for use in verification studies with large numbers of samples. In contrast, Pyrosequencing provides precise DNA methylation results at single CpG resolution but is unsuitable for screening the entire genome. A prerequisite for combining various methodologies for DNA methylation analysis is that all technologies used together must provide comparable results. This has recently been investigated in studies comparing Pyrosequencing with HRM and 454 Sequencing™, respectively (13, 14), where it was shown that Pyrosequencing can be successfully combined with these techniques, producing highly complementary methylation results.

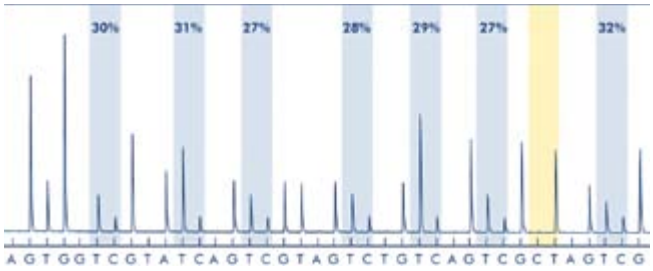

Figure 10. Pyrogram trace illustrating CpG methylation analysis of the p16 gene. Highlighted areas in the Pyrogram trace indicate variable CpG positions (light blue) and built-in bisulfite treatment controls (yellow). The methylation level of each CpG site is indicated at the top of the Pyrogram trace.

Control DNA and standard DNA — the key for standardized methylation analysis

Control reactions are a crucial part of any methylation analysis, and only a control reaction can indicate whether the DNA bisulfite conversion is complete. In addition, control reactions should be performed to ensure that the PCR primers for MSP experiments are specific for detecting bisulfite converted methylated or unmethylated DNA. Furthermore, mixtures of methylated and unmethylated DNA can be used as standards for quantifying CpG methylation in real-time PCR and HRM PCR experiments. Until recently, researchers had to prepare bisulfite converted control DNA themselves, but now, quality-controlled pre-bisulfite converted DNA — unmethylated or methylated — is available, allowing standardized and reliable control reactions for all types of methylation assays.

Conclusion

Epigenetics research has been in center stage for the last few years, and researchers have been able to correlate changes in methylation patterns with the development of severe diseases. However, many unanswered questions remain. Using the established techniques described here in a new application — epigenetic analysis — helps to overcome bottlenecks that have existed in screening, identification, and quantification of CpG sites in methylation analysis. The emergence of innovative technologies should produce greater time savings and assay standardization, which will facilitate research in topics related to human health, and lead to the development of effective biomarkers and drugs against cancer or other malignant diseases.